

A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/132293>

Copyright and reuse:

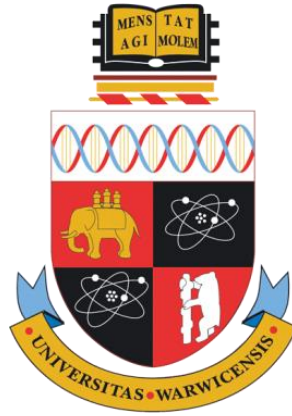
This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it.

Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk



**How well does a Preparatory Year Programme in a
Saudi university prepare students for the writing
demands of Medical Healthcare Colleges?**

**A gap analysis approach to needs analysis using the
CEFR scales**

By

Ebtesam Abdulaleem Abdulhaleem

A thesis submitted in partial fulfilment of the requirements for the
degree of Doctor of Philosophy in English Language Teaching and
Applied Linguistics

University of Warwick, Applied Linguistics

December 2018

To my family

Table of Contents

List of tables	viii
List of figures	xi
List of abbreviation	xii
Acknowledgments	xiii
Declaration	xv
Abstract	xvi
Chapter 1 Introduction	1
1.1. Introduction	1
1.2. Background	1
1.3. Context of the study	3
1.4. Statement of the problem.....	6
1.5. Research aim and objectives.....	10
1.6. Research questions.....	11
1.7. Rationale and motivation for the study	12
1.8. The significance of the study	13
1.9. My role as a researcher.....	14
1.10. Overview of the thesis structure.....	15
Chapter 2 Literature Review	16
2.1. Saudi Education.....	16
2.1.1. The implementation of the PYP	16
2.1.2. PYP appreciation and criticisms	18
2.1.3. Saudi MHCCs.....	20
2.2. Saudi students' English writing difficulties	22
2.2.1. Student-related difficulties	22
2.2.2. Assessment-related difficulties	25
2.2.3. Teaching-and-learning related difficulties.....	26
2.2.4. Effect of L1 on students' writing	27
2.3. Improving Saudi students' writing proficiency.....	28
2.4. Needs analysis/assessment (NA)	30
2.4.1. NA: purposes, approaches and types	33
2.4.2. NA: resources and tools.....	35
2.4.3. NA: reliability and validity	37

2.5. Needs analysis and language proficiency	38
2.6. Self-assessment.....	44
2.6.1. Purposes of self-assessment.....	44
2.6.2. Forms of self-assessment.....	45
2.6.3. Reliability and accuracy of self-assessment.....	46
2.7. The CEFR.....	52
2.7.1. What the CEFR is and is not	52
2.7.2. CEFR: theories, approaches and Scales.....	55
2.7.3. CEFR and pedagogy.....	56
2.7.4. CEFR and the curriculum.....	58
2.7.5. CEFR and assessment.....	59
2.7.6. CEFR and self-assessment.....	59
2.7.7. Criticism of the CEFR.....	62
2.8. Reasons for using the CEFR in this study.....	67
2.8.1. The way in which the CEFR is designed	67
2.8.2. The important role the CEFR plays as a common language.....	68
2.8.3. CEFR and its relevance to the current study.....	68
2.9. NA for MHCC students within the Saudi contexts	69
2.10. Implications for the current approach to NA	71
2.11. Defined research aims	72
Chapter 3 Research Methodology	76
3.1. Introduction	76
3.2. Research purpose and research questions	76
3.3. Research approach	77
3.3.1. Research paradigm.....	77
3.3.2. Research methodology.....	79
3.4. Research design.....	80
3.4.1. Research settings and participants.....	83
3.4.1.1. Participants in the quantitative strand.....	84
3.4.1.2. Participants in the qualitative strand	85
3.5. Research instruments.....	88
3.5.1. Quantitative instruments.....	88
3.5.1.1. Choosing CEFR scales relevant for the study context.....	88
3.5.1.2. Phase I questionnaires.....	90

3.5.1.3. Piloting Phase I questionnaires.....	92
3.5.1.4. Phase II questionnaires.....	94
3.5.1.5. Piloting Phase II questionnaires.....	95
3.5.1.6. Phase I: independent ratings of exam texts.....	96
3.5.1.7. Preparation for, and piloting of, standardisation and benchmarking session for raters.....	97
3.5.2. Qualitative instruments.....	98
3.5.2.1. Students' <i>letter to a friend</i>	98
3.5.2.2. FG meetings.....	99
3.5.2.3. Piloting FG questions and materials.....	102
3.6. Data collection.....	102
3.6.1. Ethical considerations.....	102
3.6.2. Phase I data collection procedure.....	104
3.6.2.1. Quantitative data.....	104
3.6.2.2. Qualitative data.....	106
3.6.3. Phase II data collection procedure.....	107
3.6.3.1. Quantitative data.....	107
3.6.3.2. Qualitative data.....	108
3.7. Data analysis.....	108
3.7.1. Quantitative analyses.....	108
3.7.1.1. Procedure for data entry.....	108
3.7.1.2. Phase I analysis.....	109
3.7.1.3. Phase II analysis.....	115
3.7.1.4. Phase III analysis.....	115
3.7.2. Qualitative analyses.....	117
3.7.2.1. Analysing students' letters.....	117
3.7.2.2. Analysing FGs and interviews.....	117
3.8. Limitations.....	119
Chapter 4 A present-situation analysis (Phase I)	122
4.1. Introduction.....	122
4.2. Students' CEFR levels at the end of the PYP.....	122
4.2.1. Descriptive analyses.....	124
4.2.2. Exploring and comparing participants' assessments across and within PYP levels.....	126
4.2.2.1. PYP students' self-assessment across PYP Levels.....	126

4.2.2.2. PYP tutors' assessment across PYP levels	129
4.2.2.3. Raters' assessment of students' written texts across PYP levels....	131
4.2.2.4. PYP self- and tutors' assessment within PYP levels	132
4.2.2.5. Comparing samples of self-, tutors and raters' assessment	136
4.2.3. Summary of the CEFR levels as perceived by PYP FG participants	140
4.3. Writing at the PYP	142
4.3.1. PYP Students: perceived proficiency, satisfaction and preparedness .	142
4.3.1.1. <i>Proficiency</i>	142
4.3.1.2. <i>Satisfaction</i>	145
4.3.1.3. <i>Preparedness</i>	149
4.3.2. Writing in the PYP curriculum: " <i>on the back burner</i> "	150
4.3.3. Vocabulary: " <i>the Core</i> "	152
4.3.4. Medical writing: " <i>getting blood out of a stone</i> "	154
4.3.5. Academic writing: " <i>combating plagiarism</i> "	156
4.3.6. " <i>One-size-fits-all</i> " the assessment of writing at the PYP.....	157
4.4. Phase I Discussion.....	160
4.4.1. RQ1: Students' CEFR proficiency levels at the end of the PYP	160
4.4.2. RQ2: Writing at the PYP: discussion of the qualitative findings	168
Chapter 5 A Target-situation Analysis (Phase II)	174
5.1. Introduction.....	174
5.2. The CEFR levels required of first-year MHCC students	174
5.2.1. Descriptive analysis of the CEFR levels perceived as required.....	176
5.2.2. Comparing students and teaching staff's perceptions of the CEFR levels required.....	177
5.2.3. Perceptions of the CEFR levels required across MHCCs.....	178
5.2.4. Perceptions of the CEFR levels required across students' PYP levels..	181
5.2.5. The CEFR levels perceived as required by FG/interview participants	185
5.3. Writing at the MHCCs.....	188
5.3.1. Writing in the MHCCs: three points of view.....	189
5.3.1.1. Writing for Exam Purpose: <i>highly required</i>	191
5.3.1.2. Writing for course tasks and assignments: <i>moderately required</i> ...	194
5.3.1.3. Writing for future use: <i>marginally required</i>	195
5.4. Phase II Discussion	197
5.4.1. RQ3: The minimum CEFR levels required of MHCCs.....	197
5.4.2. RQ4: Writing skills required in the MHCCs	201

Chapter 6 Mind the Gap! (Phase III)	211
6.1. Introduction	211
6.2. Misalignments between the CEFR levels achieved (Phase I) and required (Phase II).....	211
6.2.1. Misalignments between CEFR levels achieved and required as perceived by the same students	214
6.2.1.1. Overall misalignments of the students' CEFR levels	214
6.2.1.2. Misalignments of the students' CEFR levels across MHCCs.....	215
6.2.1.3. Misalignments of the students' CEFR levels across PYP levels.....	217
6.2.2. Misalignments in CEFR levels between PYP tutors and MHCC students	220
6.2.2.1. Overall misalignments in CEFR levels between PYP tutors' and MHCC students.....	220
6.2.2.2. Misalignments in CEFR between PYP tutors and MHCC students across MHCCs.....	222
6.2.2.3. Misalignment in CEFR levels between PYP tutors and MHCCs students across PYP levels.....	223
6.2.3. Misalignments in CEFR levels between raters' ratings and MHCC students.....	226
6.2.3.1. Overall misalignments in CEFR levels between raters' ratings and MHCC students.....	226
6.2.3.2. Misalignments in CEFR levels between raters and MHCC students across MHCCs.....	227
6.2.3.3. Misalignments in CEFR levels between raters' ratings and MHCCs students across PYP levels.....	228
6.2.4. Misalignment in CEFR levels between PYP students' self-assessment and MHCC teaching staff.....	228
6.2.5. Misalignment in CEFR levels between PYP tutors and MHCC teaching staff	229
6.2.6. Misalignment in CEFR levels between raters' ratings and MHCC teaching staff	231
6.3. Identifying the gaps between the writing skills perceived achieved (Phase I) and the ones required (Phase II).....	232
6.3.1. Writing at the PYP curriculum: where is the gap?	232
6.4. Causes of the gaps: a vicious circle	238
6.4.1. Lack of proper needs analysis.....	239
6.4.2. Lack of proper admission policies.....	240
6.4.3. Insufficiently rigorous assessment.....	241

6.4.3.1. The assessment of writing in secondary education: “ <i>a regurgitation of texts</i> ”	242
6.4.3.2. Writing assessment in the PYP programme: “ <i>One big blanket for all</i> ”	244
6.4.3.3. The assessment of writing in college: <i>content vs. linguistic elements</i>	249
6.4.4. Students’ proficiency in writing.....	252
6.5. Phase III Discussion.....	255
6.5.1. Does the PYP curriculum adequately prepare students to meet the writing requirements of the MHCCs?	255
6.5.1.1. General overview.....	255
6.5.1.2. The identified gaps.....	257
6.6. Implications.....	266
6.6.1. The CEFR-based-context-related standards.....	266
6.6.2. Implications for the PYP curriculum	271
6.6.2.1. The PYP curriculum: bridging the gaps between present and target situations	271
6.6.2.2. The PYP: who is it for?	273
6.6.3. Implications for the MHCCs.....	274
6.6.3.1. The need for additional writing support.....	274
6.6.4. The need for rigorous policies.....	277
6.6.4.1. Admission policies.....	277
6.6.4.2. Assessment policy	278
6.6.4.3. Institutional codes of ethics and misconduct policies.....	280
6.6.5. ‘Handle the students with kid gloves’ a concluding remark	280
Chapter 7 Conclusion	282
7.1. Summary of the thesis	282
7.2. Summary of results.....	283
7.3. The study’s contribution.....	286
7.3.1. Contribution to the Saudi context	286
7.3.2. Contribution to the field (the CEFR).....	292
7.3.3. Practical and methodological contributions.....	294
7.4. Limitations and recommendations of further studies.....	298
7.5. Concluding remarks	300
References	302
Appendices	333

Appendix A.....	333
Appendix B.....	376
Appendix C.....	381
Appendix D.....	387

List of tables

Table 1. 1 Comparison of writing scores across PYP levels.....	8
Table 3. 1 Number of students who undertook self-assessment and were assessed by their tutors in Phase I using CEFR-based questionnaires.....	84
Table 3.2 Raters' Biographical Data.....	85
Table 3. 3 Number of students and teaching staff who participated in Phase II Questionnaires.....	85
Table 3. 4 Number of students and teaching staff who participated in Phase II FGs/interviews.....	86
Table 3. 5. The distributions of the CEFR in the self-assessment grid and the numbering system/coding used.....	90
Table 3. 6 Inter-rater reliability for the 7 raters' scales.....	112
Table 3. 7 Generalized Fleiss Kappa for inter-rater agreement between 7 raters.	112
Table 3. 8 Percentage agreement between the 7 raters.....	113
Table 3.9 The Seven Raters' Intraclass Correlation ICC.....	113
Table 4. 1 Descriptive analysis of PYP students' self-assessment across the PYP levels.....	124
Table 4. 2 Descriptive analysis of PYP tutors' assessment across PYP levels....	125
Table 4. 3 Descriptive analysis of the raters' assessment of sample students' texts across the PYP levels.....	125
Table 4. 4 One-Way Analysis of variance of students' self-assessment of CEFR levels across PYP levels.....	127
Table 4. 5 Robust test of equality of mean of students' self-assessment of their CEFR levels across the three PYP levels.....	128

Table 4. 6 One-Way Analysis of variance of tutors' assessment across PYP levels	129
Table 4. 7 Robust test of equality of mean of tutors' assessment across PYP levels	130
Table 4. 8 One-Way ANOVA of raters' assessment across PYP levels	131
Table 4. 9 Paired differences between self-and tutors' assessment in each PYP level.....	133
Table 4. 10 Correlation and agreement between ratings of self- and tutors' assessment.....	135
Table 4. 11 One-Way ANOVA between self-, tutors and raters	137
Table 4. 12 One-Way ANOVA for ratings between self-, tutors and raters across PYP levels.....	137
Table 4. 13 Overall correlation analysis between self, tutors' and raters' assessment.....	138
Table 4. 14 Correlation analysis between self, tutors and raters assessment across PYP levels.....	139
Table 4. 15 CEFR level perceived achieved by the end of the PYP based on FG data	141
Table 4. 16 Phase I themes: Writing at the PYP.....	142
Table 5. 1 Summary of the CEFR levels students and teaching staff perceived as being minimally required at MHCCs.....	176
Table 5. 2 Differences between students' and staff's perceptions of CEFR levels required at MHCCs.....	177
Table 5. 3 Descriptive analysis of the CEFR levels perceived as required by students across MHCCs	178
Table 5. 4 One-way ANOVA of CEFR levels perceived as required by MHCC students across MHCCs.....	179
Table 5. 5 Robust test of mean of CEFR levels perceived as required by MHCC students across MHCCs.....	180
Table 5. 6 Descriptive analysis of the CEFR levels perceived as required by MHCC students across the PYP levels.....	181

Table 5. 7 One-Way ANOVA of CEFR levels perceived as required by MHCC students across PYP levels.....	181
Table 5. 8 Robust test of equality of mean of CEFR levels perceived as required by MHCC students across PYP levels.....	182
Table 5. 9 Students PYP Levels * MHCCs Crosstabulation.....	183
Table 5. 10 Summary of the CEFR levels selected by participants during the FGs/interviews with justifications of choices.....	186
Table 5. 11 Phase II themes: writing in first-year MHCCs.....	189
Table 6. 1 Overall misalignments in students' CEFR levels as perceived by student in Phases I and II.....	214
Table 6. 2 Misalignments in students' CEFR levels across MHCCs as perceived by the students.....	216
Table 6. 3 One-Way Analysis of Variance ANOVA of misalignments across MHCCs as perceived by students.....	216
Table 6. 4 Misalignments in students' CEFR levels across PYP levels as perceived by students.....	217
Table 6. 5 One-Way Analysis of Variance ANOVA of misalignments in CEFR levels across PYP levels as perceived by students.....	218
Table 6. 6 Robust test of equality of mean of misalignments in CEFR levels across PYP levels as perceived by students.....	219
Table 6. 7 Overall misalignments in CEFR levels between PYP tutors and MHCC students.....	221
Table 6. 8 Misalignments in CEFR levels between PYP tutors and MHCC students across MHCCs.....	222
Table 6. 9 Misalignments in CEFR levels between PYP tutors and MHCC students across PYP levels.....	223
Table 6. 10 One-Way Analysis of Variance ANOVA of misalignments between PYP tutors and MHCC students across PYP levels.....	224
Table 6. 11 Robust test of equality of mean of misalignments between PYP tutors and MHCC students across PYP levels.....	225
Table 6. 12 Overall misalignments in CEFR levels between raters' ratings and MHCC students.....	226

Table 6. 13 Misalignments in CEFR levels between raters' rating and MHCC students across the MHCCs.....	227
Table 6. 14 Misalignments in CEFR levels between raters' ratings and MHCC students across PYP levels.....	228
Table 6. 15 Overall misalignments in CEFR levels between PYP students and MHCCs teaching staff.....	229
Table 6. 16 Overall misalignments in CEFR levels between PYP tutors and MHCC teaching staff.....	230
Table 6. 17 Overall misalignments in CEFR levels between raters's ratings and MHCC teaching staff.....	231
Table 6. 18 Summary of Gap Analysis.....	261
Table 6. 19 CEFR-based-context-related standards.....	268

List of figures

Figure 1. 1 PYP Curriculum Framework for PYP MT.....	6
Figure 1. 2 PYP Students' distribution of scores in the final writing exam.....	9
Figure 3. 1 Flowchart illustrating the convergent parallel design of the study across the three Phases.....	82
Figure 3. 2 the study participants in Phase I and II.....	87
Figure 3. 3 A Screenshot (as an example) of how the CEFR scales were used by participants.....	92
Figure 4. 1 Visualisation of the quantitative analyses conducted in Phase I.....	123
Figure 5. 1 Visualisation of the quantitative analyses conducted in Phase II....	175
Figure 5. 2 Percentage of students participating in Phase II from each college across the students' PYP levels.....	184

Figure 6. 1 Visualisation of the quantitative analyses of misalignments between CEFR levels achieved in PYP and those perceived as required in first year MHCCs.	213
Figure 6. 2 Distribution of writing skills as identified by the number of students for each skill	233
Figure 6. 3 The ranking of how well students felt prepared by the PYP curriculum.	234
Figure 6. 4 The vicious circle identified from the analysis of the qualitative data affecting students' writing proficiency and contributing to the gap	254

List of abbreviation

CAMS	College of Applied Medical Sciences
CD	College of Dentistry
CEFR	The Common European Framework of References for Languages
CM	College of Medicine
CN	College of Nursing
COO	Coordinators
CPH	College of Pharmacy
ELSD	English language skills department
EMI	English as the main Medium of Instructions
FG	Focus group
HT	Humanities Track at the PYP
L1	First language
L2	Second language
MHCCs	Medical and Healthcare Colleges
MT	Medical Track at the PYP
NA	Needs analysis/assessment
PH1	Phase I
PH2	Phase II
PYP	Preparatory Year Programme
SF	Staff
SS	Students
ST	Scientific Track at the PYP
TS	Tutors

Acknowledgments

First of all, my sincere gratitude is to Allah the Almighty for bestowing the strength I needed and guiding me throughout all the difficulties.

This thesis would not have been started, progressed and reached completion without the help, support and guidance I have received from many people and individuals who have written themselves into this work in some way or another. Without them, this thesis would not have come to be.

I am deeply grateful for my supervisors: Prof Claudia Harsch and Dr Neil Murray for their guidance and support, their valuable comments and their constructive feedback throughout the different stages of my thesis. Their patience, critical and valuable feedback and continuous support throughout these years made this thesis possible. I would also like to extend my thanks to Prof Ema Ushioda for her continuous and unlimited support throughout my PhD and for her valuable comments on earlier versions of my thesis. I am also grateful to Dr Luke Harding and Dr Gerard Sharpling for their valuable comments and feedback during the viva, which helped improve the last version of this thesis.

I also wish to thank all the study participants including the students, tutors, coordinators, raters and the academic staff. Along with the deans and vice-deans of the PYP and the different medical colleges, these kind individuals provided access to the necessary data.

I am grateful for the generous sponsorship provided by King Saud University without which this thesis would not be possible.

I am grateful and thankful to Asma Al-Shehri, not only for her support as a friend, but also for her dedication to the data collection process in Phase I and her vigilance with respect to ethical requirements; Ida Salleh, for her true and trustworthy friendship, for having a great heart and ears, for listening patiently to all my research ups and downs, for her hilarious encouragements, for making sure I am having some breaks away from work and for the time spent reading my drafts; and Olfat Mirza, for her friendship, the camaraderie and solidarity, for being there through thick and thin during our last year of this journey.

My deepest thanks are extended to my friends and colleagues in CAL especially Ana K, Hah Sixian and Saeede Haghi and for their feedback and stimulating discussions.

Most importantly, I must thank my family: my father Abdulaleem and my mother Mariyam for always being there for us, for your prayers, tireless support and love towards both me and my children and for leaving all your commitments aside and being here just to help and support. Without you, Mum and Dad, this thesis would never have begun, let alone reached completion. I am also grateful to my dearest brothers (AG, Ahmed and Mohammed) and my sisters (Doody, Foush and

Amoly) who offered their support at every step of the way. My beloved husband Abdulsattar, for your patience and support, for always trying your best to help; without your support, the completion of this work would not be possible. To my heart and soul, my beloved children, Jood, Abdullah and Leenah who shared this long journey and whose presence always reinforced my commitment and dedication to the work.

Declaration

I declare that this thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy in English Language Teaching and Applied Linguistics. The thesis is entirely my own work. No portion of the work in this thesis has been submitted in support of an application for another degree or qualification at this or any other university in any previous application for any degree. I also declare that a small part of the results related to this thesis has been published as follows:

Abdulhaleem, E and Harsch, C (2018). Using the CEFR Scales to Assess Students' Proficiency Levels in a Saudi-Arabian Higher Education Context. In Brandt / Buschmann-Göbels / Harsch (eds). *Der Gemeinsame Europäische Referenzrahmen für Sprachen und seine Adaption im Hochschulkontext*. Fremdsprachen in Lehre und Forschung Bd. 51. Bochum: AKS Verlag, 167-178.

Abstract

The Preparatory Year Programme (PYP) in Saudi Arabia offers a one-year intensive English course to matriculating tertiary-level students, with the aim of meeting the high language proficiency levels required at tertiary institutions, including medical and healthcare colleges (MHCCs) where English is the main medium of instruction (EMI). However, students continue to struggle to meet the high requirements in MHCCs, particularly with respect to writing.

The purpose of the current longitudinal exploratory study was to explore students' writing proficiency at the end of the PYP (Phase I) and the requirements of first-year MHCC students (Phase II), and to identify misalignments between these two levels of proficiency, and to explore the gaps identified (if any) (Phase III).

A mixed-methods gap analysis approach to needs analysis was employed. In the quantitative component, ten CEFR scales were used to identify PYP students' CEFR levels at the end of the PYP, and these revealed writing proficiency levels between A2+ and B2. One year later, the same scales were utilized, and B2 was identified as the minimum level required in the field of writing for first-year students enrolled in the MHCCs. Concurrently, qualitative data were collected to better understand and explore students' writing proficiency and how participants perceive writing in terms of what students *can do or are required to do* with language. The qualitative analyses provided insights into the different causes of gaps that contributed to issues with learners' writing proficiency.

Comparing the two phases, significant misalignments were found, particularly in relation to the *Vocabulary range and control* and *Reports and essays* scales. Other issues included the lack of specificity in the PYP writing course, and limited practice and exposure to medical terminology and relevant writing genres.

These findings can assist PYP stakeholders to address the gaps between the two curricula and better prepare students for their first year of study at the MHCCs. The study also demonstrated the usability and reliability of the CEFR scales for needs analysis studies. Furthermore, it confirmed that the CEFR can be used with minimal training to identify students' proficiency, provided that it was done via a controlled utilization of the CEFR descriptors.

Chapter 1

Introduction

1.1. Introduction

This thesis explores the writing proficiency of pre-university students in order to identify misalignments between students' actual writing proficiency and the proficiency required for study in Medical and Health Care Colleges (MHCCs). Its purpose was to identify misalignments (if any) in students' writing proficiency in order to identify gaps in the writing course delivered as part of the preparatory year programme (PYP) at one of the most reputable universities in Saudi Arabia. By using a gap analysis approach to needs analysis (NA), data were collected in two phases. Students who participated in Phase I at the end of the PYP were revisited for Phase II data collection one year later, after their enrolment into their first year in MHCCs.

In this introductory chapter, I summarise the background and context of the current study and the problems that motivated it. The aim, objectives and research questions follow, and the significance of the study and my role in it as a researcher are set out. This is followed by an overview of the thesis structure.

1.2. Background

English has become the principal medium of instruction (EMI) in many medical, engineering and technical colleges within Saudi universities (Al-Kahtany *et al.*, 2015; Al-Jarf, 2008b; Smith & Abouammoh, 2013); this means that the teaching, learning and assessment of these subjects are all undertaken in English. The application of EMI is not fully supported by every scholar, however: a number who oppose this new direction, for example, contend that language deficiency may be detrimental to the positive outcome of EMI. Other scholars, though, remain in support of the decision, for they recognise the importance of keeping up to date with scientific research and developments around the world (Green *et*

al., 2012). In their view, EMI is valid because English is regarded as the international language, whose technical terminology is used among scientists across the globe (Al-Jarf, 2008a). Supporters of EMI also consider it necessary to demonstrate a close association with western education systems, as well as provide students with up-to-date knowledge (Abu-Rizaizah, 2010; Al-Hazimi *et al.*, 2004; Alhawsawi, 2013). According to a large number of commentators, teaching students through EMI increases access to national and international academic and professional opportunities (Al-Seghayer, 2005; Alhawsawi, 2013). For instance, students in the field of medicine are required to join fellowship programmes in the United States, United Kingdom or Canada as part of their professional development, in which English is clearly the medium of instruction.

Opponents of EMI are sceptical of this trend at Saudi universities in the absence of assessing, in detail, its potential impact. One important reason for this is that not all teaching staff possess a high proficiency in English. In the medical field, however, staff are often either native English speakers or Arabic speakers; the majority thus have excellent academic qualifications, and a significant proportion of them have been trained in western countries (Hamdy *et al.*, 2010). Other opponents emphasise the importance of Arabic as the main language of instruction for all subjects in Saudi universities.

Though many students have six years of general English exposure prior to their university education (Alhawsawi, 2013; Khan, 2011), most Saudi students' English language proficiency is insufficient to enable them to cope with university EMI studies (Abdulghani *et al.*, 2014; Alhawsawi, 2013; Al-Hazmi, 2003, Khan, 2011; Sheshsha, 1982; Zaid, 1993). This poses an obstacle to students' enrolment in these colleges (Alhawsawi, 2013), and subsequently present difficulties, particularly in the first year of university (Sabbour, Dewedar & Kandil, 2010). Thus, many universities offer an extra, introductory year – the PYP – which aims to close the language proficiency gap between high schools and universities, including medical colleges (Al-Shehri *et al.*, 2013; Yushau & Omar, 2007), to enable students to satisfy the colleges' academic requirements (Alnassar & Dow, 2013, p.51).

1.3. Context of the study

This study took place in one of the oldest and most reputable universities in the Kingdom of Saudi Arabia, which has received a high international ranking, and is considered “the highest of any Arab university” (Smith & Abouammoh, 2013, p.3). It offers a wide range of subjects and specialities including humanities, science and medicine, mostly taught through English (*ibid.*, p.3). The first medical schools in the Kingdom were established in this university in 1967 (Al-Shehri *et al.*, 2013). Currently, the university comprises five MHCCs: the *College of Applied Medical Sciences (CAMS)*, the *College of Dentistry (CD)*, the *College of Medicine (CM)*, the *College of Nursing (CN)* and the *College of Pharmacy (CPH)*. Both males and females have equal opportunity to join any of these colleges, but all prospective students must meet their college’s specific requirements. In line with Saudi policy and regulations, “a general policy of gender segregation” is strictly followed (Smith & Abouammoh, 2013, p.2). The university, like almost all others in Saudi Arabia, has two main campuses, one for female students and the other for males; the curriculum and assessment are identical (Smith & Abouammoh, 2013, p.2). Due to this segregation policy, it is impossible for female researchers to conduct studies on male campuses (or vice versa), so as a female, I was limited to studying female students.

The PYP in this university is mandatory for all prospective students, and is designed “to improve the [different] knowledge and skills of high school graduates before they join their desired majors at the universities” (Al-Murabit, 2012, p.8), including various academic and lifelong skills. Upon completion of the PYP, students enrol at one of the colleges in the University. However, admission to colleges depends on factors including the “track” these students are streamed into during the PYP. Three such tracks are available, based on students’ high school scores and specialty preference: *Science Track (ST)*, *Medical Track (MT)* and *Humanities Track (HT)*. Each track reflects the colleges students can join after completing the PYP. For example, students on the MT may choose from the MHCCs, whereas students on the ST join one of the scientific or engineering colleges, and those on the HT join one of the educational or art colleges. Students’ admission to these colleges also depends on their cumulative scores upon

completion of the PYP. Entry into the more competitive colleges (e.g. CM) requires higher scores, making PYP a high-stakes programme which determines students' educational futures.

English (worth eight credits) (Al Makoshi, 2014) is one of the main components of the PYP programme, where intensive courses are provided for different baseline proficiency levels and tracks. Therefore, the English programme at the PYP should play an important part in addressing students' language weaknesses, enabling them not only to cope with, but also to excel during their academic life at university.

The English Language Skills Department (ELSD) at PYP has several units operating together to provide a high-quality intensive English programme; those most relevant to this study are the Curriculum, Assessment, Continuous Assessment and Professional Development units.

The PYP curriculum has evolved since its inception. A few years ago, the Curriculum unit introduced the *Common European Framework of Reference for Languages: Learning, Teaching, Assessment* (CEFR) (Council of Europe [CoE], 2001) as a reference tool for designing the curriculum, selecting textbooks, writing students' learning outcomes and defining proficiency levels. The use of the CEFR at the PYP is based on the Curriculum team's intuitions about what levels, outcomes or textbooks are most suitable for the students. The English course at the PYP runs for the whole year, divided into two semesters. A general English (GE) course is offered in the first semester and English for academic purposes (EAP) and/or specific purposes (ESP) in the second (e.g. English for Medical purposes [EMP] in the MT). Each track comprises three levels representing baseline English proficiency: *elementary*, *intermediate* and *advanced*. The textbooks used and the number of class hours per week vary between proficiency levels. For example, students in the *elementary* level use more basic textbooks and complete more hours (20 hours weekly) compared to *intermediate* or *advanced* levels (about 8 hours weekly) (Al Makoshi, 2014). However, the PYP curriculum is not based on a detailed data-driven analysis of students' needs.

Prior to commencing the PYP, all matriculating students are required to take a placement test to determine their baseline proficiency levels and to assign them to groups according to track and level (Al Makoshi, 2014). The Oxford Placement Test (paper-and-pen version) is used, which examines students' listening and reading skills, grammar and vocabulary. However, this test does not assess written or oral skills, so proficiency in those areas is not identified prior to the course. The placement test is scored with a maximum of 100. Students scoring 0–45 are placed in the *elementary* level, 46–85 in the *intermediate* level, and >85 in the *advanced* level.¹

In the context of the study PYP, students' progress in language proficiency from PYP entry is mapped via a comprehensive Curriculum Framework linked to the CEFR (CoE, 2001) as a reference. Even though the linkage is currently intuitive rather than empirically based, the framework provides a roadmap that guides staff and students in the PYP regarding the proficiency levels that need to be achieved throughout the academic year. Figure 1.1 shows the curriculum framework for MT students in the PYP, indicating the alignment between the ELSD attainment levels (from elementary to advanced) and the CEFR levels (from A1 to C1), and the expected progression of students from their entry levels during the year across the rows, so expected curriculum outcomes depend on baseline proficiency level and progress over time.

¹ Placement test cut-scores were adapted from Oxford University Press (2001). *Quick Placement Test: Paper and Pen Test, User Manual*. Then the final cut-off scores were decided by the assessment team at the PYP.

Track	Baseline Levels		Quarter 1	Quarter 2		Quarter 3	Quarter 4
Medical Track (MT)	Elementary	Semester 1	Elementary A1-A2	Pre-Intermediate A2	Semester 2	Pre-Intermediate Plus A2-B1	Intermediate
	Intermediate		Pre-Intermediate A2	Pre-Intermediate Plus A2-B1		Intermediate	Intermediate Plus B1-B2
	Advanced		Intermediate B1	Intermediate Plus B1-B2		Upper Intermediate B2	Advanced C1

Figure 1. 1 PYP Curriculum Framework for PYP MT

The Assessment Unit of the ELSD designs exams and arranges test logistics. Although the expected curriculum outcomes depend on baseline proficiency, the examinations at the end of each semester are set so that all students, regardless of baseline level, take the same exam at the end of each semester.

The Continuous Assessment Unit organises two projects per semester, one focuses on oral presentation skills and the other on writing. In the latter, students write on one topic, progressing from choosing the topic to producing a final written product, over six weeks. The Professional Development Unit is the one responsible for teachers' training at the PYP.

After the PYP, students join the different colleges based on their cumulative grade point average (GPA): those with higher GPAs can join the most competitive colleges, such as the *CM* and *CD*, so the exam should ideally enable differentiation between skill levels. In such an exam, ideally "the decision makers need the test takers to be 'spread out' over the range of test scores" (Fulcher, 2013, p. 31). However, as discussed below, there was no such variations in students' scores in the current context.

1.4. Statement of the problem

While PYP has succeeded in improving students' academic skills and English language proficiency (compared to the situation before the implementation of

the PYP) (Al-Omar, 2014), many colleges, including medical colleges, have remained dissatisfied with the results (Sabbour *et al.*, 2010).

Most of the research literature related to medical students and their language levels or needs indicates the importance of high levels of English proficiency (Alfehaid, 2014; Alhawsawi, 2013; Ghobain, 2014; Shukri, 2008). For example, Ghobain mentions that students and professionals in the medical fields “must have an advanced level of English proficiency” (2014, p.154). However, what constitutes ‘high’ or ‘low’ levels of English is generally ill-defined.

Writing in English is a skill that many Saudi students find exceptionally challenging (McMullen, 2009; Shukri, 2014). This is true even in the case of highly proficient students (Shukri, 2014). In schools, students are more familiar with the rote memorisation of written essays “to be retrieved only to pass examinations” (Alnassar & Dow, 2013, p.57). This behaviour may also affect the students’ writing later on. Several informal studies have been conducted at the PYP that had identified writing as a real issue of concern. For example, an analysis of the International English Language Testing system (IELTS) scores of PYP advanced students from 2009 to 2012 revealed that writing was scored the lowest of all language skills. Taking IELTS is optional and only permitted for advanced-level students; and these students are not exempted from the PYP English course even if they score in the upper IELTS bands. An informal survey was also conducted to identify students’ preference for optional elective courses, with results showing that academic writing was participants’ first choice (82%). The above observations clearly demonstrate that writing is an area of some concern.

Although there has been much research into how Saudi students’ English skills can be improved at university level, there has been little focus on improving students’ writing, specifically to a level that would enable them to cope with the academic tasks required at tertiary level, such as writing assignments and responding to various assessment tasks in English (Alnassar & Dow, 2013). However, before we can be concerned with ways in which to improve students’ writing, it is firstly crucial to understand students’ writing proficiency and what

students *can do* and the skills they need to be able to do (are *required to do*) in their writing to cope with the specific writing requirements at university.

In order to have a general idea of students' proficiency in writing, I analysed the students' writing scores in the final exam (the scores of the same students' cohort who participated in this study). In this final exam, the students, regardless of their level, took the same standardised test. They were asked to write 120 words in 60 minutes on a general topic about their daily routine at the university. (No official permission was gained to post the actual task prompt, so it is not included in the appendix.). However, a ceiling effect has been observed from those students' final written exam (see table 1.1. and figure 1.2below).

Table1. 1 Comparison of writing scores across PYP levels

	<i>Median</i>	<i>IQR</i>	<i>N with max score</i>	<i>% with max score</i>
Elementary	9.6	9.2, 10.0	29	35.4%
Intermediate	10.0	9.6, 10.0	207	65.3%
Advanced	10.0	10.0, 10.0	194	87.8%

IQR Interquartile range

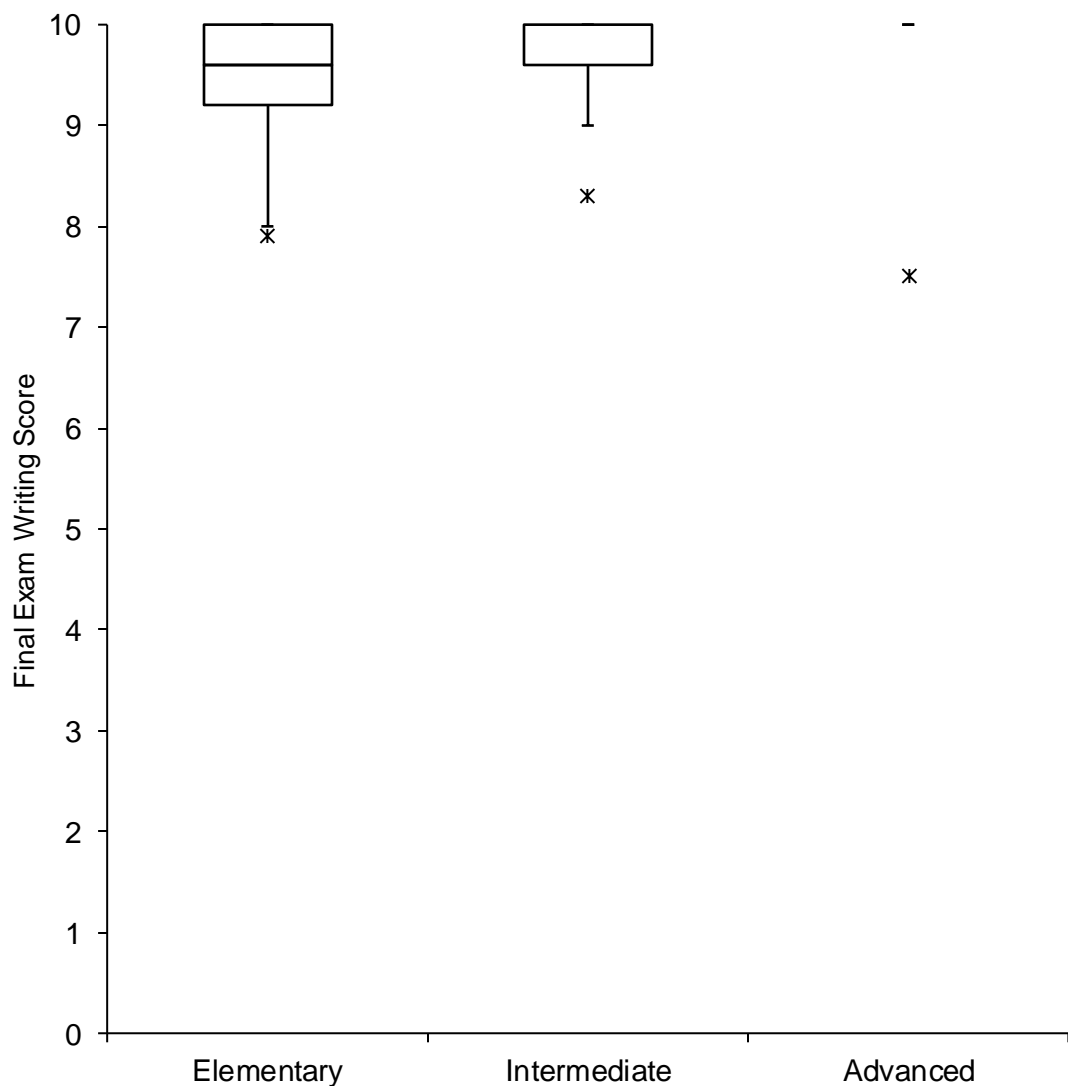


Figure 1.2 PYP Students' distribution of scores in the final writing exam

A box-and-whisker presentation of PYP students' scores distribution on the final writing exam. The boxes represent the 2nd and 3rd quartiles separated by median while outliers are indicated by crossed points. Here, the median for the elementary level is around 9.5, while the median for intermediate and advanced levels are 10. In the case of the advanced level, no box or whisker is shown as the majority of the students scored full marks, showing the ceiling effect.

This preliminary observations of the exit scores from $n=588$ students on the PYP suggest that on this exit exam, marks were not spread out across the range of scores, with 73% of students scoring full marks, and the median (IQR) scores out of 10 being 9.6 (9.2, 10), 10 (9.6, 10) and 10 (10, 10) for students starting in the elementary, intermediate and advanced levels, respectively (see table 1.1). While high scores on the exit exam could indicate students' progress due to the

teaching during the PYP, or that the exam was insufficiently challenging, it did not differentiate well between the students, nor describe the actual proficiency achieved against an internationally-recognised proficiency framework such as the CEFR.

While existing research has generated some insights into factors affecting students' writing, little (if any) is empirically known about the level of students' writing proficiency in relation to an international proficiency framework (e.g. the CEFR). The current study, therefore, seeks to explore the gap between what is being offered at the PYP and what is required in the MHCCs, focussing in particular on students' writing skills, using the CEFR (CoE, 2001) as the main tool for data collection and the gap analysis approach to NA.

1.5. Research aim and objectives

The purpose of this study is to identify and analyse any misalignment between what *is being offered* and what *needs to be* offered by the PYP, and to discuss what amendments need to be made to the PYP writing course in order to meet the needs of students intending to study in the MHCCs.

To achieve this aim, this study has the following objectives:

- 1- To identify the CEFR levels students have reached at the conclusion of the PYP, and the writing skills they have acquired, based on the perceptions of the students' themselves, their tutors and the course coordinators as well as the rating of a sample of students' written texts by raters.
- 2- To identify the CEFR levels that students would need to have reached and writing skills that students need to cope with their first-year college requirements, based on the perceptions of students and academic staff at the end of the first academic year in these MHCCs.
- 3- To specify and understand the gap(s) (if any) in the PYP writing course, by identifying any misalignment between what is being offered by the PYP and what is required by the MHCCs, and the reasons for any misalignment.

1.6. Research questions

The study aims to answer the following research questions:

Overarching research question

Does the PYP curriculum adequately prepare students to meet the writing requirements of the university Medical and Healthcare Colleges (MHCCs)?

Secondary research questions

Phase I:

1. What are the students' CEFR writing levels at the end of the PYP, as assessed by the students, their tutors and trained English language raters?
2. What writing skills do students, tutors and coordinators perceive students graduating from the PYP to have developed?

Phase II:

3. What are the CEFR levels required of first-year medical students to cope effectively with the colleges' writing requirements, as perceived by MHCC students and academic staff?
4. What writing skills are required during the first year in college, as perceived by MHCC students and academic staff?

Phase III:

5. To what extent is there a misalignment between the students' CEFR writing levels achieved by the conclusion of the PYP and those perceived as required during the first year of MHCCs?
6. To what extent are there misalignments between the writing skills perceived as having been achieved in Phase I and those perceived as required in Phase II?
7. To what cause(s) can gaps resulting from any misalignment be attributed?

1.7. Rationale and motivation for the study

The inspiration for this study stems from my observations during my time working at the PYP. Some negative comments were received about the course, particularly regarding students' writing abilities following the completion of the PYP, and particularly from those using English for academic instruction (Abomrifa, 2013). Academic staff expected students completing the PYP to have a high English proficiency level, without clear specifications of this 'high level'. In addition, I observed that many students completing the PYP still found it difficult to cope with the high expectations of the colleges in which they had enrolled (*ibid.*). I personally encountered several cases where students, despite their passion for pursuing an education in medicine, had to choose to withdraw because their language skills were insufficient to meet the demands of English use.

Furthermore, I am curious why students had such difficulty with their English written skills. Even "top-notch" students who were expected to achieve good results and were offered places in some of the top universities in the region with the most stringent acceptance criteria had difficulties with their written language.

From what I have observed, and as an academic lecturer working in higher education, realising the importance of writing as a lifelong skill, my ultimate goal in carrying out this research is to reach a better understanding of students' proficiency in writing (both achieved and required) and to explore the potential factors leading to gaps in the PYP curriculum from different perspectives, including the students themselves. This would assist PYP policymakers to improve their curriculum to help students better achieve their needs.

This concern (identifying the gaps) influenced the design of the study. Furthermore, based on the results obtained, the pedagogical and practical implications for what can be done to improve students' writing proficiency and modify the writing course in PYP curriculum, along with methodological implications of the study design, are discussed.

1.8. The significance of the study

This study contains an in-depth exploration of students' writing proficiency exhibited and needed both at the PYP and the MHCCs, which enables identification of any misalignment between the two and helps to find the gap between what *is* currently being offered and what *needs* to be offered. In relation to this, a few significant points may be highlighted:

This university is considered the highest ranked in the region and in the Middle East, and students joining this university are those who are claimed to have the highest scores in the region (Smith & Abouammoh, 2013). The purpose of this study is to provide in-depth exploration of students' writing proficiency (achieved and required) in order to identify any misalignments that lead to gaps in the PYP curriculum. It does this by looking at a specific cohort, MT students, who are claimed to be the *crème de la crème* compared to those in other tracks (ST and HT) at other universities.

Incorporating the CEFR into the study not only helps specify which writing levels and skills these students have or require, but also indicates the proficiency level(s) that first-year students in the Saudi medical field need to reach to cope with the University's academic demands. This would fill a gap in the literature, where no specific English (writing) criteria have been identified as required for admission to those MHCCs.

The quantitative data in this study introduce a new procedure (a specific set of instructions about using the CEFR grid), which I designed and called *Controlled Utilization of CEFR Descriptors*, and which helps participants assess proficiency levels using the CEFR scales. The findings contribute to the relatively sparse research literature on self-assessment and the identification of standards using the CEFR in the Saudi Arabian context. It also contributes to research concerning using the CEFR in self-assessment, where accuracy and reliability have been questioned because of students' uncertainty regarding their own levels (Luoma & Tarnanen, 2003).

The gap analysis approach to NA using the CEFR scales and the in-depth analysis contributes to the 'second generation' (Huhta *et al.*, 2013) NA, which goes beyond

the traditional approaches and which can provide a way forward for stakeholders and policy makers in the writing of standards that students in the MHCCs would need to acquire. Previously, “first generation” NA studies “focus exclusively on functions and notions and on the four skills ... speaking, listening, writing and reading” (*ibid.*, p.14). Whereas in the second generation, according to Huhta *et al.* (2013), NA follows more comprehensive approaches and focuses particularly on tasks.

Finally, the longitudinal element of the study is significant in gaining more in-depth insights into one student cohort’s perceived writing skills at the end of the PYP and, one year later, at the end of their first year at their respective medical colleges. This longitudinal element is, to date, frequently missing from existing studies.

1.9. My role as a researcher

I used a mixed-methods approach including qualitative and quantitative data. Ethical issues were strictly considered in the collection and analysis of the data. In my collection of the data, I played the role of “insider” and “outsider” researcher. The insider researcher is the one “familiar and informed about the setting of the study, while an outsider is someone who is unfamiliar with and uninformed about the setting” (Griffith, 1998 cited in Mercer, 2007, p.3). I was an insider while I was collecting the data from the PYP as I used to work in that context (although I had been away for two years before returning for data collection). However, I was an outsider when I collected the data from the MHCCs. The challenge was to maintain objectivity as much as possible (Labaree, 2002), especially as an insider. One advantage of being an insider was the ease of accessibility to students, tutors and coordinators. However, as argued in Labaree (2002) and Mercer (2007), being an insider is not without its risks. Insider researchers might come to the context with considerable rapport with the subject, which might affect the way they collect and analyse the data (Mercer, 2007). Realizing this issue, however, I was conscious in the way I asked my questions and controlled my own voice, avoiding leading questions and

comments as much as possible (Gubrium & Holstein, 2003). To avoid bias while analyzing the data, I made decisions based on patterns found in the data itself. As an outsider researcher in Phase II, I faced more challenges relating to accessibility than in Phase I.

1.10. Overview of the thesis structure

The thesis is organized as follows:

Chapter 2 reviews the relevant literature and the key definitions and concepts relevant to the current study, showing how the literature informed my study design. Chapter 3 describes the mixed-methods approach with some longitudinal aspects, and provides information about the study sample, the research tools and the procedures followed for data collection and analysis. Chapters 4-6 present the findings in Phases I, II and III, respectively, along with the relevant discussion. Chapter 7 is the concluding chapter, which discusses the implications of the study, recommendations for future research and the study's limitations.

Chapter 2

Literature Review

2.1. Saudi Education

2.1.1. The implementation of the PYP

It is important that students join MHCCs that use EMI with acceptable English proficiency (Maher, 1986; Al-Qhatani, 1999; Al-Shehri *et al.*, 2013) as students in these colleges “must absorb a great deal of information in a set of time from both written and spoken texts” in English (Al Makoshi, 2014) to succeed in these medical subjects:

English is becoming “a superior language [in Saudi universities] being an international language, and the language of science and technology, research, electronic databases and technical terminology...[and] English is more appropriate for teaching medicine, pharmacy, engineering, science, nursing, and computer science (Al-Jarf, 2008b, p.193).

In 1993, Zaid found that the subjects of his study did not attain the level of mastery in English that was expected by the Ministry of Education (MoE), i.e. students were joining universities with a lower level of proficiency than required. This gap is particularly problematic as English is the main language of instruction in many technical, scientific and medical departments in Saudi universities (Javid *et al.*, 2012; Alblowi, 2016; Al-Shumaimeri, 2011). As stated by Almulhim (2001) in his study of the needs of students attending the college of technology administrators, a ‘good’ level of proficiency is needed in all four language skills (at tertiary level).

In Saudi Arabian schools, English has for many years been taught as a mandatory core subject, starting from year seven; at intermediate school until the end of High School (a total of six years) (Al-Seghayer, 2005); and, more recently, from the sixth, fifth and fourth grades of elementary school (Alharbi, 2015). This aims

to enable the public to “make use of desired materials in English” as well as “communicate satisfactorily, according to [their] needs, in both spoken and written forms” (Ministry of Education, 2002).

However, there has been little improvement over time in students’ proficiency upon their graduation from high school (Alblowi, 2016; Al-Hazmi, 2003; Alhawsawi, 2013; Alsaif, 2011).

Therefore, there is a real need to improve students’ English proficiency to enable them to be eligible for, and have sufficient competence to pursue, tertiary level education, and, subsequently, professional careers (Alblowi, 2016, p.17). In response to this lack of proficiency, in 2010 the MoE introduced, and officially implemented in Saudi universities, an additional year of learning, the PYP (Alhawsawi, 2013; Ministry of Higher Education [MoHE], 2010).

The concept of PYP is relatively new in the Saudi Arabian context (Alaqaeli, 2014). According to Alghamdi (2015), it can be considered a “therapeutic framework to bridge the gap between general education outputs and higher education inputs in order to form multi-disciplinary skills contributing to improving the student performance” (p.121).

In addition to improving students’ English proficiency (Al Kathiri, 2014; Alblowi, 2016; Nazim and Hazarika, 2017), the other main goals of PYPs are to improve academic and personal skills, such as communication and social skills, self-learning, critical thinking, leadership and problem-solving. The PYP also includes basic maths and other scientific and business administration courses (Al Kathiri, 2014, p.65). According to Alblowi (2016 p.16), by the end of their year in PYP, students should possess:

1. advanced linguistic competence and skills in English;
2. the ability to effectively communicate in written and spoken English;
3. basic skills in academic English; and
4. be able to pass the minimum requirements in international English tests.

Beyond the PYP, there are no further English language support classes offered while students are at university (Al Makoshi, 2014); therefore, it is crucial that

students attain the necessary linguistic competence during their year undertaking that programme.

2.1.2. PYP appreciation and criticisms

In a study that covered students of both genders at three Saudi Arabian universities across geographical locations, PYP students generally realise the value and importance of the programme (McMullen, 2014).

However, some students and parents have expressed dissatisfaction over the ineffectiveness of the programme in addressing students' English proficiency (Alghamdi, 2015; McMullen, 2014; Alkubaidi, 2017).

One of the main reasons for the stakeholders' dissatisfaction and frustration with PYP programmes is that students still join university colleges with a lower proficiency level than that which is expected or required. According to Alfehaid (2014), "having limited English proficiency might lower the quality of the healthcare and demotivate students to complete their academic studies" (p. 275). Students' English proficiency is rather low in these colleges, and their graduates "had long been unsatisfactory for the Ministry of Health" in Saudi Arabia (Alfehaid, 2014, pp. 275-276). Potential reasons for the PYP's ineffectiveness stem from the learners' own lack of essential academic skills or their weak performance in relation to different subjects (Alghamdi, 2015). Alkubaidi (2017) observed that some students had joined one PYP at A1 level, and therefore lacked the "range of vocabulary" and "the knowledge of how to compose a sentence" (p.191). Both studies noted other factors, including teachers' disregard of learners' needs (*ibid*), with Alghamdi (2015) recommending that PYPs needed to improve their services, especially in the area of academic support, to satisfy students.

Other areas that have been identified as potentially affecting the efficacy of the PYP included the English curricula (Al Kathiri, 2014; Al-Omar, 2014; Shukri, 2014; Nazim and Hazarika, 2017; El Tantawi *et al.*, 2016), admissions criteria (Alblowi, 2016), and teaching and assessment (Al Kathiri, 2014; Alblowi, 2016).

PYP curricula should also be more aligned with students' future needs at the university (Al-Omar, 2014). Shukri (2014), in her study exploring L2 writing challenges and difficulties faced by students in PYPs, particularly at King Abdul-Aziz University, ascertained that most English curricula in PYPs deal only with general English, seeking to enable students to "write about themselves, their family, and a variety of topics pertaining to travel, food, daily life, etc." (p. 129). She also noted that, with the curricula's immediate focus being on the textbook, teachers have "insufficient time to implement creative writing activities" (*ibid*). Nazim and Hazarika (2017) focused on students' language needs in Najran University's PYP, and stressed the importance of incorporating the ESP course in the PYP curriculum for improving students' ESP language skills, including the professional vocabulary required for their specific discipline and professional career. They noted that "the students at the PYP have [the] language proficiency to learn ESP, but still there is a need to raise learners' proficiency level required for various professional disciplines" (Nazim and Hazarika, 2017, p.161). El Tantawi *et al.* (2016), who evaluated first-year dental students' writing after finishing the PYP, summarised the inadequacy of the PYP English curricula by stating that "one course of scientific writing in the preparatory year is not enough to develop adequate English language writing skills in undergraduate dental students" (p.153).

Alblowi (2016) evaluated the Taibah University PYP, and found that students were not meeting the expected objectives of the programme, recommending more stringent requirements for admission to the PYP. He noted, however, that students who had attended the PYP had better knowledge and skills than those who had not.

In terms of teaching and assessment, both Al Kathiri (2014) and Alblowi (2016) recommended revisiting and improving these two areas to increase the effectiveness of the PYP in general.

2.1.3. Saudi MHCCs

For students to be admitted to MHCCs in Saudi universities, they need to meet stringent criteria that are more difficult than at other specialities and colleges. Previously, admission to medical colleges was based on students' High School GPA, where only students with the highest GPA would be enrolled (Al Makoshi, 2014; Al-Shehri *et al.*, 2013). From 2002, admission to these medical colleges also requires high scores in the national Aptitude Test and the national Achievement Test (administered by the National Centre for Assessment in Higher Education). Students' scores in English High School are also considered (Al-Shehri *et al.*, 2013). These criteria are used for admission to the MT at the PYP; students then require a minimum of 4.5/5.0 GPA to be considered for enrolment at the MHCCs (Al Makoshi, 2014). The PYP is currently used as "an additional screening tool for admission to medical courses. Students are admitted to medical colleges based on the average grade obtained during the preparatory year" (Al-Shehri *et al.*, 2013, p.146).

Most studies appear to support the notion that these pre-admission criteria (i.e. Aptitude Test, Achievement Test, High School grades) provide a useful indication of students' future achievement at the colleges.

For example, Albishri, Aly and Alnemary (2012) found a positive correlation between pre-admission criteria and medical students' in-course achievement. Similarly, Al Alwan *et al.* (2013) found a positive correlation between all pre-admission criteria, especially at High School, and the medical students' GPA. Alblowi (2016) investigated the ability of pre-admission criteria to determine the students' achievement at the end of the PYP, rather than at the medical colleges, and found the strongest correlation with the Achievement Test. Al-Omar (2014) recommended considering the PYP GPA score alongside, or in preference to, the Achievement and Aptitude Tests. Al Rukban *et al.* (2010), however, discerned that three pre-admission tests did not reliably predict students' GPA in medical colleges, and therefore recommended that new admission criteria be considered. Despite EMI and the international curricula in place at Saudi medical colleges, there is no clear criterion defining the level of English proficiency required for

admission, other than having a good high school and PYP GPA, including English. This problem was identified by Javid *et al.* (2012), who recommended that Saudi Arabian universities should follow a strict admission policy that employs “an internationally benchmarked admission test” (p.65), and only students who met the requirements should be admitted. Al-Jarf (2008a) recommended changing the admission criteria at the university in general. Meanwhile, Al Makoshi (2014) proposed using international admission criteria such as the IELTS (e.g. an overall IELTS score of 7.5, with no component less than 7.0) to assess students seeking to enrol at the medical colleges.

Considering international criteria, it seems that B2 is the level internationally required, though with no empirical basis, for admission to different institutions in Europe (Deygers, Zeidler, Vilcu, & Carlsen, 2018) and around the world (Papageorgiou *et al.*, 2015). Among these few empirical studies are Harsch, Ushioda and Ladroue (2017) who noted that UK universities require a minimum of B2 level or equivalent for international students to obtain a visa in order to pursue their education at UK universities. Carlsen (2018), in a Norwegian context, investigated the minimum level of proficiency necessary to linguistically meet the demands of higher education. She correlated entrance tests with students’ achievement, and found significant positive results: tertiary-university students below B2 level struggled to achieve successful results, indicating that B2 was the minimum required level for success for entrance to higher education in Norway. Though, to my knowledge, there is no study in Saudi contexts dealing with the CEFR as a criterion for admission, Alblowi (2016), in his study of the validity of the current admission criteria in the PYP suggested the need to:

enforce a placement testing system that ensures the satisfaction of admission criteria and the acceptance of students with the same appropriate academic level [and recommended to] eventually adapt the CEFR to Saudi university requirements, thereby creating the proposed Saudi University Preparatory English Requirements (SUPER) (pp.20-21)

For many, the question now is: why are students still joining MHCCs at a lower level of proficiency than that which is required, despite the PYP?

2.2. Saudi students' English writing difficulties

Students' success in university subjects depends largely on their success in academic writing (Pecorari, 2006). Medical students need writing as this will not only form part of the academic requirements at university but, later, will also be required in their future professions (Chang, 2010). As stated in Hedge (2001), different skills need to be acquired for students to be able to write effectively, including the development of ideas, word choices, structure, and achieving coherence. Clarity is one of the most important aspects in the medical field (Messuri, 2015). All these skills make writing demanding and difficult compared with other language skills (Kroll, 1990; Hyland, 2003). Academic writing enhances critical thinking and is used to evaluate students' knowledge of content (Coffin *et al.*, 2003). Complexity in writing increases when it is in L2 academic contexts (Khuwaileh and Al-Shoumali, 2000; Hedge, 2001):

Writing in ESP [...] context propose more difficulties for the second-language learner because the student has to master the content vocabulary as well as use it in a meaningful context (Shukri, 2008, p.1).

Saudi students are stereotyped as 'poor writers of English' (Shukri, 2014, p.190), while Arab and Saudi students are characterized by their generally low proficiency in English in comparison to other nationalities (Kaur, 2003; AlFadly, 2003; Abbad, 1988), especially in speaking and writing (Rababah, 2002). The literature on the difficulties, challenges and obstacles of writing in Saudi contexts is reviewed below, in four main categories: (1) student-related; (2) assessment-related; and (3) teaching- and learning-related difficulties and (4) the effect of L1 on students' writing.

2.2.1. Student-related difficulties

Many Saudi learners are admitted to universities lacking the basic knowledge and qualifications required for "producing sound and acceptable quality writing texts" (Obeid, 2017, p. 175). The lack of basic writing skills (Shukri, 2008) affects their ability to write clearly in examinations and written assignments (Saunders

and Scialfa, 2003 cited in Shukri, 2008, p.2). Almoallim *et al.* (2010) found that over 50% of medical students had difficulty writing due to deficiencies in basic English skills. Alqahtani (2011) ascribed those difficulties to the students' lack of proficiency and lack of academic preparation.

Many students have problems with sentence structure, subject-verb agreement, using irregular verbs, coherence and cohesion, and the linkage of ideas and paragraphs (Khuwaileh and Al Shoumali, 2000). The same was reported in Ahmed (2010), where students faced problems writing thesis statements, topic sentences, and connecting different ideas, i.e. difficulties at sentence- and paragraph-level (Doushaq, 1986). Javid and Umer (2013) found that grammar, the use of vocabulary, prepositions, writing about ideas, and spelling, to be among the most problematic issues experienced by most Saudi students. Al-Khairi (2013) investigated Saudi undergraduate errors among students whose major was English, and found serious problems with vocabulary and lexical items, the use of irregular verbs, spelling and punctuation marks. He attributed these to students' low proficiency in writing, a lack of interest from teachers in assessing written tasks and giving feedback, and limited opportunities to practise writing:

Students experienced writing difficulties in using appropriate vocabulary, writing in correct spelling, following accurate grammatical rules and establishing cohesion in writing. Many students claim to have "the ideas" but have neither the L2 skills nor the pragmatic understanding to express them. (Shukri 2014, p.193)

Besides basic writing weaknesses, many Saudi students have difficulties using their own words and therefore tend to plagiarise. Plagiarism, or "immature writing" (Bereiter & Scardamalia, 1987), is a common problem among students, and even professionals, in the Middle East, as claimed in Handjani & Habibzadeh (2013). El Tantawi *et al.* (2016) studied the scientific writing skills of first-year students who had just finished a PYP at the University of Dammam. They were asked to write a 150-word assignment that was then analysed for writing problems using Microsoft Word and plagiarism detection software. They discovered that many students had resorted to plagiarism in completing this

short assignment, concluding that students were not yet prepared for scientific writing.

According to Handjani and Habibzadeh (2013), plagiarism in the Middle East is “due to a lack of linguistic expertise”. The other reason is that “in some cultures, it is not considered misconduct”. The students’ attitudes towards learning and being in a teacher-centred learning environment have contributed to their tendency to plagiarise (Bereiter & Scardamalia, 1987). For many Asian students, including Saudi students, memorisation and imitating knowledge is a mark of respect (Bereiter & Scardamalia, 1987). However, such behaviour is regarded as plagiarism in the academic tradition. The students’ desire to produce well-written assignments, to score high grades, as well as their lack of confidence and limited time allocated to a written task, are other reasons for the students’ tendency towards plagiarism (Alhojailan, 2015). Institutional leniency in taking action against plagiarism, and its ethical procedures, allows students to breach plagiarism rules (Muthanna, 2016). Muthanna analysed four policy documents and conducted interviews to examine how plagiarism is viewed. He found that the “absence of a research code, research misconduct policy, and institutional policies in the country... [has] led to the presence of several acts of research dishonesty” (p.280).

Students’ lack of awareness of their own weaknesses in writing is another factor affecting their writing. Mohammad and Hazarika (2016) studied students’ written texts in a PYP in Najran University and administered a survey (questionnaire). Results suggested that students were unaware of committing errors in relation to capitalization, punctuation, grammar and spelling, since what they indicated in the questionnaire differed from rating their texts.

Memorization techniques in writing are also prevalent in Saudi Arabia; this is explained by Shukri (2014) as part of the stereotypical feature of Arab students as ‘knowledge teller’ rather than a transferrer of knowledge (Ballard and Clanchy, 1991; Cumming 1989, cited in Shukri 2014). Most Saudi students are used to rote memorisation when writing (Alhaisoni, Al-Zuoud and Gaudel, 2015; Alhawsawi, 2013), and; therefore, tend to write only about what they have memorised. For example, the teachers Alkubaidi (2017) interviewed commented

that students usually ask: “what paragraph do you want us to remember so I can write it in the exam?” (p.195). This same issue was encountered by Mohammad and Hazarika (2016), who suggested giving students writing topics that differed from those in their curriculum to avoid memorization.

Alhojailan (2015) assessed students’ perceptions of academic writing and found that they misunderstood the role of academic writing, as most Saudi students were taught writing according to basic traditional methods (Abalhassan, 2002). Students usually lack motivation and do not understand the importance of academic writing (Alhojailan, 2015; Al-jarf, 2001; Aljamhoo, 1997). Alkubaidi (2017) studied writing challenges in one PYP and found that students perceive writing as a secondary skill compared with other skills, and also that “teachers do not give much importance to teaching writing” (p. 189). “Teachers teach so they cover the curriculum providing students with the writing topics that will come in the exam. Therefore, the teaching methods do not aid in the students becoming independent writers” (Alkubaidi, 2017, p.191).

2.2.2. Assessment-related difficulties

One of the more frequently cited difficulties among Saudi students regarding writing concerns its assessment. Assessments in Saudi Arabia are usually summative in nature (Obeid, 2017). Writing is infrequently assessed in Saudi contexts (Al-wassia *et al.*, 2015) because many consider it time-consuming (Hamouda, 2011; Javid & Umer, 2013; Obeid, 2017). Additionally, many perceive writing assessment as unfair, biased and subjectively rated (Hamouda, 2011; Javid & Umer, 2013; Mohammad & Hazarika, 2016, Obeid, 2017). For example, Obeid (2017) found that many participants were dissatisfied with the fairness of assessment methods, and therefore recommended a more holistic approach to assessment. The assessment of writing can also be affected by large class sizes, where it would be difficult to assess and give feedback to many students (Al-Jarf, 2011).

Lewin and Dunne (2000) explored the policy and practice of assessment in developing countries, concluding that assessment was not linked to curriculum objectives, and, for most of the students, the purpose of an exam was simply to pass it with a high grade.

Saudi students have also been stereotyped for only caring about attaining high scores and passing examinations with high grades. “The majority of teachers believe that students learn to gain grades and to pass the exams... as this will affect their GPA” (Alkubaidi, 2017, pp.195-196). These pressing concerns with passing examinations and scoring high grades are because the GPA, especially in PYPs, determines students’ academic and professional careers. The GPA, as expressed by Alkubaidi (2017, p.203), is “the decider for the students’ majors which in itself is evidence of the exam-centered environment that most Saudi students grew up with throughout their education journey”. Al-Sadan (2000) also described the focus on examinations in the educational system as the “crucial gateway to personal advancement” and a “killer of pupils”. He further notes that the assessment process should instead “be concerned with the pupil...his learning, his aptitude and his personality, not only his learning progress” (Al-Sadan, 2000, p.154).

The task being assessed is also another factor affecting assessment. The difficulty of the task and how demanding it is affects the way students perform and, hence, affects assessment. Kuiken and Vedder (2008), in their study on the relationship between the cognitive complexity of a task and the students’ linguistic performance in L2 writing, found that “written products of the cognitively more demanding task turned out to be more accurate, with significantly lower error ratios...than those of the cognitively less demanding task” (p.48).

2.2.3. Teaching-and-learning related difficulties

In the Saudi context, instructions are rigid when it comes to teaching writing. There is no focus on teaching different genres in most Saudi high school curricula (Obeid, 2017). The focus of writing lessons usually lies on teaching the

mechanics of writing more than writing as a process. For example, Al-Haq and Ahmed (1994), in their study of Saudi students' argumentative writing, found that, when writing in English, Arab students in general, and Saudi students in particular, tend to focus on the technicalities and the formalities of writing more than what the text should communicate. According to them, the reasons behind this limitation in their writing are due to the teachers' lack of experience and/or qualifications in teaching writing.

Limited time allocated for writing instruction is another factor affecting students' writing proficiency. Usually, very little time is allocated to teaching writing in various Saudi contexts (Obeid, 2017). This could be because teaching writing is considered time-consuming and complex to assess. Alkubaidi (2017) highlighted a "lack of time and the pressures of finishing the curriculum within a designated timeframe" among the constraints faced when seeking to improve students' writing (p.243).

Another important factor affecting students' writing is lack of practice. Writing, like any other language skill, needs to be practised in order to yield improvements. Fageeh (2003) investigated the writing difficulties of students whose major was English. He found that students had difficulties in writing because of lack of practice or involvement in writing in different genres. Other factors that tend to affect students' ability to write include the appropriateness of materials used in teaching writing and the fact that students are asked to write on topics that are of little interest to them (Alkubaidi, 2017).

2.2.4. Effect of L1 on students' writing

Mother tongue interference is another factor that affects Saudi students' writing competence (Wang 2012). According to El-Hibir and Al-Taha (1992), L1 interference is the main reason for students' difficulties in spelling. They recommended teaching spelling with reference to these common errors, as well as highlighting the difference between the pronunciation and orthography of words when teaching writing. This finding was reinforced by Alhaisoni *et al.*

(2015), as the majority of students' most frequent spelling errors in written texts concern silent vowels, owing to the differences between the Arabic and English language systems.

There are, moreover, other differences between the two languages, including writing, rhetoric and structure that can also affect students' English writing (Al-Haq and Ahmed, 1994; Fageeh, 2003). However, students may have similar difficulties with their writing in L1 (Doushaq, 1986). Khuwaileh & Al Shoumali (2000) investigated students' writing skills in both English and Arabic, and found that "poor writing in English correlates with similar deficiencies in the mother tongue" (p.147). Fageeh (2003) studied 37 male students' beliefs concerning the difficulties in writing from different perspectives, including the students' experiences of writing both in English and Arabic. He found that limited exposure to writing in both languages contributed to students' difficulties.

2.3. Improving Saudi students' writing proficiency

According to Shukri (2014), students need to be exposed to materials relevant to their needs, and to shift from rote memorisation attitudes to more creative and self-independent writing practices: "Trying to get students to move away from antiquated exam-centered behaviour to learner-centered behaviour is an uphill battle, yet it is a goal we must continue to address in order to turn these high school graduates into "life-long learners" (Javid *et al.*, 2012, p.65). El Tantawi *et al.* (2016) similarly noted that "There is a pressing need to increase awareness of students about what is acceptable writing and which unethical practices [that] should be avoided" (*ibid*, p.152).

Other studies reveal the role of teachers. Alnassar and Dow (2013) emphasized the importance of effective teaching in higher education in Saudi Arabia in developing students' confidence and good learning skills. Alkubaidi (2017) emphasized improving methods of teaching writing at the PYP and not simply blaming the students for their difficulties in writing where "teachers take a back seat and criticize this attitude without attempting to make the material interesting" (Alkubaidi, 2017, p.211). Teachers need to understand the

difficulties students face when writing in L2 (Cox, 2014). Cox (2014) also asked those faculty members to read about academic writing and to understand the challenges the students face. Khafaji (2004) believed that if students are guided with the right supervision and made aware of their weaknesses, they can overcome the complexities of writing (Khafaji, 2004). According to Alhojailan (2015), staff members in specialised colleges need to help improve students' writing by continuously asking students to write, and engaging them in the process (Alhojailan, 2015). Aljumah (2012) studied students' writing using blogs as a new way to teach writing in the Saudi context. Students preferred weblog writing to traditional methods, as it was more motivating. Alkubaidi (2017) found that the choice of topic is important in motivating students to write (p.230).

In Saudi contexts, the purpose of writing in the first year of tertiary education is mostly to "take notes, write reports, answer exam papers" (Doushaq, 1986; Fageeh, 2003) and, therefore, "there is a need for the students to acquire basic writing skills and be aware of the relevant academic genres" (Shukri, 2008, pp.1 - 2). Al-Eisa and Smith (2013), in their review of issues related to Saudi universities and higher education, emphasized the importance of acquiring high proficiency in English to deal with subjects taught in English in Saudi universities.

Additional support, especially while students are studying their majors at university, seems to be an important consideration, since many students join medical and healthcare schools with insufficient proficiency in writing. Ariail *et al.* (2013) investigated the effectiveness of connecting instructions with language expertise in a writing centre in one health science school. Over 90% of students agreed on the effectiveness of the writing centre and that the quality of writing of those students who attended the centre was twice as good as those who did not attend. The study recommended the importance of medical and other healthcare students using the services of an additional writing centre.

Based on the reviewed literature, there is not, to my knowledge, any study exploring why the PYP is not helping the students to meet their expected proficiency, especially in writing. This proficiency is, furthermore, ill-defined, with no clear definition of the proficiency students have or require to cope with university demands. This study, therefore, aims to address this gap. The PYP

constitutes one intensive year that aims to prepare students for their specific colleges. Therefore, following a NA approach, I sought to identify the students' actual proficiency and compare it with the specific proficiency required. With this, I would be able to specify the gaps and give more specific recommendations to help different stakeholders at the PYP and the university to improve the writing course and help the students meet the requirements.

2.4. Needs analysis/assessment (NA)

Needs analysis and *needs assessment* are used interchangeably to refer to the same concept, with the abbreviation *NA* adopted both cases (Brown, 2016).

NA was first introduced into language teaching and learning through the development of ESP in the 1960s (Richards, 2001, p.51). Since then, there has been an increased demand for the development of 'specialized language programs', and applied linguists began to introduce NA to language teaching and learning (Richards, 2001, p.51) to avoid general language courses teaching too many skills and vocabulary that may not be needed or excluding other necessary skills (Long, 2005).

Since the introduction of NA, different types of ESP have been made available based on the learners' specific needs (Belcher, 2009), such as those that are specifically designed for one purpose, like EAP, EMP or English for Occupational Purposes (EOP), or two purposes, for example, English for Academic Medical Purposes (EAMP) (Belcher, 2009, p.2). The area best known among language educators is EAP, which usually targets higher educational levels (Hyland, 2006 and Belcher, 2009).

According to Maher (1986, p.112), EMP refers to English courses designed specifically for medical personnel to meet the specialised language needs of future medical practitioners. Here, themes and topics are selected from the medical field and skills developed are restricted to those needed by students within the field. Instead of "a one-size-fits-all approach", every language course should be treated as involving specific purposes related to its context (Long, 2005, p.19).

Various studies support NA as the basis for any ESP course (Coffey, 1984; Hutchinson & Waters, 1987; Johns & Dudley-Evans, 1991; Munby, 1978; Nunan, 1988; Robinson, 1991) and it can also be used to modify and improve an existing programme (Royse *et al.*, 2009; Wang & Bakken, 2004).

Yet, in NA, the term *needs* is not as straightforward as it might appear. Needs can be used to refer to *wants, desires, demands, necessities, expectations, motivations, lacks, constraints, gaps, deficiencies, requests, prerequisites, essentials* and *requirements* (Brindley 1984 in Richards, 2001, p.54 and Brown, 2016, p.13). Thus, it is difficult to find an agreed definition for NA in the applied linguistics literature (West, 1997). However, different definitions in the literature share commonalities.

Richards (2001) used the term *needs* to express what learners can do and what they should be able to do with the language. Based on a detailed review of definitions for NA in the literature, Brown (1995) provided a 'formal definition' in which he tried to include "everything by combining a number of different definitions" (Brown, 2016, p.3). His definition is as follows:

the systematic collection and analysis of all subjective and objective information necessary to define and validate defensible curriculum purposes that satisfy the language learning requirements of students within the context of particular institutions that influence the learning and teaching situation (Brown, 2016, p.36).

Brown (2016) later added the following points to clarify his earlier definition of NA:

(1) '*[S]takeholder*' refers to anyone who has an interest in the curriculum such as students, teachers, administrators and parents.

(2) A '*defensible curriculum*' refers to a curriculum that can satisfy the requirements of the students and teachers in a specific context "in such a way that it can be successfully defended to and accepted by all stakeholder groups" (*ibid.*, p.4).

(3) The '*necessary information*' refers to all possible information quantitatively or qualitatively gathered to define and validate the 'defensible curriculum' (*ibid.*, p.4).

It is important to identify the exact and specific language needs of the targeted learners, and to arrange these needs "according to priorities" (Richards and Schmidt, 2010, p.389). Simultaneously, NA should "identify general to specific language needs, which can be addressed in developing goals, objectives and content for a language programme [and] provide data which can serve as the basis for reviewing and evaluating existing programme" (*ibid.*, p.24).

The current research focuses on identifying what students can do and identifying "the gap between what is and what should be" (Brindley, 1989, p.65). Learners are an important factor when defining and conducting NA, because this potentially leads to an effective learner-centred curriculum, as both learners and teachers are involved in the process of NA (Nunan, 1988). Learner involvement and a learner-centred curriculum are crucial because, should the programme be irrelevant to learners' needs, they would adopt strategies to expend minimal effort towards achieving the course objectives, consequently affecting their progress and assessment (Murphy, 2003).

NA has limitations which are summarised in Robinson (1991, pp.111-112). These are:

(1) It is not easy to predict professional needs with accuracy, especially with 'pre-experienced learners'.

(2) Usually, when conducting NA, we deal with non-homogeneous groups of learners with different needs to consider.

(3) With all the methods available to conduct NA studies, "no fool-proof method of collecting and analysing data on needs" is available.

(4) NA usually focuses on the product and not the process (Robinson, 1991).

Participants might not be totally aware of their needs because they may be influenced by previous teaching experience (Wang & Bakken, 2004).

2.4.1. NA: purposes, approaches and types

According to Royse *et al.* (2009), there are four main reasons for conducting NA studies. Firstly, if there is a need for “a new program, intervention, or agency” (Royse *et al.*, 2009, p.17); secondly, for “modification or revision” of a programme or policy (*ibid.*, p.18); thirdly, to improve the service delivered (*ibid.*, p.18) and/or fourthly, to develop new services (*ibid.*, p.19). It is used to ensure that the programme or course is relevant and useful to learners (Nation & Macalister, 2009, p.24).

Richards identifies the following purposes of language NA:

- “to find out what language skills a learner needs in order to perform a particular role;
- to help determine if an existing course adequately addresses the needs of potential students;
- to determine which students from a group are most in need of training in particular language skills;
- to identify a change of direction that people in a reference group feel is important;
- to identify a gap between what students are able to do and what they need to be able to do;
- to collect information about a particular problem learners are experiencing” (2001, p.52).

Brown (2016) identified eleven ‘*analysis strategies*’ that can be used to identify students’ needs. The five most common types are:

(1) *Target-situation use analysis*: to identify what “students should be able to do at the end of instruction” (*ibid.*, p.18). This can be conducted by collecting and analyzing examples of language from the target ESP context.

(2) *Target-situation linguistic analysis*: to identify the specific ‘linguistic features’ students will need in their ESP classes by identifying the specific vocabulary, genre, pragmatics etc. used in the ‘real-world’ ESP context in which students are expected to operate.

(3) *Target-situation learning analysis*: to identify “the sort of learning that students will need to engage in, in target ESP situations” (*ibid.*,p.19).

(4) *Present-situation analysis*: “what students can do with the language of the particular ESP...with respect to use, linguistics and/or learning” (*ibid.*,p.19).

(5) *Gap analysis*: to assess “the disparities between the students’ current abilities and what they need to be able to do in the ESP [context]” (*ibid.*,p.19).

International in scope and specific in purpose, ESP has encouraged many researchers and linguists to conduct NA to discover the different needs of language learners (Johns & Dudley-Evans, 1991); however, little has been published on how to implement NA (Long, 2005,p.24). NA can be classified *subjectively* or *objectively* (Brindley, 1989,p.65). *Objective* or *perceived* needs can be conceptualized by stakeholders or participants other than the learners. In other words, information is collected from participants other than the students (Branden, 2006; Berwick, 1989). NA can be about learners’ “use of language in real-life communication situations as well as their current language proficiency and language difficulties” (Brindley, 1989, p.70). *Objective needs* can be determined via questionnaires, interviews, analysis of documents and materials, and observation (Nation & Macalister, 2009, p.25). On the other hand, *subjective* or *felt* needs are identified by the learners themselves (Branden, 2006; Berwick, 1989), “using lists and scales” (Nation & Macalister, 2009, p.25). This “refers to the cognitive and affective needs of the learner in the learning situation...such as personality, confidence, attitudes, learners’ wants and expectations with regard to learning English and their individual cognitive and learning strategies” (Brindley, 1989, p.70).

NA is useful when combining both subjective and objective data (Richards & Schmidt, 2010, p.389). These data can then be translated into learning objectives to inform programme amendment, course design and material preparations (West, 1994).

2.4.2. NA: resources and tools

According to Dudley-Evans and St. John (1998, p.32), Long (2005, pp.24-40), Richards (2001, p.59) and West (1994, p.7), different sources and tools can be utilized to conduct NA research. Learners, former students, teachers, employers, people working or studying in the field, relevant documents and tasks can be the main sources employed, whereas the tools that can be used may include questionnaires, surveys, interviews, observations, tests, document analysis, checklists or rating scales and previous research.

Many modern NAs, “second generation NA” (Huhta *et al.*, 2013, p.14), have focused on “task as the unit of analysis” and usually identified and analyzed the “target tasks [that are] relevant for the communicative needs” of certain groups or learners (Long, 2005, p.4). This brings into the picture the CEFR as a potential scale to be used to carry out NA studies, since the descriptors in the CEFR focus on tasks that should be performed by learners in each level of proficiency (CoE, 2001; North, 2014; Little, 2006). Hence, the CEFR can be used as “a model and a set of benchmarks for a great deal of course development and evaluation”, and can allow for “a coherent needs analysis...that can accommodate different stakeholder perceptions and an analysis of needs from the macro- to the micro-level” (Huhta *et al.*, 2013, p.7).

Nevertheless, as mentioned earlier, the term “needs” is not always clear; furthermore, needs are always changing. It is important, therefore, that “needs are looked at from a variety of perspectives at a variety of times” (Nation & Macalister, 2009, p.30). These perspectives can differ based on the type of need (lack, necessities, wants, objectives or subjective needs), the source of information (present or past learners; teachers; present or future tasks and materials; future colleges, assessors or teachers), the data-gathering tools (text and discourse analysis; frequency counts; interviews; questionnaires; observation; negotiation and discussion; reflection on experience), and the type of information (learning goals; preferred styles of learning; learners’ commitment to learning) (Nation & Macalister, 2009, p.30). When a NA is

conducted also depends on its purpose (before, during, or at the end of a course) (Nation & Macalister, 2009, p.30).

It is important to include various stakeholders in the different stages of a NA exercise (Brown, 2016). The CEFR suggests that learners, teachers and employers could be involved in determining learners' needs (CoE, 2001). This will involve the activities, tasks, language functions involved and situations where the language skills are needed (Richards, 2001, p.33).

Based on the CEFR, Huhta *et al.* (2013) proposed an approach to NA in which they advocate the creation of CEFR Professional Profiles based on involving quantitative and qualitative methods, the importance of linking the collected data to the CEFR levels, and the ethnographic principle of "thick description" of Geertz (1973) (*ibid.*).

Dudley-Evans and St. John (1998) and Richards (2001) discussed quantitative techniques in NA, including designing questionnaires and surveys and reporting results, and some studies used quantitative NA methods (Chostelidou, 2011; Aliakbari and Boghayeri, 2014); others have used qualitative methods, such as Holliday (1995), or combined qualitative and quantitative data sources in a mixed-methods approach (Bosher & Smalkoski, 2002; Önder Özdemir, 2014; Robinson, 1991).

Collecting data from different sources can have contradictory results (Hutchinson & Waters, 1987). "Needs as interpreted by the sponsors may indeed conflict with the needs felt by the learners" (West, 1994, p.12); results should therefore be treated with caution (Robinson, 1991). However, it is vital to include different sources for NA studies. For example, student data on its own could be insufficient for decision-making, since students "cannot be expected to make pedagogical, linguistic and content decisions concerning their studies. The students as pre-experienced learners, may also not be fully aware of their professional needs, as well as their academic needs... and might confuse needs with wants" (Robinson, 1991, p.19). Depending on only one data source could lead to bias; comparison of results between participants should resolve the issue (*ibid.*).

2.4.3. NA: reliability and validity

Like any research method, NA tools should be tested for validity and reliability before application. NA could be “affected by the ideology of those in control of the analysis” (Nation & Macalister, 2009, p.32); it is important, therefore, to consider a wide range of “possible viewpoints when deciding on the focus of the needs analysis, and seek other viewpoints on where change could be made” (Nation & Macalister, 2009, p.32) in order to maintain reliability. Considering only one source of information could lead to an incomplete picture; triangulation is therefore recommended, i.e. information is collected from two or more sources (Richards, 2001, p.59). Triangulating sources and NA methods improves the reliability and validity of NA findings (Long, 2005; Dörnyei, 2007). Effective NA includes a ‘multiplicity of perspectives on the professional contexts’, thus encouraging triangulation to reach the ‘thick description’ of language needs from different perspectives, involving ‘sequential or concurrent use of both quantitative and qualitative methods’ (Huhta *et al.*, 2013, pp.24-26). “The more pieces of observation and the more people who are studied, the more reliable the results” (Nation & Macalister, 2009, p.30). The reliability of NA can be increased by having a well-designed tool for collecting NA data: “using well-thought-out, standardised tools that are applied systematically” (Nation & Macalister, 2009, p.30).

With respect to the validity of NA, researchers should look “at what is relevant and important” when conducting a NA study (Nation & Macalister, 2009, p.24). Therefore, it is important to consider the “type of need that is being looked at and the type of information that is being gathered” (Nation & Macalister, 2009, p.24). To be able to discriminate what is important from what is not, researchers should rank the “activity to decide what type of need should get priority in the needs analysis investigation” (Nation & Macalister, 2009, p.30), since “better-conducted needs analyses...will enhance the quality of language teaching programs based upon them and, thereby, success rates for language learners” (Long, 2005, p.12).

2.5. Needs analysis and language proficiency

This section briefly discusses the definitions, concepts, models and assessment of language proficiency in order to understand this concept and how it can be integrated as part of this study NA. This is important especially that the needs analysis of this study is focusing on the identification of learners' writing proficiency (both attained and required). For this reason, it is important to review and understand this concept to implement the NA followed in the current study.

The concept of language proficiency started in the 1970s (Harsch, 2016); nonetheless, the literature has no clear agreed definition of the concept, which is not universally understood (Murray & Hicks, 2016), although it is largely assumed to be the ultimate goal when teaching or learning a language (Harsch, 2016). There is no specific threshold definition that can differentiate between proficient and less proficient learners (Murray, 2016a), although it is assumed that it can be divided into levels, such as 'elementary', 'intermediate' and 'advanced' (Harsch, 2016).

One of the first attempts to define language proficiency was related to the concept of communicative competence of Hymes (1972), where language proficiency referred to learners' knowledge of language and their ability to use it (Canale & Swain, 1980). A decade later, Bachman (1990, p.16) defined it as "the knowledge, competence, or ability in the use of a language, irrespective of how, where, or under what conditions it has been acquired". Bachman's communicative language ability is divided into different competences, as summarised by Moore and Harrington (2016):

It separates communicative language ability into knowledge structures, language competence, strategic competence, psychophysiological mechanisms and context of situation. Language competence itself is separated further into grammatical and textual (organisational competence), and illocutionary and sociolinguistic (pragmatic competence) (pp.388-389)

Cummins (1980) differentiated between two concepts: 'basic interpersonal communication skills' (BICS) and 'cognitive academic language proficiency' (CALP). The former is needed to interact with people in everyday language communication. The latter broadly refers to academic skills required in school or university. Hulstijn (2011, 2015) also proposed a distinction between basic language cognition (BLC) and higher language cognition (HLC). BLC deals with the basic linguistic features, which "may occur in any communicative situation, common to all adult L1-ers, regardless of age, literacy, or educational level" (Hulstijn, 2015, p.230). HLC includes all the features of BLC but has additional knowledge and the use of higher complexity and less frequently used words in spoken and written language. "HLC discourse pertains to topics other than simple everyday matters, that is, topics addressed in school and colleges, on the work floor, and in leisure-time activities" (Hulstijn, 2015, p.231).

Lea and Street argued for a new model of language proficiency with particular – although not exclusive – relevance to understanding writing proficiency in an academic context (Lea & Street 1998; 2006). They categorised proficiency into three models: (1) study skills, (2) academic socialization and (3) academic literacies models. The study skills model is based on the concept that language skills (i.e. study skills) are transferable and "students can transfer their knowledge of writing and literacy easily from one context to another" (Lea & Street, 2006, pp.368-369). In this model, the focus is on the 'surface features' of the linguistic elements of writing, such as grammar, spelling, punctuation and the other mechanics of writing. The academic socialization model focuses on "the students' acculturation into disciplinary and subject-based discourses and genres" (Lea & Street, 2006, p.369). This model presumes that disciplinary approaches and genres are generic across contexts, and if students "have learnt and understood the common ground rules of a particular academic discourse they are able to produce it unproblematically" (Lea & Street, 2006, p.369). This model, which assumes that writing is transferrable across different writing contexts, is commonly accepted by many educators (Leki, 2007); however, two experimental studies in Leki (2007), where students who had enrolled in academic writing courses and attained high scores failed to produce writing texts in specific disciplines, have disproved this assumption.

The academic literacies model is “concerned with meaning making, identity, power and authority and foregrounds the institutional nature of what ‘counts’ as knowledge in any particular academic context” (Lea & Street, 2006, p.369). This differs from the academic socialisation model in that academic literacy is viewed “as more complex, dynamic, nuanced, situated, and involving both epistemological issues and social processes including power relations among people and institutions, and social identity” (Lea & Street, 2006, p.369). As well as the second model, the academic literacies model examines additional genres specific to certain disciplines or subjects, while also considering the context and institutional rules and power that can govern the students’ skills and writing (e.g. regarding plagiarism, rules of writing, feedback etc.), and “in more specific contexts such as variations across individual faculty members’ requirements and even individual student assignments” (Lea and Street, 2006, p.369).

These three models are not “mutually exclusive, rather they overlap” (Lea & Street, 2006, p.369), and can be applied to any academic context. There are also various overlaps at a theoretical level. For example, both the academic socialization and academic literacies models focus attention on the relationship between epistemology (knowledge) and acts of writing and literacy in subject areas and disciplines (Bazerman, 1988; Berkenkotter & Huckin, 1995).

Based in part on Lea and Street’s three models of language proficiency, Murray (2010, 2013, 2016a) proposed a tripartite model of English language proficiency: (1) general proficiency, (2) academic literacy, and (3) professional communication skills. Murray postulates that ‘general proficiency’ is the basic and a prerequisite for the other two types of proficiency (Murray, 2010). It refers to a set of general skills “that enables its users to express and understand meaning accurately, fluently and appropriately [...] proficiency is reflected in learning that includes a focus on grammar, phonology, vocabulary development, general listening, reading and writing skills, communication strategies” (Murray, 2011, pp.154-155). Academic literacies refer to a set of literacy practices specific to a domain or discipline where students “need to become conversant if they are to develop and perform effectively as students of that discipline” (Murray, 2011, p.156). Professional communication skills “refers to that range of skills and

strategies that bear on communicative performance in professional settings” (Murray, 2011, p.157). In this model, students need additional skills and competencies in the workplace associated with a specific profession.

The current study explores participants’ different views on writing proficiencies, as students transition between requiring basic general skills towards academic literacies within their medical disciplines.

There are two common perspectives on the assessment of language proficiency: (1) the assessment performed by an ‘outside agent’ through tests and examinations by teachers or trained examiners; and (2) assessment performed through “the learner’s own perspective” as an ‘internal’ or ‘self-directed activity’, where learners are asked to assess their own abilities (Oscarson, 1989, p.1). The former assessment is perceived to be more reliable, whereas the latter is sometimes perceived as unreliable to determine learners’ language proficiency (Oscarson, 1989), as discussed in section 2.6.3 below.

In relation to the ‘outside agent’ of the identification of language proficiency, Thomas (1994) discussed four proficiency measurements. First, proficiency according to institutional status, i.e. identifying the students’ proficiency based on “hierarchically-organized social structure” (1994, p.317). This measure of proficiency is not based on any theoretical grounds and is thus criticised for its low validity (Callies *et al.*, 2013, p.5). Second, proficiency can also be determined by an ‘impressionistic judgment’. This means using recommendations from others without the administration of tests. It also can be referred to as “word of mouth”. However, this is also an unreliable indicator of proficiency “due to the subjectivity of the judgments” (*ibid*, p.6). Third, standardized proficiency tests such as IELTS, TOEFL (Test of English as a Foreign Language) and PTE (Pearson Test of English) can also be used. These tests “report clear construct definitions and validity arguments that include relatively high reliability coefficients” (*ibid.*, p.6). Fourth, “in-house” customised tests are another measure of proficiency. However, these need very careful planning to maintain validity and reliability (Thomas, 2006).

Language proficiency frameworks, such as the CEFR (reviewed in detail in section 2.7 below), are used as “relatively complex and multidimensional conceptualization of language proficiency” (Harsch, 2016, p.2). In this current study context, the CEFR emerges as a potential framework which offers a descriptive scheme in different general aspects of the language and what learners can do using a set of statements in six ascending levels of proficiency (CoE, 2001). The CEFR has achieved “traction across Europe and, indeed, around the world as the fashionable way to compare and contrast actual and target language proficiency across users, courses, examinations, institutions, etc.” (Callies *et al.*, 2013, p.7). But it is important to emphasise here that the CEFR focuses on general language proficiency (Alderson, 2007). It only partially reflects academic literacy in some B2 descriptors in some of the CEFR scales (McNamara *et al.* 2018). Adapting and understanding the CEFR scales and descriptors in academic writing research is important (Byrnes, 2007; Alderson, 2007). Some studies have incorporated the use of the CEFR to analyse academic writing, which included linking specific academic writing examinations to the CEFR (Callies *et al.*, 2013); others focused on examining the adequacy of the CEFR in representing academic literacies in their descriptors (McNamara *et al.* 2018).

Present-Thomas, Weltens and de Jong (2013) mentioned that the CEFR can provide ‘rough descriptions’ of individual learners’ proficiency. Different ways, according to them, can be used to identify learners’ writing proficiency. For example, the use of CEFR-based self-assessment is considered a valid and cost-effective method to get a general overview of students’ proficiency. However, the use of self-assessment to identify proficiency is not without its concerns as discussed below (section 2.6.3). The CEFR can also be used by teachers to identify their learners’ language proficiency (Fleckenstein, Leucht & Köller, 2018). Again, this is not without its concerns (see section 2.6.3). Using holistic ratings/judgements, if well-planned, can give reliable results (Present-Thomas *et al.*, 2013). However, this method is less practical in terms of cost and time (*ibid.*).

Understanding the concept of proficiency is vital to identifying learners’ proficiency in any specific context. Also, NA is one possible approach to identifying and understanding learners’ proficiency (Long, 2005). The

emergence of different concepts in language proficiency, and specifically academic literacies, has shifted the focus of language courses, emphasising the particular writing requirements of each discipline (Murray, 2016b). For this reason, policy-makers in language programmes, such as the PYPs, need to identify and consider the different skills and academic literacies required when planning their curricula. Policy-makers must also “decide how much, and what nature, of English proficiency is sufficient to ensure optimum health outcomes...or academic learning” required of students at certain levels in specific disciplines (Elder & Harding, 2008, p.34). The CEFR implies “a degree of harmony about how proficiency is currently conceived” (Harsch, 2016, p.4), and that it can be used to identify learners’ proficiency, but it has limitations, discussed below (section 2.7.7).

In the current study context, the PYP did not permit administering an additional test to the students, apart from the CEFR self-assessment. Therefore, self-assessment was used, along with tutors’ judgements of the same students using the CEFR, triangulated with raters’ ratings of texts from the standardized exam.

The importance of identifying the students’ levels (required and acquired) is vital in the current NA study for identifying gaps between the two perspectives. This is especially important because, according to Gunn, Hearne, and Sibthorpe (2011), most university staff expected that matriculating students should arrive on the programme with the academic literacies required of their discipline. An example of this is the Jefferies *et al.* (2017) systematic review assessing the importance of academic literacy among nursing students, which concluded that “educators should not assume academic literacy skills upon commencement of an undergraduate nursing programme” (p.84).

Since the expectation of students arriving with the required academic literacy already is rarely met (Jacobs, 2007; Jefferies *et al.*, 2017), and separation between language and content is found to be extremely challenging (Brunfaut, 2014), collaboration between language instructors and discipline specialists is an effective way to improve students’ literacies. For example, Canale & Swain (1980) emphasized the importance of collaboration between ESP teachers and content teachers to achieve students’ language goals. Jacobs (2007) also found that

thorough and sustained communications and interactions between language teachers and 'disciplinary specialists' are vital to improving students' academic literacies (p. 59). Crocker (1981) suggested negotiation between subject teachers and language specialists to decide on specific learning objectives. Beside the collaborations between subject specialists and language teachers, "one-on-one writing support, embedded literacy, and writing tutorials...have been successful in supporting students with lower English proficiency levels" and help improve students' academic literacies (Jefferies *et al.*, 2017, p.90). This collaboration is important, not only amongst academic staff and language teachers, but also for policymakers, in order to understand students' proficiency and how this should be assessed, for policymakers may have misconceptions regarding who is responsible for testing the linguistic elements of the content (Pill & Harding, 2013; Taylor, 2009). Is this the responsibility of the language teachers, or the content teachers?

2.6. Self-assessment

Self-assessment is defined as "procedures by which the learners themselves evaluate their language skills and knowledge" (Bailey, 1998, p.227). This means learners reflect on their abilities, skills and knowledge of a foreign/second language (Little & Perclová, 2001). The CoE defined self-assessment as learners' "judgements about [their] own proficiency" (CoE, 2001, p.191). Self-assessment can involve both the assessment of the process of learning and the product (Brown & Harris, 2013). Learners can describe their abilities (i.e. "these are the characteristics of my work") or evaluate their abilities and skills (i.e. "this is how good my work is and what is it worth") (Brown, Andrade & Chen, 2015, p.444).

2.6.1. Purposes of self-assessment

Educators and researchers use self-assessment as an instrument of formative assessment (Oscarson, 1989), for selection and placement purposes (LeBlanc

and Painchaud, 1985) or for certification purposes. Dickinson's (1982) participants' self-assessments aligned closely with their tutors' assessments. Chalhoub-Deville and Deville (1999) also considered self-assessment to be an initial indicator of students' language ability. As a lifelong skill, students can improve their evaluation and assessment skills (Oscarson, 1989) using self-assessment in their learning (Jacobs & Farell, 2003) and can share the burden with teachers and become more responsible for their own assessment (Oscarson, 1989).

Self-assessment is also used in language teaching and learning and can be an important element of learner autonomy (Dickinson, 1982; Harris, 1997; Little, 2002). Hung (2009) investigated students' writing using electronic portfolios, and found that self-assessment encouraged autonomous learning and was an effective teaching and learning strategy. Self-assessment can also have positive impacts on students' attitudes to learning (Little, 2005), their development (Stefani, 1994; Taras, 2001), and their awareness of their skills and proficiency (Little & Erickson, 2015, p.121). For example, when self-assessment practice was related to assessing writing, it was found to be beneficial, as it fostered the metacognition of the students, who reflected on their writing, observed their progress, and realized their weaknesses and difficulties in writing (Campillo, 2006; Falchikov & Bound, 1989). With self-assessment, learners can improve their perceptions of their abilities (Zimmerman, 2000) and identify gaps in their learning (Ekbatani, 2000).

2.6.2. Forms of self-assessment

Well-constructed self-assessment questionnaires are important "to produce high quality results" (LeBlanc & Painchaud, 1985, p.683). Forms range from simple to more detailed and extensive instruments (Oscarson, 2013a; Denies & Janssen, 2016). An example of a simple self-assessment tool is the use of cards to self-assess a specific task, activity or skill (Oscarson, 1989). Using questionnaires, rating scales and check-lists are other common methods of self-assessment. "The

basic requirement for any questionnaire is to give learners the opportunity to indicate what they think they can do with the language they are studying” (LeBlanc & Painchaud, 1985, p.677). The questionnaire might include open-ended or closed questions and can include a rating scale (e.g. with a range of 5 to 10 points). More recently, can-do self-assessment statements have become commonly used methods, the best-known example of which is the self-assessment grid in the CEFR document (CoE, 2001). The European Language Portfolio (ELP) uses the CEFR self-assessment grid as part of the portfolio to encourage learner autonomy (Little, 2005). The DIALANG test, which is “an assessment system intended for language learners who want to obtain diagnostic information about their proficiency” (CoE, 2001, p.226), uses self-assessment can-do statements based on the CEFR global and other descriptors from the reading, writing and listening illustrative scales of the CEFR (Alderson, 2005). Other studies have adapted the CEFR scales and descriptors for self-assessment. For example, Hasselgreen (2005) adapted the CEFR descriptors to design can-do statements for on-going self-assessment in their class.

2.6.3. Reliability and accuracy of self-assessment

Many concerns have been raised about the accuracy and reliability of self-assessment, because it is subjective. However, as stated by Oscarson (1989), “subjectivity does not necessarily invalidate the practice of self-assessment techniques in language testing and evaluation and...self-assessment may be motivated by reasons that go beyond mere evaluation” (p.2). There are concerns that learners might not have reached the proficiency level needed for them to judge their abilities (Blue, 1988, p.100). LeBlanc and Painchaud (1985) have noted that “[a] fine-tuned self-assessment ability certainly does not come automatically to all students...students simply do not have the tools to cope with a self-assessment situation that requires them to describe with some precision their level of proficiency” (p.675). Blue (1994) found no relationship between students’ assessment and their language test scores, concluding that self-ratings were unreliable. Therefore, to improve reliability, training “under proper

guidance” is required (Oscarson, 1989, p.3), as self-assessment can be “more accurate when learners receive some training” (CoE, 2001, p.191; Leach, 2012; Jafarpur, 1991). Chen’s (2008) findings supported this, indicating that students’ assessment of their speaking aligned more closely with that of their teachers in the second round of a longitudinal study, following training.

In undertaking any self-assessment with students, it is important to take their language proficiency into consideration, since this does affect the accuracy of self-assessment (Heilenman, 1990; Oskarsson, 1984; Shimura, 2006). The more proficient learners are, and the higher the academic ability they possess, the more accurate their self-assessment will be (Barnett & Hixon, cited in Brown, Andrade & Chen, 2015).

Blanche and Merino (1989) conducted a meta-analysis, and concluded that self-assessment can be very accurate. However, they found that the more proficient students tend to underestimate their abilities, whereas those who are less proficient tend to overestimate their proficiency. A decade later, another meta-analysis conducted by Ross (1998) confirmed these findings. In 2006, Ross found a positive correlation between self-assessment and other measures of students’ proficiency, and that more proficient students were found to be more accurate in their assessment than less proficient students. Üstünlüoğlu *et al.* (2012) found that students’ perceptions regarding their language competency level was affected by their actual proficiency level at the PYP in a Turkish university. Pre-intermediate-level students felt more competent compared to higher-level students.

This tendency is evident not just in terms of general competencies but also specific skills. In Sahragard and Mallahi (2014), 30 Iranian upper-intermediate EFL students learning English at one of the country’s English institutes in Iran were asked to write an essay and then to self-assess their writing by completing a checklist. The self-assessments and researchers’ assessments were compared, and the findings again showed that the more proficient students tended to underestimate their writing abilities, while the less proficient students overestimated them.

This behaviour has been explained differently by various researchers. McLeod, cited in LeBlanc and Painchaud (1985), claimed that more proficient students underestimate their language proficiency because “they have some notion of all that remains to be learned”, whereas less proficient students tend to overestimate their proficiency as “they cannot perceive a need for improvement” (p.675). Boud and Falchikov (1989) reviewed studies comparing students’ and teachers’ assessment to explore areas relevant to self-assessment. They found that “Able students working in a new subject are likely to be aware of their own deficiencies ...weaker and less mature students tend to overrate themselves and the weaker they are...the greater the degree of overrating” (p.544).

This could be due to the “Dunning-Kruger effect” (Kruger & Dunning, 1999), according to which less competent learners have less ‘metacognitive’ ability to recognize their own proficiency. Kun (2016) and Hodges, Regehr and Martin (2001) conducted two studies on self-assessment aiming to confirm or reject the Dunning-Kruger effect. Both studies confirmed the hypothesis, where higher-achievers tended to underestimate their levels and lower-achievers to overestimate them.

The inaccuracy of self-assessment of low-proficient learners can also be attributed to the tendency of learners to be more optimistic and potentially not pay attention to important information required to self-assess properly (Dunning, Heath & Suls, 2004). The class environment can also be a factor where students tend to overestimate their levels, as learners feel pressure to overstate their abilities (Saavedra & Kwun,, 1993). Boud and Falchikov’s (1989) meta-analysis found that students tended to inflate their results when self-assessment had an impact on their grades. Tejeiro *et al.* (2012) found that self-assessments were significantly higher than marks given by the professor, especially in the group where self-assessment counted towards the final mark. Possible “reasons for this are the desire to obtain the highest possible grades and the stress associated with self-assessment” (p.790).

To improve the accuracy and reliability of self-assessment, it can be combined with other measures. For example, Alderson, Brunfaut and Harding (2014) called for future studies to combine subjective measures such as self-assessment with

objective ones to improve assessment. Self-assessment can be accompanied with teachers' assessments (Oscarson, 1989), or with other judgements, such as those of qualified raters or fellow learners, or with performance on tests (Brown, Andrade & Chen, 2015, p.445).

Accuracy increases when students are aware that their assessment will be compared to others (another peer or assessor) (Fox & Dinur, 1988). Lejk and Wyvill (2001) compared self-assessment with peer assessment, and found that secret assessment (where students do not know who they are assessing) gave more accurate results than agreed assessment (students do know who they are assessing).

In various studies, however, the correlation found between self-assessment and teachers' assessments or performance in a test was often weak (Brown & Harris, 2013; Brown, Andrade & Chen, 2015). For example, Brown and Harris (2013) found that the correlation between self-assessment and other measures was weak to moderate, with an average of no more than $r=.60$. Falchikov and Boud (1989), in their meta-analysis, found a moderate correlation ($r=.39$), with a tendency for students to overestimate their proficiency.

However, teachers trained in methods of assessment had a higher correlation with students' academic achievement scores. Südkamp, Kaiser and Möller's (2012) meta-analysis, including 75 studies comparing teachers' judgement on students' academic achievement and the students' actual scores in academic achievement tests, found that the overall mean effect size was 0.63, and informed judgments (where teachers were given clear instructions for assessment) gave higher correlations than uninformed judgements.

Teachers/raters can also identify students' language proficiency. However, we need to be cautious with the results, especially if the teachers are untrained or inexperienced. Davis (2015) asked 20 raters to rate TOEFL speaking tests and investigated two factors, rater training and experience, which usually affect the consistency of raters' judgements. He found that "training resulted in increased inter-rater correlation and agreement as well as improved agreement with established reference scores". Experience, on the other hand, had a small effect

on raters' consistency, "although the level of agreement with reference scores continued to increase" (p.117). Fahim and Bijani (2011), in their study of the effectiveness of training in reducing bias among raters, found that "raters who were identified as being highly severe/lenient and biased in particular categories of the rating scale were no longer biased after training" (p.1).

Training regarding the self-assessment tool and scoring/rating criteria is crucial not only for teachers, but also for students taking part in self-assessment. Babaii, Taghaddomi and Pashmforoosh (2016) studied the (mis)match between students' self-assessment of speaking and teachers' assessment before and after training. This revealed that providing the learners with the scoring criteria and a follow-up practice session "minimised the existing mismatch" between students and teachers (p.413).

Sebba *et al.*'s (2008) systematic review highlighted the importance of students understanding the concept and tools for self-assessment, and of receiving training prior to using the tools to achieve reliability. This means the design of the tools is very important. Ross (1998) found that a statement in the self-assessment tool which "exemplifies achievement functional ('can do') skills" helps students in self-assessment and improves reliability (Ross, 1998, p.16). Learners are more accurate in estimating their performance when there are specific and concrete criteria, goals or standards they can use for self-assessment (Andrade & Valtcheva, 2009).

Bachman and Palmer (1989) explored the reliability of different types of questioning for the self-rating of grammatical, pragmatic and sociolinguistic competencies among 116 non-native speakers. They found that "self-rating can be reliable and valid measures of communicative language abilities...[and] some measures proved to be reasonably good indicators of specific language abilities" (p.14). The most effective types of questions are the ones "which asked about subjects' perceived difficulty, with various aspect of the language. This suggests that foreign/second language users may be more aware of the areas in which they have difficulty than they are of the areas they find easiest" (p.23). Age and experience can be other factors affecting accuracy. Blatchford (1997), for

example, in his longitudinal study, found that accuracy is less between ages 7–16 than with older students.

Giving rewards, incentives and feedback for self-assessment motivates students and increases accuracy. Miller and Geraci (2011) found that giving incentives and feedback increased accuracy, especially among lower-performing students. However, giving incentives might not always work and therefore “this tactic should be employed with caution” (Brown, Andrade & Chen, 2015, p.449).

Studies have also showed positive associations between self-assessment and achievement tests (Brown & Harris, 2013), though few have addressed the validity of self-assessment (Brown, Andrade & Chen, 2015), i.e. to what extent students are truthful in their assessment (Brown, Andrade & Chen, 2015). Self-assessment is considered particularly useful at the culmination of a course or programme (Blue, 1994, p. 5).

Giving all the limitations mentioned above with using self-assessment to identify language proficiency, the triangulation of data obtained is the most appropriate solution to increase the reliability of the current study results. In this study, self-assessment was combined with teachers’ assessments and rating the same students’ texts. I made sure to use well-designed tools with clear, detailed instructions, especially given insufficient time for proper training. Different incentives were also provided to participants (including personalised feedback, explained in Chapter 3) to encourage participation and accurate responses.

The CEFR has been identified as one of the tools that can be used both for self-assessment and for NA. Since the CEFR is the proficiency framework used in the current study, the following section gives an overview of the CEFR and its different uses reported in the literature.

2.7. The CEFR

2.7.1. What the CEFR is and is not

The CEFR is “a framework, which describes language learners’ ability in terms of speaking, reading, listening and writing at six reference levels” (Cambridge ESOL, 2011,p.1). This framework comprehensively describes what learners need to do at each level to communicate effectively and the language knowledge and skills required to do so (CoE, 2001). It also ‘defines levels of proficiency’, which allows for the measuring of learners’ progress at their different stages of language learning (*ibid.*, p.1).

The CEFR was the product of different meetings, discussions and symposia held by members of the CoE to discuss and agree upon issues related to European language teaching and learning (Morrow, 2004). These meetings aimed at developing tools that could help in planning and constructing courses adapted and adjusted to the learners’ “needs, motivations and characteristics” (Trim, 1978,p.1).

The CEFR was first drafted in the 1990s by the CoE, and published in English and French in 2001. The CoE is a ‘political organization’, established in Europe in 1949. The Council has different foci, including the development of international understanding and the protection of ‘human rights’ (North, 2014,p.8). The CoE recognized the needs for a common reference that brings ‘understanding’ and ‘tolerance’ to the diversity of European languages and culture (CoE, 2001), and identified the need for the development of a framework to “make the process of language learning more democratic by allowing learners to ‘steer and control’ their own progress” (Trim, 1978, p.1). Alderson (2007) mentioned that the framework is “genuinely European” and it was intended to serve European countries.

Since its publication in 2001, the CEFR has been translated into 40 languages and has influenced the work of different sectors and institutions related to language learning, teaching and assessment around the world (North, 2014). The CEFR has not only attracted attention and interest within Europe, but its effects have been

felt further afield (Little, 2007), including in America, Asia and Australasia (Byram & Parmenter, 2012).

At a micro-level, practitioners use the CEFR to reflect on their practice and track their learners' progress, with a view to improving students' learning and assessment (Cambridge ESOL, 2011,p.6). Learners can “analyse their own needs, conduct an informal audit of their current level of proficiency, identify their learning goals and construct their own programme of learning” (Little, 2011, p.388). At a macro-level, stakeholders can use the CEFR to critically reflect upon existing practices, and to compare practices to one another (Broek & Van den Ende, 2013, p.27). It can be a useful tool to allow communication and enable practitioners to talk about objectives and language proficiency levels in a ‘more coherent way’ (Cambridge ESOL, 2011,p.6).

The CEFR is also seen as a catalyst in promoting cooperation among educational institutions, not only regionally but across countries (CoE, 2001, p.6). By offering “a sound basis for the mutual recognition” of learners' language qualifications (*ibid.*, p.6), the CEFR can “facilitate reflection, communication and networking” among different educational parties (North, 2014, p.9). Furthermore, it can be used as ‘a common basis’ for the design and elaboration of “language syllabuses, curriculum guidelines, examinations, textbooks, etc.” (CoE 2001, p.1). In this way, the CEFR can assist various stakeholders in educational organizations, including learners, teachers, curriculum and course designers, examining personnel and educational administrators. This is because by providing a ‘common basis’, it allows all these stakeholders “to situate and co-ordinate their efforts” (*ibid.*, pp.5-6). In turn, it provides a “metalanguage” that can be used among all language and educational practitioners. This “common language” assists understanding of objectives, aims or levels that learners need to achieve. Practitioners can also use the CEFR for planning and designing language programmes, defining “assessment criteria” and “raising learners' awareness” of their present knowledge or skill level (CoE, 2001, p.6).

However, using the CEFR does not mean that curriculum outcomes should be stated solely based on the CEFR. For example, stating that “school leavers are expected to achieve B2” by the end of the programme (Little, 2011, p.388), as is

now the case in many institutions, is not what was intended when considering the use of CEFR scales. The use of 'can-do' statements should be based on empirical evidence tailored to learners' needs in specific contexts. The CEFR can also be used "to analyse learners' communicative needs, and to describe the language they must learn in order to fulfil those needs" (Little, 2006, p.174). This can be done when stakeholders "reflect on their current practice with a view to situating and coordinating their efforts and to ensuring that they meet the real needs of the learners to whom they are responsible" (Little, 2011, p.382). The CEFR "can-do" descriptors should assist in defining learning outcomes and, perhaps, areas of the learning process, while tailoring curricula to learners' needs, which could encourage initiative and engender a degree of control in their learning (Little, 2011, p.388).

Furthermore, the CEFR can be used to "provide principles and approaches to the teaching, learning and assessment of all languages" (Saville, 2010). The CEFR is considered innovative in its ability to bring these areas together (curriculum, pedagogy and assessment) in more interdependence than traditionally the case (Little, 2006; 2011).

However, the function of the CEFR should not be misconstrued. The authors made it clear from the outset that CEFR does not tell users "what to do or how to do it" (CoE, 2001, p.xi); the Council states that "it is not the function of the Framework to lay down the objectives that users should pursue or the methods they should employ" (*ibid.*, p.xi). The Framework neither functions to enhance a particular teaching strategy or technique, nor does it support one methodology over another (CoE, 2001; Coste, 2007; Little, 2006). It is also not designed to achieve "precision" to specific skills or specific contexts, and it is not a tool for "precise equating" and through which "to impose standardised solutions" (Saville, 2010). The CEFR should not be seen as an "alternative system" for grading (Little, 2006). In addition, the CEFR document was not designed to offer specific recommendations, suggestions or guidelines; rather, it is a framework to "describe and not to prescribe" (Morrow, 2004; Little, 2006). It is not intended to be used as a tool for centralization and harmonization (Jones & Saville, 2009); on the contrary, it is intended to be flexible and adaptable to different contexts. Its

function is simply to provide options, “to judge between them” and to encourage users to choose and decide on what best suits their interests and needs and then to indicate “what consequences their choices have for their practice” (CoE, 2001, p.113).

2.7.2. CEFR: theories, approaches and Scales

The CEFR is based on theories and approaches in language teaching and learning (including the Communicative Approach), while driven by learners’ needs (Little, 2006, p.175). The CEFR takes an action-oriented approach. It considers learners as “social agents” who have an important role in society and who have tasks they need “to accomplish in a given set of circumstances, in a specific environment and within a particular field of action” (CoE, 2001, p.9). This action-based approach takes into consideration the “cognitive”, “emotional” and “volitional resources” of the individual as a “social agent” (*ibid.*, p.9). According to the CEFR authors:

Language use, embracing language learning, comprises the actions performed by persons who as individuals and as social agents develop a range of competences..., The monitoring of these actions by the participants leads to the reinforcement or modification of their competences (CoE, 2001, p.9).

The authors developed the “common reference levels of language proficiency”, organized in a grid (*ibid.*, p.16). “A proficiency level is a band which allows some variation, but still, a given level has some characteristics that distinguish it from the level below and the one above” (Carlsen, 2012, p.163). On the vertical dimension, the six levels are presented and arranged from *basic* (A1 and A2) to *intermediate* (B1 and B2) to *proficient* (C1 and C2). On the horizontal dimension, there are descriptors of what learners can do with the foreign language at each level. In addition to the six proficiency levels, CEFR authors suggested three additional “more or less equally sized, coherent...plus levels” (CoE, 2001, p.31), e.g. A2+ between levels A2 and B1, B1+ between B1 and B2, and B2+ between B2 and C1.

Today, the CEFR is a product of several decades of work, considering learners' needs and the notional and functional approach designed to promote communicative competence and autonomous learning (Little, 2006). The CEFR document distinguishes between receptive skills (reading and listening) and productive/interactive skills (writing and speaking) (Little, 2011). Levels A1 to B1 reflect basic communicative activities, whereas the levels from B2 upwards are more associated with "academic or professional L2 use" of language (Little, 2011, p.386). This division, as explained by North (2007), was inspired by Cummins's (1984) distinction between BICS, based on the social incidental situation, and CALP, based on more intentional formal learning (Hulstijn, 2003).

A proficiency scale is "a series of ascending bands of proficiency. It may cover the whole conceptual range of learner proficiency, or it may just cover the range of proficiency relevant to the sector or institution concerned" (CoE, 2001, p.40). The CEFR document has different proficiency scales: the global scale; the self-assessment grid; and the scales for the language activities with different sub-skills along with the linguistic skills (Morrow, 2004; CoE, 2001; North, 2014).

2.7.3. CEFR and pedagogy

The CEFR has had a major impact on the education policy of many countries in Europe (Jones & Saville, 2009) and elsewhere (Byram & Parmenter, 2012). The impact has been most apparent in curriculum development, test formulation and certification, but less so in teacher training and classroom applications (Jones & Saville, 2009); this is because teachers and students are unfamiliar with the framework, and "are not equipped to make use of it" (Jones & Saville, 2009, p.53), which could be attributed to a lack of training in CEFR use in classrooms (Little, 2007). Many teachers are unfamiliar with its applications (Moonen *et al.*, 2013). Also, using CEFR in the classroom can be time-consuming and might be viewed as an unnecessary 'add-on' for both teachers and students (Faez *et al.*, 2011b).

Nonetheless, the CEFR is a useful tool that can be used in English language teaching; for example, linking B2 CEFR descriptors to specific course objectives,

which are then utilized to establish the criteria for measuring students' performance (Neff-van, 2013). This procedure helped teachers and students to be more critical and to focus on "feasible advancement in critical discourse" more than "elimination of errors", helping students to take an active role as critical citizens in society (Neff-van, 2013, p.207).

The CEFR can also be used to improve students' (Neff-van, 2013) and teachers' (Hismanoglu, 2013) awareness of specific language skills that might cause concern. Faez et al. (2011a) used "Can-Do" statements to design class activities for teaching French as a second language. Participating teachers gave positive reactions, and Faez recommended making such tasks available to teachers for developing future curricula for the teaching of foreign languages. In another study, Faez *et al.* (2011b) investigated students' and teachers' perceptions of using the "Can-Do" descriptors in their classes. They found that using CEFR-informed activities increased students' motivation, self-confidence, self-awareness of their levels and abilities, and brought authenticity and autonomy into the classroom (Faez *et al.* 2011b). Students "tak[ing] charge of [their] own learning" (Faez *et al.*, 2011b, p.9), and becoming autonomous is one of the main purposes of using the CEFR, as Little (2006) has observed. The 'authenticity' mentioned by Faez *et al.* (2011b) refers to students' use of authentic language in the classroom to express what they can do in a foreign language using *Can-Do* statements. It also refers to the kind of "goal-oriented authentic activities in the classroom" (*ibid.*, p.16) that were positively highlighted by the teachers. The CEFR is also widely used in "CEFR-based textbooks" (Moonen *et al.*, 2013).

However, it is important to note that the CEFR is neither language- nor context-specific, and does not promote "a functional approach or a task-based approach at the expense of ensuring that learners acquire a knowledge and mastery of a system of the language" (Little, 2007; North, 2006). Therefore, CEFR users need to adapt it to make it suitable for the targeted language and context (Cambridge ESOL, 2011, p.12). The CEFR is a relevant reference tool that is open to different 'methodology and teaching style[s]' (*ibid.*, p.8) to suit the needs of different learners and contexts.

2.7.4. CEFR and the curriculum

Policy-makers often use the CEFR to design and develop language curricula for their programmes or institutions; e.g. setting minimum language requirements for the programme (Cambridge ESOL, 2011, p.2), reflecting the programme aims and objectives (*ibid.*,13), and setting curriculum goals and entry requirements (Hulstijn, 2007).

In these instances, curriculum designers need to identify the CEFR levels of students in each particular institution or programme. For example, Bechger et al. (2009) asked administrators at an institute of higher education to decide on the minimum CEFR levels and scores that students had to achieve to meet their institutional English language requirements in spoken and written skills. A questionnaire with 114 “can-do” statements was used. Participants were asked to choose from 4 scales (from *certainly not required* to *definitely required*). The required level(s) was based on the majority selecting that level as “required” (*ibid.*,p.134). Level B1 was required in Part I of the programme and B2 in Part II.

After designing a CEFR-based curriculum, it is necessary to measure its impact. Üstünlüoğlu *et al.* (2012), implemented a CEFR based-curriculum for tertiary-level students in a PYP in a Turkish university. The CEFR was successfully used to state general English language objectives; however, it was necessary to include specific academic objectives. Lowie, Haines & Jansma (2010, p.153) found that implementing a CEFR-based curriculum provided “a single common structure upon which [they] can base [their] interpretations of the linguistic performance of students”. Üstünlüoğlu *et al.* (2012) assessed the ‘effectiveness of the programme’ at a PYP in Turkey, using a scale based on the CEFR descriptors as a data collection instrument in the four language skills, and focus group (FG) meetings with participants. Students’ perceptions of their competencies correlated with their scores in the language proficiency exam at the end of the PYP. However, while faculty members believed that the PYP met students’ needs, the students themselves felt they were struggling to deal with tasks that required higher-order thinking skills.

An area of study that has utilized the CEFR in a variety of ways is assessment, including the utilization of the CEFR as a reference to examine or identify entrance requirements and standards for university education in different contexts (Deygers, Van Gorp & Demeester, 2018; Fleckenstein *et al.* 2018; Carlsen, 2018).

2.7.5. CEFR and assessment

Though assessment is only one of the three dimensions of the CEFR document, the CEFR has been used more intensively in this area than for teaching and learning (Jones & Saville, 2009; Little, 2007; Figueras, 2007). Using the CEFR within the context of assessment poses specific demands, which have led to the development of materials supporting the CEFR document. For instance, the CoE produced a manual (2009) explaining how to link and align tests to the CEFR. It has increasingly become a “key reference” tool for test designers who want their test(s) to be recognized and accepted within Europe (Harsch & Rupp, 2011, pp.1-2). Researchers have also used the CEFR in areas related to assessment, including designing test specifications (Davidson & Fulcher, 2007), tasks (Harsch & Rupp, 2011) and assessment rubrics (Harsch & Martin, 2012).

The assessment area most relevant to the current study is that of self-assessment using CEFR. In 2001, the CoE defined self-assessment as “judgments about [someone’s own] proficiency” (p.191) solely from the learners themselves, and should not replace standardised tests (Ünaldi, 2016, p.68), but complement them, as well as teachers’ assessment (CoE, 2001, 191, p. 54).

2.7.6. CEFR and self-assessment

Researchers have used the CEFR in different ways for self-assessment. The most common is the self-assessment grid in the CEFR document (CoE, 2001, pp.26-27), as used in Atai and Shoja (2011) and Dragemark Oscarson (2009). Others

adapted the illustrative CEFR descriptors related to language activities or competences for self-assessment (Ünaldı, 2016), or used self-assessment instruments based on the DIALANG diagnostic self-assessment system (Alderson & Huhta, 2005; Luoma & Tanana, 2003). The DIALANG is designed from and directly connected to the CEFR (CoE, 2001). It is to be used with “adults who want to know their level of language proficiency” (Luoma & Tanana, 2003, p.226). Others, such as Muñoz (2014), used self-assessment CEFR scales adapted from EAQUALS (Evaluation and Accreditation of Quality in Language Services) to measure students’ perceived progress in different skills on an EMI course at tertiary level. In this study, ten CEFR scales related to writing, including two from the DIALANG, were used. The same scales with their exact descriptors stated in the CEFR document were used for self- and tutors’ assessment following a guided procedure explained in Chapter 3.

There are different purposes for which the CEFR has been utilized as a tool for self-assessment, including to predict students’ proficiency levels (Atai & Shoja, 2011; Dragemark Oscarson, 2009; Ünaldı, 2016). For example, Ünaldı (2016) asked tertiary level students in one PYP in Turkey to use the CEFR self-assessment grid to assess their proficiency. He then compared the results with their teachers’ scores and their placement test scores using multiple regression. He found that students’ proficiency levels could be predicted by self-assessment and teachers’ assessment (2016, p.78). Alobaid (2016) examined the accuracy of a group of ESL learners using the CEFR self-assessment grid with reference to their gender and level of proficiency by comparing their self-assessment to their TOEFL scores. He found no significant correlation, but the qualitative data suggested that the grid accurately reflected their levels of language proficiency. Engelhardt and Pflingsthorst (2013) compared the Quick Placement Test (QPT- pen and paper-based version) with CEFR-based self-assessment, aiming at evaluating “the predictive power of self-assessment based on the global CEFR ‘can-do’ descriptors in the context of a university language centre placement test” (p.75). They found that the CEFR global scales could be reliably used as a placement tool to a certain degree. Ünaldı (2016) also found that it was possible to predict students’ proficiency levels using a CEFR criterion-referenced self-assessment tool, although with greater caution at lower proficiency levels.

The CEFR-based self-assessment, as used in the DIALANG project, can be used for the purpose of eliciting “feedback on the strengths and weaknesses” of learners’ proficiency (Luoma & Tarnanen, 2003, p.440). Brantmeier *et al.* (2012) used the DIALANG criterion-referenced self-assessment tool tailored to the course objectives and found that self-assessment was “a powerful low-stakes assessment tool”, which could benefit both individual learners and language programmes (p.153).

CEFR self-assessment can also support the development of learner autonomy as it “promotes reflection, helps learners to take responsibility for their own learning, enables learners to see gaps in their learning, and...to take risks” (Ekbatani 2000, p.6-7, cited in Little 2006, p.186). Self-assessment can also be used as a “useful additional tool for learning, which can help learners reflect on their own [language skills], for example, writing” (Luoma & Tarnanen, 2003,p.461). Furthermore, Luoma and Tarnanen (2003) reported that learners were happy dealing and working with the self-rating instruments.

CEFR self-assessment can be used to compare different participant groups. Ashton (2014) studied the use of can-do statements for self-assessment in three different languages: German, Urdu and Japanese. In general, participants used the statements well for self-assessment, but some differences were found in their use of Urdu and Japanese, which required further investigation. Another study, by Fleckenstein *et al.* (2018), investigated the accuracy of CEFR-based judgements of students’ proficiency by comparing the assessment of EFL teachers in an upper-secondary school in Germany to students’ CEFR level based on their scores in TOEFL and ITP scores using the CEFR scales. The CEFR-based judgment was appropriate, although teachers tended to overestimate their students’ level of achievement. The CEFR has, therefore, been used for different purposes, including self-assessment and teacher assessment, and can be compared with other measures, such as objective tests.

In my study, I used eight CEFR scales and two from the DIALANG, as they are in the CEFR document without any adaptation of the descriptors. However, I designed the scales in a specific way to help students in their self-assessment (which I called *Controlled Utilization of CEFR Descriptors*: see details in Chapter

3). I used the CEFR in this study for self-assessment, tutors' assessments and for the identification of the levels required. I also used the scales to compare these different usages to identify any potential misalignments. A rating scale based on the CEFR scales was also used by the raters rating students' exam texts, for the results to be compared with the students and tutors' assessments (see Chapter 4). The use of the CEFR scales by different participants, and for different purposes, in this study contribute to the literature. For example, very few studies, to my knowledge, have compared the use of self-assessment across more than two proficiency levels. One such study (Brantmeier *et al.*, 2012) was conducted in Spain, where a comparison of self-assessment across three proficiency levels (beginning, intermediate and advanced) with achievement tests were conducted. This study found that advanced-level students were the most accurate. The current study will contribute to the literature showing how the CEFR scales are used for different purposes (identifying the levels achieved, required and misalignments between the two) with different participants (students, tutors and teaching staff), across different proficiency levels (elementary, intermediate and advanced) and across different contexts (PYP and MHCCs).

In my study, students' self-assessment across the three levels was compared with their teachers' assessment using the same scales, and then compared with raters' rating a sample of the same students' written texts. Additionally, students' self-assessment of their own proficiency was compared with their assessment of the proficiency required to identify the misalignment statistically. This intensive way of using the CEFR scales for self-assessment is, to my knowledge, the first of its kind. In addition, the way I designed the *Controlled Utilization of CEFR Descriptors* aims to overcome the difficulties (indicated in Luoma & Tarnanen, 2003) that participants often encounter when choosing the appropriate level.

2.7.7. Criticism of the CEFR

Although the use of the CEFR has become widespread, it has also been criticized by users, in terms of its format, development, descriptive scales, levels, and theoretical and empirical underpinnings.

Morrow (2004) reported that the document itself is not 'user-friendly': the font and layout are small and condensed, and the document is long and contains 'ponderous' and specialized words and terminology, making it difficult to read. However, it deals with an inherently complex issue, related to learning, teaching and the assessment of a foreign language, so we should not expect the material to be easy to comprehend. Morrow (2004) encourages CEFR users to work on the parts of the CEFR relevant to their needs and interests, and to redesign them, where necessary, in ways that are more suited to their particular context.

Others have criticized the CEFR for being an "imprecise and theoretically limited framework" (O'Sullivan & Weir, 2011). Its scales and descriptors were based on teachers' judgements (North, 2014); it is not linked to language theories such as the second-language acquisition theories (SLA) (Fulcher, 2004; Little, 2006). However, not all researchers agree with these criticisms. Firstly, the CEFR is based on "linguistic theory and measurement theory" (a theory of language as communication, for example), but the authors were careful not to make apparent and explicit connections between the CEFR and any specific theories (North, 2014, pp.16-22). CEFR proponents decided to leave it to users to decide on the best way to approach the document based on their particular context and interests (CoE, 2001; North, 2014). Secondly, when the descriptors were developed, Second language acquisition SLA theories were not comprehensive enough (North, 2014) to support its development. Nonetheless, the CEFR authors acknowledged that within the areas of categorisation and description, the CEFR "needs to be [more] theoretically grounded" (CoE, 2001, p.21).

Others have expressed concerns over the validity and applicability of the CEFR levels and scales, with some scholars (e.g. Alderson, 2007, Fulcher, 2004, Hulstijn, 2007) believing that they "are not suitable for direct application to many particular contexts". Meanwhile, Callies, Zaytseva, and Present-Thomas (2013, p.1) stated that although the CEFR is flexible enough to be applied in a variety of situations, it is too abstract to be applied "comprehensively to any one in particular". This is also reflected by Jones (2002, p.181) who claimed that "different people tend to understand 'Can-do' somewhat differently". Weir (2005) also suggested that, especially for testing purposes, more details and test

specifications are required. Yet, Nagai and O'Dwyer (2011, p.141) caution against uncontrolled adaptations of the CEFR components, since "the more it is adapted to a specific context, the greater the possibility that the CEFR will lose its validity and the original language proficiency scales will be altered in an unhelpful way".

The CEFR has also received criticism in terms of its suitability for use in academic and specific contexts, with descriptors focused on general language learning, teaching and assessment, rather than academic or specific language (Charpy & Carnet, 2014), and there has been limited incorporation of the academic literacies required in higher education (McNamara *et al.*, 2018). Consequently, it is suggested that there is a need to consider other options besides CEFR in contexts where academic or specific language is the language of interest (Üstünlüoğlu *et al.*, 2012). One of the reasons behind this criticism is the descriptors failing to articulate higher cognitive intellectual academic skills (i.e. EAP) (Hulstijn, 2011). "It is underspecified in terms of the domain of academic literacy" (McNamara *et al.*, 2018, p.25). Therefore, if these skills are important to any institution, they need to be identified and explicitly articulated (McNamara *et al.*, 2018).

Other studies also question the validity of the CEFR descriptors and their impact in language tests and examinations (Fulcher, 2004, 2008a, 2008b; Fulcher & Davidson, 2007; Fulcher, Davidson, & Kemp, 2011). Some researchers view the framework as a "product of political forces" and not the outcome of an "academic argument" (McNamara, 2006, p.37) or "empirical validation" (Harsch, 2014, p.161), since the CEFR was "presented within the context of political and policy issues in Europe" (Fulcher, 2004, p.253). In the area of assessment and testing, the scales are said to be too general and insufficiently specific to be used for examination purposes (Davidson & Fulcher, 2007; Fulcher, 2004; Fulcher & Davidson, 2007; Fulcher, Davidson, & Kemp, 2011; Harsch, 2014; Weir, 2005).

Another concern regarding assessment and testing is that scales might be used as "a one-size-fits-all approach to measuring language ability" (Saville, 2010, p.7) when they should only be used as a guiding tool. Yet, though the validity of the CEFR descriptors has been questioned (Alderson 2007, p.660), the generally accepted viewpoint is that practitioners should continue to "reside in the CEFR

while researchers are constructing the poles underneath” (Hulstijn,, 2007, p.666).

Another area of contention is the fact that little attention has been given to using the CEFR with L2 writing versus other skills, possibly due to the difficulties with the writing descriptors. Lowie, Haines and Jansma (2010, p.153) noted that “the CEFR levels are described in general terms, using phrasing that leaves much room for interpretation by the individual assessors”. These scales also do not “represent all aspects of written production that may be relevant in defining proficiency” (CoE, 2001, p.61). In addition, the levels in the writing scales are “very broad and thus academic writing may vary in terms of its linguistic quality even within a certain overall band” (*ibid.*, p.154). Furthermore, it is not always easy to find ‘illustrative samples’ of the same level and domain when teachers want to apply CEFR to students’ written work (*ibid.*, p.154).

Based on the criticisms that have been levelled at the CEFR, it is important that it is utilized with care if users are not to misunderstand or misuse the framework (Saville, 2010). As noted by North, “The CEFR as a whole and the CEFR descriptors in particular should be considered to be a learner model” (2014, p.23). North suggests that the CEFR descriptors should also be treated as “a simplified description of selected aspects of infinite varieties of skills and knowledge that characterise real students” (Mislevy 1995, p.343 cited in North, 2014, p.23).

North (2000, 2007) provided a detailed description of how the scales were developed and empirically validated following extensive qualitative research. He argued that the validation process differed from the traditional quantitative process conducted in language research (i.e. positivist point of view), but followed a different qualitative validation process, basing it on practising teachers’ judgements (North, 2000).

The CEFR helps with understanding skills and knowledge at different levels of language proficiency, while respecting and acknowledging differences among learners (Saville, 2010). Therefore, learners need to progress not only vertically, based on the CEFR scales, but also horizontally (Little, 2006). There is also a need

to take into consideration not just learners' language levels (Green, 2012) but also how well they "can do" the different elements of language. It is also important to emphasize that CEFR is used "to promote profiling and not leveling" (North, 2014, p.13). This means that it is not necessary for a learner to be at a specific level in one skill (B1 for example) and to be in the same exact level in all the other skills; it depends on their particular language learning needs and how, and in what context, they intend to use the language.

As stated by Little (2006, p.187), "it is possible to use the same 'can-do' descriptor to identify a learning target, shape the learning/teaching process, and guide the assessment of learning outcomes". The CEFR also helps and encourages practitioners "to reflect on their current practice, particularly in relation to learners' practical language learning needs, the setting of suitable objectives and the tracking of learner progress" (North, 2006). However, users may need to add features relevant to their context which are not mentioned in the CEFR (Cambridge ESOL, 2011, p.12). It is better to view it as a "general theoretical framework that needs to be supplemented with language-specific and context-specific descriptors" (Deygers, Zeidler, Vilcu, & Carlsen, 2017, p.13).

The application of the CEFR framework to all languages is also fraught with difficulty, as each language has its own lexico-grammatical means (Callies *et al.*, 2013). It is said that "those statements are often too global and underspecified, and thus, of limited practical use for language assessors... [and therefore] there is a growing awareness among researchers of the need to specify the CEFR framework by developing more explicit descriptors anchored in language use" (Callies *et al.*, 2013, p.7; Hawkins & Buttery, 2010).

Another limitation in the CEFR framework is that "[it] is also lacking with respect to the register of academic writing" (Neff *et al.*, 2008). "The current CEFR descriptors for writing proficiency have not been empirically validated, nor do they claim to represent all aspects of written production that may be relevant in defining proficiency. In fact, they have largely been reconstructed from scales describing different skills entirely" (CoE, 2001, p.61, Callies *et al.*, 2013, p.7).

While the shortcomings of using the CEFR are thus acknowledged, we cannot ignore the body of research that has already provided valuable insights into the use of the CEFR in different domains, including the identification of students' language proficiency.

2.8. Reasons for using the CEFR in this study

There are three principal reasons for the use of the CEFR in the current study, namely:

(1) the way in which the CEFR is designed; (2) the important role that the CEFR plays as a '*common language*' via which to understand and measure proficiency; and (3) its relevance to the context of the study, namely the PYP.

2.8.1. The way in which the CEFR is designed

The CEFR "can be presented and exploited in a number of different formats, in varying degrees of detail" (CoE, 2001, p.36). The descriptors correspond well with the communicative teaching paradigm (Green, 2012). The CEFR descriptors can "specify learning objectives in terms of situation, activities, functions and notions" (Green, 2012, p.21); and each descriptor "is worded in positive terms, even for lower levels" (North, 2014, p.55). With the ceiling effect of students' writing scores at the end of the PYP, identifying writing proficiency using the CEFR scales can be a useful alternative to gain insight into students' actual and required levels of written proficiency.

2.8.2. The important role the CEFR plays as a common language

The CEFR can be used as a *common language* to “foster mutual understanding” across different users (Tannenbaum & Wylie, 2005, p.41). Alhawsawi (2013) notes that its use is crucial when following international curricula in Saudi Arabia, and particularly for MHCCs (Al-Shehri *et al.*, 2013). Having an international “mutual understanding” of students’ proficiency levels in English is important in light of “the adoption of many current international trends” in Saudi medical education and curricula (*ibid.*, p.141).

It can be used as a reference tool for identifying learners’ needs prior to designing the curriculum (North, 2006; Little, 2007), and as “a point of departure” (North, 2014) to start the reflection, analysis and discussion of potential university standards and admission criteria (Harsch, 2018).

2.8.3. CEFR and its relevance to the current study

The CEFR is already used at the PYP, in its curriculum document to identify the programme’s objectives and to choose textbooks for each of the PYP levels. The PYP follows a communicative approach to teaching English, which complements that of the CEFR. The CEFR can therefore be used as a *common language* between me (the researcher) and the PYP and university stakeholders.

Therefore, in this study, the CEFR is used to identify both (1) the proficiency levels students have reached at the end of the PYP, as perceived by the study participants, and (2) the levels they *need to* have reached in order to be adequately prepared for the colleges requirements.

The next section considers the various studies conducted in NA specifically considering medical students, including those in the Saudi Arabian, Middle Eastern and international contexts. There is an evident gap in the literature around the use of the CEFR framework in NA studies and is almost non-existent regarding medical students. The implication of this gap for the context of this study will be explored in section 2.10.

2.9. NA for MHCC students within the Saudi contexts

Although it has been acknowledged that the first step for a successful English programme is the identification and consideration of learners' needs (Alfehaid, 2014), many language programmes omit proper NA (Al-Tamimi & Shuib, 2010), and few studies have utilized NA to examine students' needs in Saudi Arabia (Ghobain, 2014). Of those that have, some focused on science (Alhojailan, 2015) or engineering students (Hellmann, 2013). There are very few NA studies in Saudi Arabia with a medical/healthcare focus. Of these, some have focused on English needs in medical workplaces (Alharby, 2005; Ghobain, 2014) rather than on students' needs in medical colleges, and most were performed prior to the implementation of PYP.

Among the studies that were conducted prior to the implementation of the PYPs in Saudi universities is Al-Ghamdi's (2006), which used target and present NA to consider the views of stakeholders on the products and processes of the Medical English course for medical students, including objectives, methodology, effectiveness and appropriateness. He found that the EAP course was effective and successful on the whole, but needed further improvements to meet the students' needs.

Al-Eissa (2008) focused on learners' needs in their current situation (at medical colleges) and their future workplace as perceived by educators at King Abdul-Aziz University Medical Colleges, with reference to how these perceptions affected the course design and selection of materials. He investigated learners' needs from an academic and professional perspective, focusing on reading, writing and conversational skills. He reported that writing is the primary skill required at university level, whereas speaking, with a focus on fluency and pronunciation, was more important in the medical professional workplace (e.g. in hospitals). It was thought that language course designers were not aware of learners' needs due to their unfamiliarity with the medical courses and the requirements of future medical workplaces.

Shukri (2008) investigated first-year medical students' perceptions of their writing needs, including *necessity*, *lacks* and *wants*, using a mixed-methods

approach (semi-structured interviews and questionnaires) to collect data from different stakeholders. Writing was a primary need for tertiary-level students in the medical and healthcare colleges in Saudi universities. Writing practice and more focus on grammar were the students' *'wants'*, while vocabulary, spelling accuracy and applying grammar in writing were their *'lacks'*. High expectations were found among the teachers regarding the students' *'necessities'* when the latter were still struggling with the basics of language proficiency.

After the implementation of the PYP, studies targeting medical students' needs focused on their needs in workplaces rather than at tertiary level at university. For example, Ghobain (2014) studied medical workplaces to investigate students' and practitioners' needs, attitudes and motivation in using English. She found that there was a shift in the requirements of spoken English in the medical workplace. With the increasing influx of Saudis in medical workplaces, there has been a reduction in the need for 'nativisation' and the issue of possessing a native-speaker accent was seen as a secondary requirement compared to other needs.

Alqurashi (2016) explored the English language needs of Saudi medical students and fellowship doctors enrolled at medical programmes in Australia. The study identified students' need to improve more language-related aspects like fluency and accuracy of structure. The study also found that "there is a critical need to structure English for medical purposes programs in Saudi Arabia to make better course design, content, and materials responsive to target language learners' own future goals" (p.243).

As can be seen, there is a clear gap in the literature with regard to medical students' English needs at tertiary level, especially in writing. More specifically, none of the above studies, to my knowledge, has focused on the identification of medical students' proficiency levels achieved at the PYP and required in the first-year in MHCCs. Therefore, this study will contribute to the sparse literature in this area.

2.10. Implications for the current approach to NA

This literature review has discussed the multifaceted approaches to NA, the different perspectives on language proficiency, the various uses of self-assessment and the CEFR scales and descriptors in language teaching, learning and assessment that have shaped the design of the present study. It also summarized NA studies related to the study context in order to identify what is lacking and to understand what is happening in terms of writing in the context of the present study. The purpose of the current study is not to test or validate the CEFR scales or to confirm the applicability of any of the language proficiency models reviewed in the literature. It is exploratory in nature and aims to describe the students' proficiency, in relation to the study context, based on the CEFR framework and to provide a description of their proficiency (both achieved and required). The literature reviewed in this chapter has informed the NA approach taken to pursuing the three overall aims of the research: firstly, to explore students' writing proficiency and the CEFR levels achieved by the end of the PYP, prior to enrolling at medical colleges; secondly, to explore students' writing proficiency and the CEFR levels required of MHCC students in their first year at college; and thirdly, to explore the misalignment between students' proficiency and levels achieved and the ones required, and in doing so to identify the gaps in the writing curriculum at the PYP. By drawing on the literature reviewed in this chapter, the following assumptions are made:

- The CEFR, which is based on theoretical developments in applied linguistics concerning the nature of language proficiency, can be used as a tool to identify learners' language proficiency (both required and achieved).
- The CEFR scales can be used as a reference tool to empirically conduct systematic NAs (Harsch, 2018); i.e. gap NA in the current context.
- The CEFR has limitations, such as focusing only on general functions of the language. For this reason, it is important to carry out in-depth analysis (thick description) (Geertz, 1973; Huhta *et al.*, 2013) of language proficiency in a specific context. The current study, therefore, employs both statistical quantitative methods to identify the CEFR levels achieved

(in PYP) and required (in MHCCs), and to postulate the gaps between them, as well as qualitatively exploring the relationship between participants' perceived and actual proficiency, and analyses qualitative data collected from different stakeholder groups, exploring their views and perceptions of writing abilities and requirements across PYP levels and contexts (i.e. PYP and MHCCs).

- Self-assessment can be used as a reliable measure of proficiency if certain factors are taken into consideration. This includes, but is not limited to, proper design of the tools, the acknowledgement of students' proficiency, providing incentives that are suitable to the study participants, and (most importantly) combining self-assessment results with other measures such as test scores and/or teachers' assessments.
- The CEFR and its descriptors are not meant to be utilized in any one specific way, and, as suggested by its proponents, the CEFR is merely a guide to help us navigate through learners' language levels and proficiency. The CEFR scales, if adapted with care and piloted, can be used for different purposes, such as self-assessment or the identification of expectations and standards. The literature reviewed also indicates the possibility of using the CEFR in contexts outside Europe (Saudi Arabia in this case).

2.11. Defined research aims

The literature review informs the approach to exploring the three overall aims of the study:

- 1- Exploring PYP students' CEFR levels and achievement in writing proficiency.

Even though the CEFR has been introduced to some PYPs in Saudi universities as part of their curriculum framework to reference entry and exit levels, to the best of my knowledge, no study has hitherto explored or examined its different uses in Saudi contexts.

2- Exploring first-year MHCC students' CEFR levels and the writing proficiency required of them.

Few NA studies on first-year medical students' writing proficiency have been conducted in the context of Saudi Arabian universities, and most found that students join MHCCs with low proficiency, which does not meet the colleges' requirements. However, there is no clear definition of what 'low proficiency' means. For example, Ghobain (2014) stated that medical students "must have an advanced level of English proficiency in order to properly understand the medical field" (p.154). Alhawsawi (2013, p.144) indicated the importance that "the English language proficiency it offers in its EFL programme meets a high standard to ensure students to communicate in English, and that it responds to the students' academic needs and provides the desired competency". Shukri (2008) suggested that "the target proficiency should be matched against the learners' existing proficiency" (p.25), and the only definition she gave of advanced-level students is that they are those "who already have acquired basic reading and writing skills to become aware of the different kinds of [writing] genres" (p.16).

3- Comparing Phases I and II to identify misalignments between achieved and required CEFR levels and writing proficiency, identify gaps in the PYP writing course.

Dissatisfaction with students' proficiency and their failure to demonstrate conversancy in the academic literacies expected in medical colleges even after the implementation of the PYP is a concern. It is very important to spot the gap(s) (if any) in the PYP English writing course to understand this problem and recommend solutions. As stated by Bawazeer (2015), "In order to improve from the current situation to the desired one, one needs to take steps that acknowledge and address the current situation and build on it rather than completely ignore its existence" (Bawazeer, 2015, p.32). Therefore, to identify and understand the gaps in the PYP writing course, I need to understand and specify students' writing proficiency at the end of the PYP ('the current situation') and compare it to the proficiency required in the MHCCs ('the desired/target situation'). I decided to follow Brown's (2016) gap analysis approach. According to Robinson (1991), this process is used to combine "target situation and present situation analyses" (p.9).

This would allow the needs demanded and required of first-year students in the MHCCs to be compared with the students' identified current abilities "in order to function effectively in the target situation" (Hutchinson & Waters, 1987, p.55) and enable us to identify, improve and rectify the gaps in the PYP writing curriculum (Royse *et al.*, 2009).

Though there are different tools that might be suitable to conduct NA for the current study, using the CEFR scales was the most suitable for the current context for the following reasons. First, the CEFR has already been introduced to the PYP curriculum. PYP created their curriculum framework without empirical data, based on the CEFR levels, and tracks students' progress based on the CEFR throughout the year. Second, though the entrance and exit CEFR levels for each of the PYP levels (elementary, intermediate and advanced) are stated in the PYP curriculum framework, the PYP exit standards and the English language entry requirements (and, in particular, writing requirements) are not clearly defined in PYP policy nor clearly articulated in the admissions policies of the medical colleges. A ceiling effect was also found in students' final writing exams, and it is insufficient to rely on the students' scores in the PYP as indicators of students' proficiency in writing.

Due to the constraints related to the policy at the PYP, the option of giving the students a writing test benchmarked against the CEFR to measure their proficiency was not available. Therefore, using the CEFR for both students' self-assessment of levels achieved and for identifying the proficiency required for the first year of Medical College was found to be the only reasonable option available. Acknowledging the limitations as identified in the literature, I tried to maintain the reliability and the validity of the data collected (see Chapter 3).

In my NA, I purposely chose to collect the data at the end of the academic year to give students the opportunity to engage with the language throughout the year, acquire a clearer understanding and better awareness of what is expected of them, and provide more comprehensive and accurate data, accordingly. This is important because, as stated by Robinson (1991), "It is vital to make students more aware of language and professional needs...as students become more involved with the course, their attitudes and approach may change" (p.102).

Defining language proficiency in the context of Saudi universities not only enables a common language to be used with different participants and stakeholders, both nationally and internationally, it can also be used as a starting point for establishing specific language criteria for admission. Moreover, instead of deciding on those criteria intuitively, the systematic and thorough analysis conducted in the current study promises to provide a better informed, more relevant basis on which to establish them.

In the literature review, it was found that NAs in the Saudi Arabian context have mainly focused on identifying language needs for medical students in the workplace and in medical professional fields, and studies were performed prior to the implementation of the PYP when language courses were integrated with the curriculum for the first year in college. Another interesting finding is that, although writing is an important skill to be considered in the medical colleges (Al-Eissa, 2008), to the best of my knowledge there has yet to be a study that focuses on medical students' writing proficiency needs in their first-year of college except for Shukri's (2008; 2014); nor has there been any study conducted in which the CEFR is utilized for this purpose.

In addition, I have yet to find any study exploring misalignments between what is currently being offered in PYPs and what needs to be offered to prepare students for their academic studies. To the best of my knowledge, none of the studies followed gap analysis for their NA, or used longitudinal follow-up of participants, as performed in this study. My study, therefore, adds to the sparse NA literature on first-year medical students, with the aim of improving the PYP curriculum and thereby better preparing students to meet their writing requirements at medical colleges. Most importantly, identifying the needs around the CEFR will allow for a *common language* that can be used nationally and internationally in discussions pertaining to students' needs and language proficiency levels in the medical and healthcare academic field.

Chapter 3

Research Methodology

3.1. Introduction

This chapter presents the study methodology, including the purpose, research questions (RQs), paradigmatic position, study settings, selected participants, and design, piloting, validity and reliability of the instruments used. This is followed by a description of the two data collection phases of this study, ethical considerations, data entry and analysis, and the limitations and constraints of the study.

3.2. Research purpose and research questions

The main aim of the study is to identify the gaps between students' writing proficiency (i.e. the CEFR levels and the writing skills perceived achieved) at the PYP and what is actually required by students as they progress to first year in various MHCCs. Identifying gaps in the writing curriculum should allow identification and establishment of a set of minimum standards, focussing on students' requirements in the first year of MHCCs at the university. These standards are significant not only for PYP stakeholders but also stakeholders in other universities in Saudi Arabia who are interested in students' writing needs.

Phase I identifies the PYP students' writing abilities in terms of the CEFR proficiency levels, perceived as achieved, including writing skills and writing-related problems. Phase II explores students' writing skills required including the CEFR levels in their first year at the MHCCs.

Phase III explores the misalignments between phases I and II and examines the situations and problems associated with writing which led to such gaps in the PYP curriculum.

With these aims in mind, this study seeks to answer the following questions:

Overarching question

Does the PYP curriculum adequately prepare students to meet the writing requirements of the university Medical and Healthcare Colleges (MHCCs)?

Secondary research questions

Phase I:

1. What are the students' CEFR writing levels at the end of the PYP, as assessed by the students, their tutors and trained English language raters?
2. What writing skills do students, tutors and coordinators perceive students graduating from the PYP to have developed?

Phase II:

3. What are the CEFR levels required of first-year medical students to cope effectively with the colleges' writing requirements, as perceived by MHCC students and academic staff?
4. What writing skills are required during the first year in college, as perceived by MHCC students and academic staff?

Phase III:

5. To what extent is there a misalignment between the students' CEFR writing levels achieved by the conclusion of the PYP and those perceived as required during the first year of MHCCs?
6. To what extent are there misalignments between the writing skills perceived as having been achieved in Phase I and those perceived as required in Phase II?
7. To what cause(s) can gaps resulting from any misalignment be attributed?

3.3. Research approach

3.3.1. Research paradigm

Given the nature of the research questions, this study embraces a pragmatist paradigm, in which "quantitative and qualitative research can be meaningfully integrated" (Bryman, 2006, p.114); both qualitative and quantitative methods are useful (Teddlie & Tashakkori, 2009) as long as they answer my research

questions. Pragmatists believe that the decision regarding which method to use depends on 'the current statement' of the research questions (Teddlie & Tashakkori, 2009, p.87). As stated by Johnson and Onwuegbuzie "research approaches should be mixed in a way that offers the best opportunities for answering research questions" (2004, p.16).

My research builds on my "positivist" ontology (i.e. in terms of the nature of reality) on the other hand, where I believe there is "one single reality" which is "independent of our mind" (Snape & Spencer, 2003, p.16) and my "constructivist" believe on the other hand, that realities 'are multiple' and can be constructed by the world around us (Teddlie & Tashakkori, 2009, p. 85). It also draws on my view of epistemology (i.e. in terms of knowledge). I believe that knowledge can be subjectively constructed by the interactions which we perceive between the world and the individuals who inhabit it, based on reality and the experience gained from the world around us. It can also be understood objectively (Teddlie & Tashakkori, 2009). I believe that being objective and having statistical data is valuable and adds significantly to the nature of the research, and I also believe in the importance of having in-depth qualitative analysis of the context, since reality is context-constructed.

My research questions lend themselves to the use of a mixed-methods approach, especially since these questions "could not be answered in any other [better] way" (Tashakkori & Teddlie, 2003, p.x). These questions require quantitative analysis of the CEFR levels perceived as achieved and required, and the gaps between these, as well as statistical comparisons between self-assessments, tutor evaluations and independent ratings of texts. In addition, qualitative analyses can enhance the understanding of participants' views on writing proficiency, writing difficulties, preparedness for MHCC, satisfaction with PYP and the reasons for gaps. Thus the mixed-methods approach is appropriate.

This study is not concerned with the statistical generalizability of the findings to a global context, but concerns a context- and cohort-specific understanding of writing requirements. However, the findings may still be generalizable to similar contexts in other universities in Saudi Arabia. Nonetheless, what can be generalized outside the specific context from the current study are the methods

used to conduct a gap analysis of NA using the CEFR scales in a mixed-methods approach with a longitudinal aspect.

3.3.2. Research methodology

Three interconnected objectives need to be addressed to answer the study research questions. First, it is important to evaluate the PYP's writing programme to explore "what is happening in the organization that [may] suggest there is a need to change the performance/behaviour of individual(s)" (Bee & Bee, 2003, p.140) This could be described as 'present situation analysis' (Brown, 2016). The second objective is to identify the writing skills students are required to possess in order to cope with the requirements of students studying in the MHCCs by exploring the colleges' writing requirements, a 'target situation analysis' (Brown, 2016). The third is to "identify the gap" between the "required performance" (students' writing requirements at the MHCCs) and the students' "current level of performance" (Bee & Bee, 2003, p.140) and the reasons behind these gaps (Brown, 2016) (if any).

The exploratory and evaluative nature to this research led to the adoption of a mixed-methods approach to NA, combining qualitative and quantitative research in a longitudinal approach. Phase I explores PYP students' current writing performance, thereby addressing RQs 1 and 2, whilst Phase II explores the first-year writing requirements of students enrolled in the university's MHCCs addressing (RQs 3 and 4). Any misalignments identified between these two phases then address RQs 5, 6 and 7 leading to answer the overarching question regarding the level of adequacy of the PYP in preparing students to meet the writing requirements of the university MHCCs.

The quantitative approach is useful to statistically identify and compare the CEFR levels which the students have achieved (Phase I) and ought to achieve to cope with the writing demands of the first year at college (Phase II).

The CEFR scales can be useful in quantitatively identifying students' perceptions of their language proficiency (Atai & Shoja, 2011) and to invite students and

tutors to qualitatively reflect on the students' language skills and contextual requirements (Luoma & Tarnanen, 2003). However, CEFR descriptors are neither language- nor context-specific (Little, 2007; North, 2006) but need to be contextualised. Therefore, using qualitative approaches can "increase the scope, depth and power of research" (Punch, 1998, p.243). In this study, qualitative methods are used to gain further insights into the students' current and required context- and language-specific writing proficiency, their needs and limitations, and additional writing skills which may not be covered by the CEFR scales (Cambridge ESOL, 2011). Identifying CEFR levels (achieved and required) alone would provide a limited insight into what students *can* do or are *required* to do within those identified levels in my specific context and setting. Supplementing quantitative with qualitative methods promises to increase the level of detail elicited in both phases of the research, and provides an alternative means through which participants can talk about their experience of, and proficiency in, writing.

3.4. Research design

This study explores the students' writing from the perspective of different stakeholders using a mixed-methods approach with some longitudinal aspects. It follows a convergent parallel design (Creswell & Plano-Clark, 2011). This design was considered beneficial to this study which aims to gather varied but related data on a common topic in order to understand and explore the gaps in the writing programme at the PYP (Morse, 1991, p.122) following a gap analysis approach to NA (Brown, 2016). The same design was applied in the two phases of the study. In each phase, the research questions, instruments, data collection and data analysis techniques have quantitative and qualitative dimensions (Teddlie & Tashakkori, 2009). The collection of quantitative and qualitative data in each phase is "concurrent but separate" (Creswell & Plano-Clark, 2011, p.78) and they carry equal importance for addressing and answering the study's research questions (Creswell & Plano-Clark, 2011). According to Alderson *et al.* (1995), using quantitative and qualitative approaches is of benefit to NA and

helps avoiding bias in the results gained that might be caused if one approach alone was used in NA studies.

In Phase I, quantitative data were collected towards the end of the PYP year via two questionnaires: (1) a CEFR-based self-assessment questionnaire for PYP MT students, and (2) a CEFR-based tutors' assessment questionnaire. Focus groups (FGs) conducted with PYP MT students, PYP tutors and PYP coordinators were one of the main methods used to collect qualitative data in this phase. In addition to the FGs, a *letter to a friend* written by the students, in which they wrote about their writing skills at the PYP, constituted a further dataset. This letter was requested before working on the self-assessment questionnaire.

The students who participated in Phase I were tracked in Phase II one year later, following their initial academic year in the MHCCs. Both the first-year college students and their teaching staff completed CEFR-based questionnaires as part of the quantitative data in Phase II. Towards the end of the academic year, participants in this phase, were asked to specify the levels they felt were minimally required to cope in that academic year. FGs with the students and interviews with the academic staff formed the qualitative part of this phase.

Figure 3.1 (adapted from Teddlie and Tashakkori, 2009, p.152 and Creswell & Plano-Clark, 2011, p.79) shows the parallel quantitative and qualitative strands of the study in Phases I and II, which were subsequently brought together in Phase III.

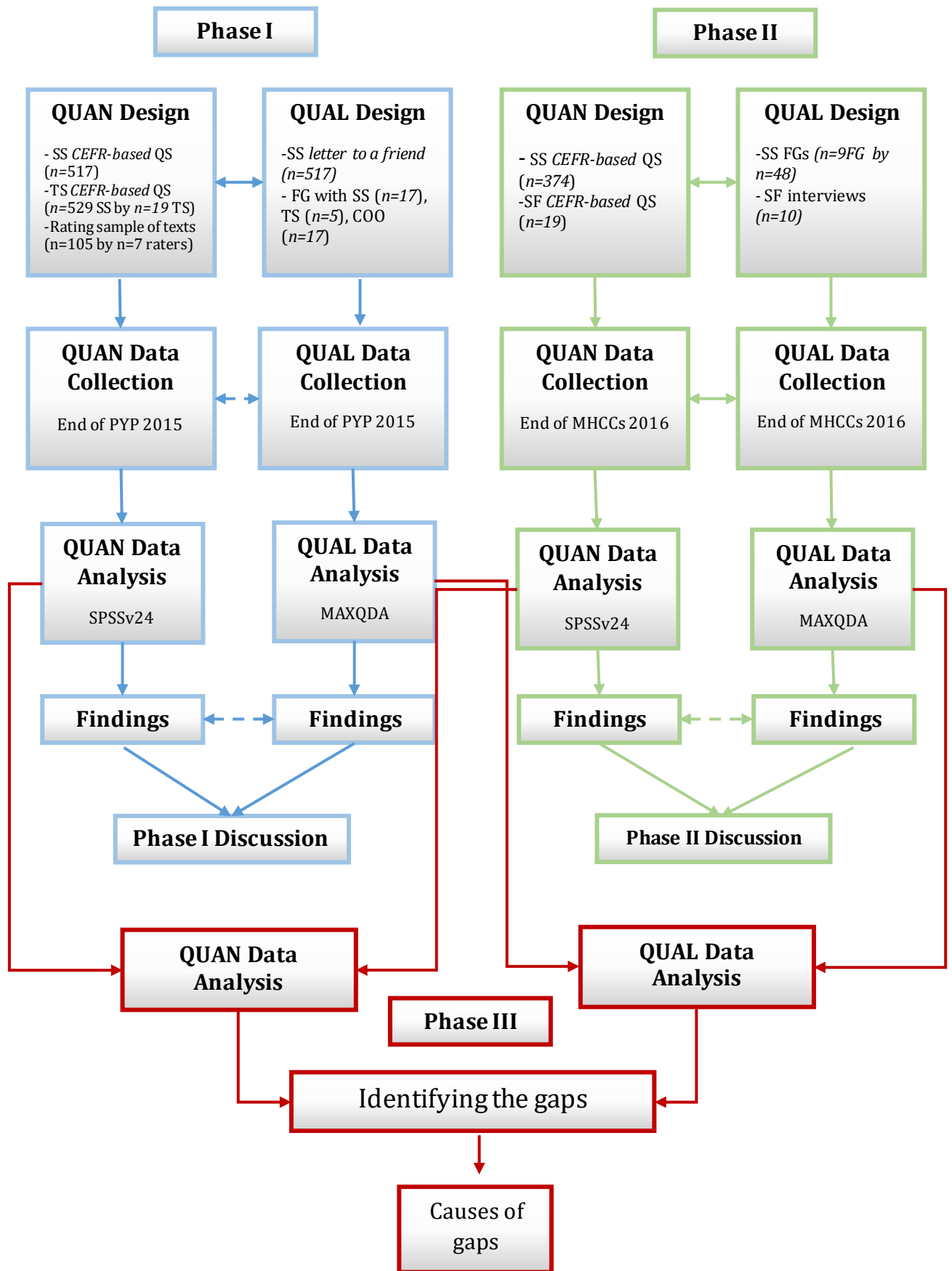


Figure 3. 1 Flowchart illustrating the convergent parallel design of the study across the three Phases.

The two parallel and independent quantitative and qualitative strands were “planned and implemented to answer related aspects of the same overarching mixed-methods research question” (Teddlie & Tashakkori 2009, p.152). Then, inferences based on the results from each strand were “integrated to form meta-inference at the end of the study” (Teddlie & Tashakkori 2009, p.152).

This type of parallel design allows for the gathering of “different but complementary data on the same topic” (Creswell & Plano-Clark, 2011, p.77). It also provides a more comprehensive account of participants’ views and opinions on the different issues related to the study. This is because combining quantitative and qualitative approaches leads to better “captur[ing] of the trends and details of a situation” (Ivankova, Creswell & Stick, 2006, p.3).

3.4.1. Research settings and participants

The settings for data collection differed in each phase. For Phase I, the study was conducted at the ELSD in one of the oldest, most reputable Saudi universities between April and May 2015, after students had attended one year of an intensive English programme at the PYP.

One year later, the data for Phase II were collected at the main campus of the same university in five different MHCCs. These included *CM, the CPH, CAMS, CN* and *CD*. Most students who participated in Phase I of the study had transitioned into these colleges based on their accumulative final GPA on the PYP. Data collection for Phase II took place between April and May 2016. In both phases, the rationale for collecting data towards the end of the academic year was that participants would probably be able to meet the various skills of writing required and understand the various needs relating to writing skills, and therefore offer greater insight into the topic at hand.

3.4.1.1. Participants in the quantitative strand

In Phase I, the whole female cohort of the MT students in PYP ($N=640$) was invited to participate in the CEFR-based self-assessment questionnaire. After checking and cleaning the data, a total of $n=517$ participated, giving a response rate of about 80% of overall participants across the three PYP levels (*elementary*, *intermediate* and *advanced*), 90% were Saudi and 10% non-Saudi, aged 18–19 years. All PYP tutors ($N=24$) teaching English to the students in the PYP MT were also invited to participate in the CEFR-based tutor questionnaire to assess their students' writing. A total of $n=19$ tutors accepted the invitation to participate. They assessed a total of $n=529$ students out of the $N=640$ across the three PYP Levels. The tutors were of different nationalities (native and non-native English speakers) with varying levels of experience. Table 3.1 provides further details of Phase I participants in the quantitative data collection process for each of the three PYP levels. Seven raters (see table 3.2 for more information), who are experienced English language teachers from two language institutes in the UK, agreed to participate in the study (in October 2017) by attending the benchmarking training session (prepared and presented by me) and then rating a random sample of the study participated students' written texts.

Table 3. 1 Number of students who undertook self-assessment and were assessed by their tutors in Phase I using CEFR-based questionnaires

PYP Levels	# of Students Self-assessed	# of Students Assessed by Tutors
Elementary (n=92)	73(79.3%)	90(97.8%)
Intermediate (n=324)	268(82.7%)	249(76.8%)
Advanced (n=224)	176(78.6%)	190(84.8%)
Total	517/640 (80.8%)	529/640 (82.6%)

Table 3.2 Raters' Biographical data

Raters	Years of experience assessing English L2 writing	Years of experience using CEFR scales	First language	Current role	Experience in test design
PN	35	15	English	Teaching fellow and director of language studies	Yes
MA	8	6	French	Language institute manager/ ELT teacher	Yes
MT	13	12	English	English language teacher	Yes
DIM	10	10	English	English language teacher	Yes
DV	6	6	English	Associate English Tutor	Yes
ID	13	6	Farsi	TESOL lecturer	Yes
AN	4	2	English	English Language teacher	No

One year later, following the students' enrolment in the MHCCs (Phase II), a total of 374 out of 568 students (66%) and 19 academic teaching staff participated across the MHCCs and completed a CEFR-based questionnaire (numbers and percentages participating shown in Table 3.2). There were 71 students (about 19%) who completed Phase II questionnaires but did not originally participate in Phase I.

Table 3.3 Number of students and teaching staff who participated in Phase II Questionnaires

Colleges	<u># of Students</u>	<u># of Academic Staff</u>
CM	85 (73%)	9
CPH	79 (79%)	5
CAMS	106 (48%)	0
CN	54 (87%)	5
CD	50 (71%)	0
Total	374 (66%)	19

3.4.1.2. Participants in the qualitative strand

Two sources of data were collected in Phase I of the study: (1) students' written letters to an imaginary friend about their writing experience in the PYP ($n=517$)

(a letter to a friend) and (2) FGs with different stakeholders in the PYP. A total of 17 students from the intermediate level participated in three FGs and 5 tutors participated in a further two FGs. In addition, four coordination units (17 coordinators) were asked to take part in FGs made up of 4 coordinators from the Curriculum Team, 5 from the Assessment Team, 5 from the Continuous Assessment Team and 3 from the Professional Development Team.

During Phase II, qualitative data were primarily obtained from FGs with first-year MHCC students. It was intended that the FGs should involve both students and staff members (in separate FGs); however, due to staff's time constraints and different teaching timetables, individual interviews with the academic staff were conducted instead. Table 3.3 summarizes the number of participants in the qualitative part of Phase II. Figure 3.2 below summarises the study participants in both Phases.

Table 3.4 Number of students and teaching staff who participated in Phase II FGs/interviews

Colleges	<u># of Students (# of FGs)</u>	<u># of Academic Staff</u>
CM	17 (3)	2
CPH	12 (2)	3
CAMS	9 (2)	no participants
CN	10 (2)	5
CD	no participants	no participants
Total	48 (9)	10

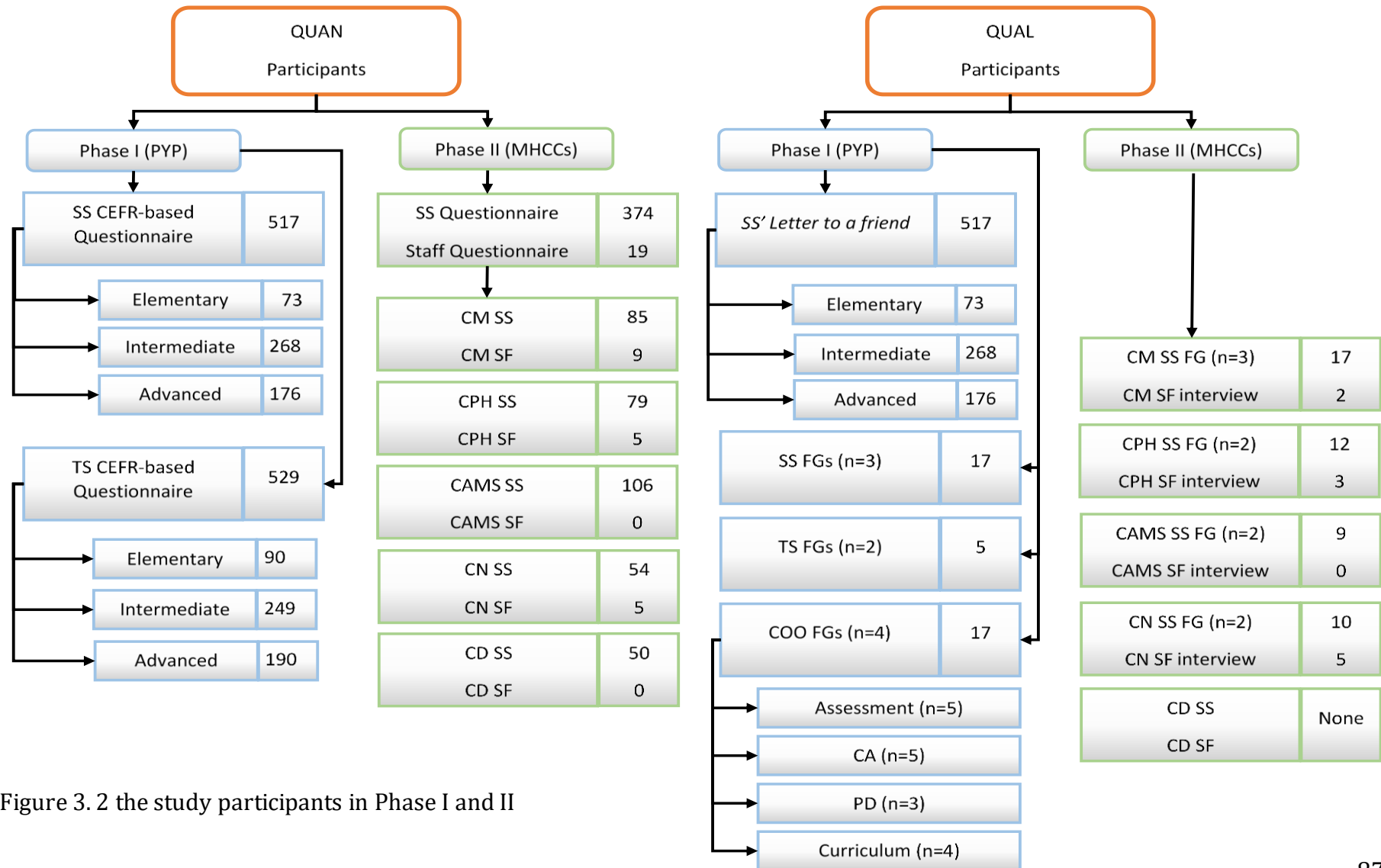


Figure 3. 2 the study participants in Phase I and II

3.5. Research instruments

In this section, a detailed description of the different instruments used for data collection is presented, including development, piloting and establishing reliability and validity.

3.5.1. Quantitative instruments

During Phase I, quantitative data involved the use of the CEFR-based questionnaire designed for self-assessment by students through 10 CEFR Scales (see Part Two of the students' questionnaire in Appendix A1). This questionnaire was translated into Arabic and distributed to the students (available upon request). A second questionnaire was designed in English for PYP tutors to assess their students using the same CEFR scales (available upon request).

In Phase II, both students and staff answered a CEFR-based questionnaire to choose the levels which reflect the descriptors of the skills that they would need rather than their achieved levels (See Appendix A2). Again, the staff questionnaire differed only slightly from the students' questionnaire to reflect the different participants. Students and Arab staff were given the questionnaires in Arabic (available upon request). The next section describes the design and reasoning behind the use of CEFR scales as employed in the study questionnaires.

3.5.1.1. Choosing CEFR scales relevant for the study context

There are 53 CEFR scales representing different language skills and these must "be used selectively" (North, 2014, p.11) to suit the context in which they are employed. For this reason, I used the following procedure to identify the scales that were most suited to the context of this research.

First, I listed all the CEFR scales related to the skill of writing in the CEFR (CoE, 2001). Their relevance (face validity) to this study's context was then checked with three colleagues, two of whom were teachers on the PYP and the other a member of academic staff working in one of the medical colleges.

The decision was made, for example, to exclude the *correspondence* scale (CoE, 2001, p.83) and the *creative writing* scale (CoE, 2001, p.62) as they are not related to the study context. On the other hand, the writing scales from the DIALANG project (CoE, 2001, p.240-241) were added, since they were considered relevant. This was an important step as it helped ensure that the study did not include irrelevant scales or exclude relevant ones.

After designing the questionnaire and before piloting, more feedback was sought from the same teachers and from colleagues from the applied linguistic field. Based on this feedback, further scales were either eliminated or combined. For example, some scales shared the same descriptors (e.g. the descriptors in the *self-assessment grid*, p.26-27) appeared in other scales as well), and thus were excluded. The vocabulary scales (*vocabulary range* and *Vocabulary control*, p.112) were combined to minimize the number of scales used in the study, to reduce the burden on participants (Faez *et al.*, 2011b) and therefore increase the likelihood of their engagement in the assessment and providing more accurate answers. Equally, however, there was a need to ensure that the maximum number of relevant writing CEFR scales were covered to gather a more complete picture of the students' writing levels and needs. A balance, therefore, needed to be obtained.

The following are the CEFR scales (CoE, 2001) that were selected for inclusion in the questionnaires:

- 1) Overall written production (p.61).
- 2) Overall written interaction (p. 83).
- 3) Types of texts the students can write (p.240-241).
- 4) What the students can write (p.240-241).
- 5) Vocabulary range and control (p.112).

- 6) Grammatical accuracy (p.114).
- 7) Orthographic control (p.118).
- 8) Processing texts (p.96).
- 9) Reports and essays (p.62).
- 10) Note taking (p.96).

3.5.1.2. Phase I questionnaires

The Phase I student questionnaire comprised three main parts (see Appendix A1). The first focused on qualitative data, described in Section 3.5.2.1 below (Qualitative Instruments). The second focused on quantitative data using the CEFR scales and required students to use the ten CEFR scales listed above to self-assess their own levels in writing. The third gathered demographic information on participants.

The self-assessment part of the questionnaire (see Appendix A1, Part Two) was designed as a grid. Each of the ten CEFR scales was placed in one row in the grid. The CEFR descriptors for each scale were lined up in columns representing the CEFR levels. Some CEFR scales included the ‘plus levels’ while others did not, hence empty boxes were left in some columns. ‘Plus levels’ refer to additional levels “between each of the main levels (A2+, B1+, B2+)” (North, 2014, p. 71) and indicate an increased level of proficiency (a strong level) within a particular band (i.e. A2+ indicates a slightly stronger performance than A2). The numbers in the grid refer to the CEFR levels (Table 3.5).

Table 3.5. The distributions of the CEFR in the self-assessment grid and the numbering system/coding used

1	2	3	4	5	6	7	8	9
A1	A2	A2 +	B1	B1+	B2	B2+	C1	C2

The descriptors in each scale were restated in '*I Can*' statements, thereby bringing the "self-assessment into a much closer relation" to its users than when following other procedures, according to Little (2006, p.185). However, the wordings of the descriptors remained the same. Under each descriptor (in each column), there were two options from which students were required to choose, either *Yes I Can do* or *Not Sure*. Similar to Ashton's argument (2014), providing a third option (e.g. cannot do) would make the analysis more complex and difficult to interpret and was thus avoided. I chose the '*not sure*' option to allowing space for doubts regarding their abilities, since "[i]f somebody says that they can do something sometimes, does that mean they are at a given level or not?" (Alderson, 2005, pp.211-212). When the students choose '*Yes I can*', this, in my view (by adopting a more 'conservative' approach), indicates that students are most probably confident enough to complete what is designated in that chosen descriptor. This means, in my opinion, that being at a certain level indicates the students have achieved mastery of all the skills mentioned within that specific level – which they ticked as '*Yes I Can*' – before they can be deemed to be at that particular level. Conversely, when they choose '*not sure*', two possible assumptions can be made: first, that the students cannot perform the skills mentioned in that specific level's descriptor; second, that they are in doubt, which suggests that they are not able to carry out what is designated in a specific descriptor. In both situations, we would not be confident in assigning that level to the student; hence, one level lower would be the level assigned to the student if they chose the '*not sure*' option.

Students had to read the descriptors starting from number 1. If they felt they *can do* what the descriptor described, they would tick (*I can do*) and move to the second descriptor, where they followed the same procedure. They would continue reading the descriptors in the same row of the same category until they reach a descriptor that they felt they were *not sure* they were capable of doing (either cannot do it or unsure). In this case, students would tick *not sure* and would not need to continue reading the remaining descriptors in that particular row. Instead, they would proceed to the next row (i.e. the following CEFR scale) and follow the same process. I called this procedure *Controlled Utilization of CEFR Descriptors*, where participants need to follow the instructions to perform the

assessment using the grid (see figure 3.3. below for an example of how it can be used).

	1	2	3	4	5	6	7	8	9
Overall written production	I can write simple isolated phrases and sentences.	I can write a series of simple phrases and sentences linked with simple connectors like 'and', 'but' and 'because'.		I can write straightforward connected texts on a range of familiar subjects within my interest, by linking a series of shorter discrete elements into a linear sequence.		I can write clear, detailed texts on a variety of subjects related to my field of interest, synthesising and evaluating information and arguments from a number of sources.		I can write clear, well-structured texts of complex subjects, underlining the relevant salient issues, expanding and supporting points of view at some length with subsidiary points, reasons and relevant examples, and rounding off with an appropriate conclusion.	I can write clear, smoothly flowing, complex texts in an appropriate and effective style and a logical structure which helps the reader to find significant points.
	<input checked="" type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input checked="" type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input checked="" type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure
Overall written Interaction	I can ask for or pass on personal details in written form.	I can write short, simple formulaic notes relating to matters in areas of immediate need.		I can convey information and ideas on abstract as well as concrete topics, check information and ask about or explain problems with reasonable precision. I can write personal letters and notes asking for or conveying simple information of immediate relevance, getting across the point I feel to be important.		I can express news and views effectively in writing, and relate to those of others.		I can express myself in writing with clarity and precision, relating to the addressee flexibly and effectively.	As C1
	<input checked="" type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure

Figure 3. 3 A Screenshot (as an example) of how the CEFR scales were used by participants

The purpose of the tutor questionnaire was to assess the writing skills of the same students. The same assessment grid was slightly amended so the “I can” statements for each descriptor were changed to “the student can”. Tutors used the same procedure to assess the students using the CEFR grid.

3.5.1.3. Piloting Phase I questionnaires

After designing the assessment grid, feedback was obtained from colleagues on its structure, design and use. The grid (CEFR scales) and descriptors were written in Arabic, taken from the Arabic translated version of the CEFR (CoE-Arabic

version, 2008). When both questionnaires (students' and tutors') were ready, they were piloted, including instructions on usage to minimise errors, to test the reliability of the developed instruments and to improve efficiency before applying them to the targeted sample. The students' questionnaire was piloted with three PYP ST classes ($n=67$ students) from different levels. The tutors' questionnaire was piloted with $n=3$ tutors in the same track with a total of $n=30$ students assessed by those tutors. In the actual study, tutors assessed each student in their class using the tutor questionnaire. However, in the piloting stage, each tutor was asked to assess the first ten students in their alphabetical class list.

After completing the questionnaire, students were asked to fill in an *Evaluation Form* (see Appendix A3), to gather further feedback about the quantitative and qualitative parts, including the CEFR grid.

Detailed instructions were provided to the participants in the questionnaire itself and on a separate instruction sheet on how to use the assessment grid. Once the pilot data were collected, entries which did not adhere to the instructions correctly were excluded, leaving 36 out of a total of 67 student questionnaires for analysis and 25 tutor questionnaires. Having 46% erroneously-completed student questionnaires in the piloting stage and from comments received from participants, the design and instructions were revised and modified. For example, some students commented that "*the instructions for using the questionnaire were not very clear*". Therefore, the questionnaire's instructions were revised to be clearer after the pilot run.

Further, reliability of the collected data was tested using Cronbach's alpha, which measures how closely related are a group of items that purport to measure the same aspect (e.g. individual questions making up a summary scale) (Cronbach, 1951). Alpha represents the average correlation between pairs of items and takes values between 0 and 1 (Field, 2009). Higher alpha scores indicate that the items are related and therefore likely to be measuring the same underlying construct. The 10 CEFR scales rendered a high degree of reliability for both students' self-assessment and tutors' assessment questionnaires, with Cronbach's alphas of 0.884 and 0.951 respectively. This indicates that the 10 items (10 CEFR scales)

measure the same construct, and Cronbach's alpha does not improve if any scale is deleted. This means that the 10 scales (items) can be treated as a one multi-item-scale and the average scores can be considered in the analysis (Bland & Altman, 1997).

The pilot run led to several adjustments being made to the instrument content and layout. For example, based on the results and the feedback received, a decision was taken to improve the design of the grid to include more visual interaction and guidance. In addition, an audio-video presentation guide was designed for students to demonstrate how to use the grid in the actual data collection round. Though a good number of the students had praised the use of the questionnaire (e.g. one commented "*I like the idea of your questionnaire, it is new, different and exciting*"), some found it "*long*" and "*boring*," so it was considered important to further motivate students to complete it. Consequently, a decision was taken to provide student participants with the option of receiving an individual report about their writing levels based on their self-assessment. This was a crucial step in encouraging them to spend time on the questionnaire and provide more accurate data.

3.5.1.4. Phase II questionnaires

The questionnaires in Phase II aimed (1) to identify the CEFR Levels that the students minimally required to perform well in their colleges; and (2) to identify the writing skills required of students (See Appendix A2). The questionnaire comprised background information on the participants. It required students to list the different writing activities and skills they had used/required during their first academic year at college and to indicate how well the PYP had prepared them for each of these skills and activities. The questionnaire also contained the CEFR grid, where students were asked to read each scale (on each row) and choose only one level (descriptor) that best captured the minimal required skills in their college. There were also open-ended questions where participants were

invited to record comments and feedback about the writing required and writing difficulties in the first year of college.

A very similar questionnaire was used with the teaching staff (available upon request), with minor changes to address the particular (teaching staff) audience. The biographic questions in the first part were also different. The questionnaires were initially checked by colleagues in Applied Linguistics and revised prior to being translated into Arabic and then back into English. They were then distributed for piloting in Arabic.

3.5.1.5. Piloting Phase II questionnaires

All students in their first academic year and teaching staff from a College of Medicine in a different university in Saudi Arabia were invited to participate in the Phase II pilot stage. A total of $n=38$ students and $n=9$ teaching staff participated. After screening the questionnaires and excluding those incorrectly completed, $n=32$ student questionnaires and $n=8$ staff questionnaires were considered for data analysis.

In the first part of the questionnaire, participants were asked to provide some biographic data. In the second part (*The Required Writing Skills*), participants were asked to list the writing skills and activities they were required to do in their first year at college and then rate how well they were prepared for these before joining the college. Because participants had to provide the different writing skills required in this part and then rate them, it was important to assess the practicality of this step in the questionnaire. When I examined students' responses to this part, the results were generally positive. In addition, certain trends and patterns were found in the responses of the students, which allow for the skills to be categorised effectively. The rationale for positioning this part of the questionnaire before the CEFR grid task was to prevent students from being influenced by skills mentioned in the CEFR grid.

In this part of the questionnaire, and as stated in O'Cathain and Thomas (2004), some qualitative and quantitative features were combined where participants

“write whatever they want in their own words, with little structure imposed by the researcher” (p. 4). Following this procedure for data collection would allow the participants the opportunity to “voice their opinion” instead of having a specific list of questions which “represent the researchers' agenda and minimise the “power balance between researcher and research participants” (O’Cathain and Thomas, 2004, p.2).

Regarding the CEFR grid used to identify the minimum CEFR level(s) required, a reliability analysis was run using Cronbach’s alpha and the results rendered an acceptable level of overall reliability (0.847). Alpha was not increased by removing any scale, which indicates high reliability. All the scales can be treated as a one multi-item-scale and the average measures can be considered (Bland & Altman, 1997). Regarding the staff questionnaire, only $n=9$ academic staff participated in piloting, which is a small sample to consider for reliability analysis (Yurdugül, 2008).

3.5.1.6. Phase I: independent ratings of exam texts

Independent raters used a CEFR-based scale to rate random samples of students’ texts written for their final standardised exam at the PYP (the same texts that had been marked by the PYP tutors which gave the ceiling effect mentioned previously). This CEFR-based scale was taken from the *Manual* (CoE, 2009, p.187) and used to compare with the results collected from the self- and tutors’ assessments. The *Manual’s* CEFR-based rating grid (2009) consists of six categories (six scales), from which I selected five: *Range*, *Coherence*, *Accuracy*, *Description* and *Overall* (see Appendix A4). *Argument* is the category that was not included in this study because it does not suit the task being assessed. Texts were rated on 9 levels (A1 to C2) as outlined in Table 3.4 for consistency and ease of comparison to coding used for self- and tutors assessment.

3.5.1.7. Preparation for, and piloting of, standardisation and benchmarking session for raters

In order to prepare the raters to rate the exam texts, several steps were followed. First, to maintain the validity and the reliability of the data collection tools and to prepare for a standardised benchmarking training session, representative samples from the actual data (CoE, 2009) were identified. To select these samples, I chose texts that represented different PYP levels ($n=20$ samples) (see Appendix A5 for some samples) to be rated by the piloting group. The piloting group for this stage consisted of $n=5$ raters (different from those ones participating in the main study ratings). They were English language teachers with experience in rating L2 students' written texts and showed interest in participating in piloting. The purpose of this stage was twofold: firstly, to pilot the rating scale (that would be used later with different raters in the actual data collection run) with similar texts to those that would be rated in the actual study. Secondly, it served to select samples to be used with the actual study raters in their training session. The samples selected for the training session would be the ones with the highest agreement on the CEFR level assigned by the five piloting raters. The samples with many discrepancies would be eliminated from use in the training session.

The piloting group was sent the 20 samples, the rating scale and an excel file where they could enter their ratings easily (see Appendix A6 for screenshots of the email sent to participants and the excel file received from one participant after completing the ratings). They were asked to comment on the use of the CEFR-based scale and their rationale for their selected levels. After receiving the ratings and the comments from the pilot participants, I compared their ratings and comments and chose only the samples with exact/high agreement among participants. $N=8$ samples out of 20 were then considered for use in the training session with the actual raters.

The second step was to prepare for the training session with study raters. Following the steps in the *Manual* (CoE, 2009), different materials were prepared for the standardisation and benchmarking session with the study raters. This is a crucial step to standardise the process of rating and to ensure that all raters are consistent (CoE, 2009).

I followed the steps stated in the *Manual* (CoE, 2009) to prepare for this benchmarking training session, since “the interpretations of [CEFR] levels in the project does reflect the common interpretation illustrated by the illustrative samples” (CoE, 2009, p.9). The session was divided into four main parts: (1) *Familiarisation with the CEFR scales* where the CEFR and its scales were presented. (2) *Familiarisation with the rating scale* where the rating scale was presented to participants to familiarise them with its format and descriptors. (3) *Rating samples* which have already been benchmarked against the CEFR-based scale in the pilot run. (4) In the last part of this session, participants were given *more written samples to benchmark* against the CEFR-based rating scale and to discuss their ratings and their rationale with other raters. Before commencing the training session with the actual raters, I piloted the materials and the presentation that I prepared with two fellow students studying for a PhD in Applied Linguistics who have experience in English language teaching and who gave their comments and suggestions which helped improve the materials further.

3.5.2. Qualitative instruments

The two sources of qualitative data in Phase I were: students’ *letter to a friend* and FG meetings. The three sources of qualitative data in Phase II were student FG meetings, individual interviews with staff members, and open-ended questions in participants’ questionnaires.

3.5.2.1. Students’ *letter to a friend*

In Phase I, the students were asked to write a letter to an imaginary friend joining the PYP the following year. A set of questions to guide the students on what they should include in their letter were provided (see Part One of the students' questionnaire, Appendix A1).

The purpose of the letter was to collect qualitative data that would provide insights into the writing situation in the PYP, including the skills that students perceived they had developed while studying in the PYP. In addition, they were asked to describe the writing skills students felt they had acquired (*can do*) or still needed to acquire (*cannot do*). The rationale for using this method was that it could provide additional "grounded data" that could be used as a new and "different lens" through which to investigate the current study, which focuses on the students' perceptions of their writing abilities (Abbasabady, 2009, p.84). Asking the students to write the letters using their own words encourages critical thinking and allows us to treat students as active participants (Benesch, 2001).² The instructions and questions included in the letter-writing task were checked and feedback was elicited from colleagues in Applied Linguistics, resulting in several refinements. The letter's instructions and questions were then translated into Arabic, after which the Arabic version was translated back into English, differences checked and instructions amended as necessary. Students were given the choice and allowed to write the letter either in Arabic or in English.

3.5.2.2. FG meetings

In Phase I, PYP students, tutors and coordinators from different units participated in the FGs.

Five main questioning routes, as suggested by Krueger & Casey (2000) and Litosseliti (2003), were prepared starting from the *opening question*, *introductory questions*, *transition questions*, then moving to the *key questions* and

² The use of the letter was also suggested by participants during *the European Association for Language Testing and Assessment (EALTA) Special Interest Group (SIG) Assessing Writing/English for Academic purposes EAP Meeting* in February 2015 at the University of Warwick, where the research plan of this study was presented.

finishing with the *ending questions* (see the FG interview guide in Appendix A7). The questions in the first and ending routes were general with a set of 'common questions' whereas the questions and the FG activities in the middle routes were following 'unstructured approaches' to allow for more data to emerge (Alderson, Brunfaut & Harding, 2014, p.204). Generally, the questions and the activities for all FG members were similar, to help with subsequent data analysis (Krueger & Casey, 2000). However, there were slight differences in some of the questions and activities to tailor them to the participant groups.

Additionally, the CEFR was employed as a reference tool for data collection throughout the FGs, because the framework aims to provide 'objective criteria for describing language proficiency [which] will facilitate the mutual recognition of [different] qualifications among participants' (CoE, 2001, p.1). Thus, I wanted to pinpoint the varied outlooks of participants regarding the current students' CEFR levels at the end of PYP, how well they were prepared and what else they still needed to do or learn concerning their English writing skills.

Participants in the FGs were asked to 'define and articulate' the specific writing skills, activities and knowledge they had in each of the categories in the CEFR scales. By doing this, the students' level(s) and descriptions of what they can do in each category were described using participants' own words related to their own contexts. The idea was that the students' language proficiency should be *described* and not *prescribed* (North, 2005). Therefore, English writing proficiency was described in a way that aims to lead to mutual understanding and recognition of the CEFR levels between different groups of PYP users.

During the FGs, some activities were used to collect data and promote discussion with participants. In the first activity, I cut up the descriptors of the six CEFR scales related to writing in the self-assessment grid (CoE, 2001, pp.26-27). The students and tutors (but not the coordinators) were asked to collect all the descriptors they felt the students can do on one side of a blank paper and the one they felt students cannot do on the other side. The rationale here was to identify the students' CEFR level based on the consensus of the participants and to check their justifications for why they feel the students can do certain descriptors but not others.

In the second activity, participants were handed cards with the categories used in the CEFR scales with some prompts related to each category (see Appendix A8 for a sample of these cards) to encourage them to reflect on the students' writing in terms of these skills. This is an effective way to "promote discussion and generate data" (Hennink, 2014, p.62). Alongside these activities, I prepared a set of questions "to tap into participants' thought process and to encourage them to critically reflect on issues raised by the activity" (Hennink, 2014, p.62).

Similar procedures were followed for the Phase II FG meetings (see Appendix A9). In Phase II, the same question route was used with both the students and staff, with certain minor alterations to address each sample. In this phase, however, the focus was on the descriptions of the writing skills the students have required in these MHCCs and their reflections on how well the PYP programme prepared them to meet these requirements. Regarding the activities used in this phase, the students as a group were first asked (on A3 paper) to brainstorm all the writing activities they were involved in during their first year in college, including the topics they wrote about and the types of texts they were engaged in (see Appendix A10 for some examples). Using brainstorming techniques in FGs provides 'a synergy' which helps encourage the group participants to 'recount their inner experiences' in relation to the topic and "add to those commentaries as they hear what other group members contribute" (Lederman, 1990, p.119). The other activity used was the self-assessment descriptors that were cut into six boxes (as used in Phase I); however, this time the purpose was to choose only one descriptor that best represented the minimum required level in their college and to discuss their justifications for their choice. The third activity was using the cards with the CEFR categories to promote discussion with participants about what students actually require in relation to the skills mentioned on each card. Appendix A10 includes a sample of the cards used in this phase, with screenshots of the students' brainstorming activity during the FG meetings.

3.5.2.3. Piloting FG questions and materials

In both phases, student FG questions and activities were trialled with volunteers prior to the piloting stage in order to establish a general sense of how they would be delivered: three colleagues from the Applied Linguistics department for Phase I materials and three friends from different departments (one each from the applied linguistics, business and medical schools) for Phase II materials. Further modifications were applied to improve the questions and materials used, based on the experience and feedback received.

In addition to the above-mentioned pre-pilot, in Phase I, a few days before the actual study commenced, the developed FG guide (including the questions and materials) was piloted on (1) a group of six students from the ST at the PYP (2), a group of tutors from the HT and (3) two coordinators from the Professional Development unit related to the HT at the PYP. Findings and notes received and/or observed during this pilot run were considered and used to enhance the guide.

Piloting the questions and activities a second time in Phase II was not possible, as each college has different regulations and they only gave permission to collect data in a limited period (one week in each college).

3.6. Data collection

This section presents the ethical considerations of the study, followed by a detailed account of the procedures followed for data collection in each phase.

3.6.1. Ethical considerations

Acknowledging the importance of upholding sound ethical practices in research, every possible effort was made to perpetuate the highest ethical standards throughout my research's different stages, as suggested by Creswell (2007) and

Richards (2003). As a first step, official permission was sought and granted by the University of Warwick regarding the application, instruments and data collection of the two study phases. The second step was to gain official permission from the relevant educational authorities for data collection.

In Phase I, I applied for formal permission from the PYP Deanship to collect data on the female campus, which was signed and provided by the Dean of the PYP and the PYP research committee. This allowed me to collect data from the PYP including questionnaires, FG meetings, interviews and documents related to the study. Approval to analyse the students' final exam written texts came later.

In Phase II I applied to two educational authorities within the university: (1) the Deanship of Higher Education and (2) the Research Ethics Committee, plus the Dean/Vice-Dean of each of the five MHCCs to obtain their formal approval for access and data collection which were received (Appendix A11).

Prior to data collection, I fully informed all participants about the aims of the research and the consequences of their participation (Punch, 2005). I clarified that it was possible to withdraw from the study at any time during or after participation and participants were also given the chance to ask any questions regarding the study.

In each phase, all participants received an information sheet about the study, including all relevant contact information (see Appendix A12), and a consent form to be signed (see Appendix A13). Both were translated into Arabic to ensure the full comprehension of participants (available upon request).

During Phase I data collection, I visited each class and discussed with students their ethical rights concerning participation in the study.

Similar ethical procedures were followed for the FG meetings and interviews. All participants were assured that any data would remain confidential and would be anonymised. The aim of the study was explained, as well as the purpose of recordings. Participants were assured that they maintained the right to participate or withdraw from the study at any time without any negative impact resulting from their decision. Those who agreed to participate were provided with consent forms to be signed (see Appendix A14) before the commencement

of the FGs. Some staff members in Phase II were willing to be interviewed, but not recorded. For this reason, while only notes were taken during the interviews, the consent form was nonetheless signed indicating agreement to participate. All collected data from each phase was stored on a secured and password-protected hard disk drive accessible only by me (the researcher).

3.6.2. Phase I data collection procedure

3.6.2.1. Quantitative data

To ensure that the data could be collected as planned, first, an email was sent to all tutors in the MT via the head of the English department inviting them to meet me. In the email, I introduced myself, my research topic, and the purpose of the meeting. I scheduled three meetings with the tutors to suit their availability.

During the meetings, I introduced the purpose of my study and the study instruments with which their help would be needed during data collection. I explained that they were welcome to participate in either, both or neither parts of the study. I also stressed that they could withdraw from the study at any time without any negative consequences.

The first part of their assistance included the distribution of the questionnaire to students during their class time. I mentioned that the whole process would take between thirty and forty-five minutes to fully complete, and that formal approval had been gained to contact students during their class time. Furthermore, it was emphasised that they maintained the right to disallow me from entering their class or to choose the time and the class which best suited them and their students. They were also offered the option, if unwilling to monitor the administration of the questionnaire but willing to take part, for me to come and administer it. The tutors were then presented with the procedures they should follow to administer the questionnaire to those students who agreed to participate (described below). Tutors were reminded to mention to students the benefits of participation. This included an individual report summarising each

student's proficiency level based on their own self-assessment using CEFR Scales (see Appendix A15). In addition, students were entered into a draw if they chose to participate.

The second point presented in the meeting was the tutors' participation in the study. They were informed that a similar questionnaire should be used to assess their students, and that this would require the tutor to complete a questionnaire for every student in their class. Hands-on materials were given to the tutors during the meeting to practise using the CEFR scales in the questionnaire. Throughout the meeting, a PowerPoint presentation was used to emphasize the main points and explain how the questionnaire would be used by both students and tutors. The tutors were given one to two weeks to complete these questionnaires. It was clarified that their participation would be valuable and that they would receive a voucher as a gesture of appreciation for the time taken to fill in the questionnaires, as well as being entered into a prize draw for a mini tablet. Tutors were also required to have been teaching the students for a minimum of one semester and to have been involved in assessing the continuous assessment writing project.

Tutors were invited to ask any questions and seek clarification regarding the research and its administration. Some tutors showed immediate willingness to help and support the study. Others agreed to help with monitoring the process of students' questionnaire completion but opted not to assess the students. Others decided not to participate at all. During the discussion with the tutors, some asked to receive recognition (e.g. a certificate) from the department as a reward for their participation. The head of the department offered to provide such certificates to participating tutors at the end of the study.

To collect data from the students, I introduced the research to the classes, then tutors administered the questionnaires. Data collection took place during revision weeks towards the end of the academic year, when almost all students had completed their course syllabus, so it caused no disruption to classes.

To collect raters' data, interested English language teachers with experience in rating L2 texts (in UK) were invited to join a benchmarking training session to be

trained on rating students' samples using the CEFR rating scale. Due to participants' busy schedules and the difficulty in having them all in one day for the training session, two sessions were conducted (with three and four raters, respectively). Participants were involved, during the two-hour session, in familiarisation, standardisation and benchmarking activities (adapted from the *Manual*, CoE, 2009). During the session, participants rated and discussed a few samples and agreed/disagreed on their CEFR levels on the scale. After the session, participants were emailed the materials (i.e. the rating scale, sample texts and the excel sheet with drop-down menu to make it easy to choose the CEFR levels) and detailed instructions of what was required to rate the texts. They were asked to rate 105 texts which comprised 10% of the total available, randomly selected from different proficiency levels: elementary (n=14), intermediate (n=55) and advanced (n=36). Raters were given two to three weeks to submit their ratings. Each of the 105 texts were rated by seven raters.

3.6.2.2. Qualitative data

A separate email was sent to tutors inviting them to participate in FG meetings. Scheduled times were suggested for FGs according to PYP levels. A similar email was sent to the coordinators in each unit inviting them to participate and to suggest a suitable timeframe. Simple refreshments and vouchers would be provided, to encourage participation. The students' FGs were scheduled with the coordination of the department's student administration office. FGs were scheduled to ensure that no more than two occurred per day. Throughout FG data collection, structures and procedures found in the literature (Krueger, 1998; Krueger & Casey, 2000) and the FG guide (Appendix A7) were utilised. A total of nine FG meetings were recorded totalling approximately 18 hours.

3.6.3. Phase II data collection procedure

Phase II data collection differed between colleges as each has its own regulations for dealing with researchers.

This phase did not allow the same luxury of time or flexibility to communicate and contact students and academic staff as the first Phase did.

3.6.3.1. Quantitative data

In each college, I had to deal with an intermediary to liaise with students and academic staff when scheduling and organising data collection. In some colleges, it was necessary to contact the students' class leader (the contact point between students and all other academic or administrative matters related to the department).

One week was planned for data collection in each college. First, I had to liaise with the person responsible for arranging the time and date to distribute the questionnaires to the students. The only time available in almost all the five colleges was during students' breaks or, sometimes, staff members offered five to ten minutes at the beginning or end of their lectures to discuss the research and distribute the questionnaires. When I managed to reach the students, I introduced myself and explained the purpose in conducting this research. I stressed confidentiality and that the students had the full right to participate or not and to withdraw from participation at any time. I explained the benefits of participation: students who participated in this phase would be entered into two prize draws: one for a simple gift within the college itself (a voucher) and another prize draw for a mini tablet among all five of the participating colleges. It was also mentioned that each participant would immediately receive a certificate of participation countersigned by my supervisor. Students then took the questionnaires and returned them after completion.

Members of staff had to be personally approached in some of the colleges. Unfortunately, very few opted to participate due to time limitations and other

pressing responsibilities. For those who agreed to participate, I explained how the CEFR scales should be used (and instructions were included).

3.6.3.2. Qualitative data

Similar procedures were followed to arrange for FGs and interviews with participants in this phase. All meetings were arranged based on participants' availability. FGs were conducted as above. One-hour FGs were scheduled with the students (about 9 hours in total) and thirty-minute interviews with staff members (a total of approximately 5 hours of recordings).

3.7. Data analysis

This part summarizes the procedures followed to analyze both the quantitative and qualitative data of each phase of the study.

3.7.1. Quantitative analyses

3.7.1.1. Procedure for data entry

The quantitative data in this study was analysed using IBM SPSSv24. After collecting the data, all responses were assigned numerical coding before being entered (Pallant, 2013).

The quantitative data in Phase I was based on the use of the CEFR scales. These scales consist of six main levels from A1 to C2; some of these scales contain additional "plus" levels (e.g. A2+), giving a total of 9 possible score levels (see Table 3.4).

On each scale, when a participant ticked *not sure* under a level, they would be marked one level lower. For example, if one of the participants ticked *not sure* under level B2 (number 6 in the coding grid [Table 3.4]), that student was

marked at the next level lower (B1+ i.e. number 5 if the particular scale contained a B1+ level, or 4 equalling B1 if there was no “plus” level for that scale). Data entry was done manually and was randomly checked for accuracy.

There is consistency in data entry and coding throughout the CEFR scales. This represents an important step, especially as all the scales have “the same attributes; [and it would not be] appropriate if the items have different response scales, or if the values are coded differently” (Pallant, 2013, p.36). This increases the ease of data analysis. Checks of the assumptions of any statistical methods applied to the data were also carried out (described below). During piloting, this procedure showed high reliability (Cronbach’s alpha of 0.88 and 0.95). After data entry, an audit was run to ensure that the data were accurately entered by running quantitative frequency analysis (Green, 2013; Pallant, 2013). The same data entry procedure was followed with the tutors’ data.

In Phase II, participants selected the CEFR level that they felt most accurately represented the minimum requirements during their first academic year in college. The same coding scheme as in Phase I was used, and the code corresponding to the level chosen by the respondent was entered in the dataset. If a participant ticked more than one level on the same scale, this item was excluded from data analysis.

For coding the student’s free-list responses about required writing skills, the responses were grouped into categories: the Likert scale ratings (their evaluation of the PYP programme in preparing them for the skills they listed) for each of these categories were entered for descriptive analysis. The four-point Likert scale was coded as 0 (not at all prepared), 1 (slightly prepared), 2 (well prepared) and 3 (very well prepared).

3.7.1.2. Phase I analysis

The purpose of the quantitative data collected in this phase was to ascertain the students’ CEFR levels attained at the end of PYP, as perceived by PYP students,

tutors and raters. Students self-assessed their levels using the CEFR scales and the tutors assessed the same students separately using the same scales, whereas the raters' rated a random sample of students' texts using a different rating scale based on the CEFR.

A descriptive analysis was conducted to check the normality of the data distribution. This step was important to decide which type of test to apply (generally *parametric* for normally distributed data or *non-parametric* otherwise) (Green, 2013; Field, 2009; Pallant, 2013). However, based on the arguments offered in Green (2013), a parametric test can be used even if there is a violation of the assumption of normality when certain other conditions are met, including having a large sample and equality of variances (Lumley *et al.*, 2002). Thirty participants or more counts as a large sample and can allow for the use of a parametric test (Pallant, 2013). Since over 500 students participated in this study, parametric analysis of the data was used. *Levene's test* can be used to assess whether the data has heterogeneity (inequality) of variance, in which case non-parametric tests should be used. After deciding on the type of test to apply, the data were analysed as follows:

First, for an overview of how participants used the CEFR scales in each of the PYP levels, descriptive analyses were performed to find the means (*M*) and standard deviations (*SD*) of students' self-assessments, tutors' assessment and raters' ratings regarding each of the CEFR scales.

It was also important at this stage to identify whether PYP participants (students and tutors) perceived students' CEFR levels differently or similarly across the three PYP levels (elementary, intermediate and advanced). Therefore, one-way ANOVA was used to identify differences across the PYP levels separately for the students and tutors' assessments. After performing the analysis, Levene's test was checked. This test "tests whether the variance in scores is the same for each of the three groups" (Pallant, 2013, p.262). Where Levene's test indicated there was no violation of the assumption of homogeneity of variance, ANOVA was used; when the assumption of equal variances was violated, the non-parametric analysis of variance (e.g. Brown-Forsythe and Welsh Test), as mentioned in Green (2013), was used instead.

If the significance (*P*-value) was <0.05 , this indicates a significant difference between the mean scores between the three groups. However, this does not show “which group is different from which other group” (Pallant, 2013, p.262). For this reason, a post-hoc test, i.e. Tukey’s Honestly Significant Difference (HSD) test (Pallant, 2013) (if there is no violation to the assumption of homogeneity) or Tamhane’s T2 (Green, 2013) (with heterogeneity of variances), needs to be used to check the significance between each pair of the three PYP groups. Post-hoc tests are only utilised if significant differences in means are identified (Pallant, 2013, p.263).

The next step was to compare the students’ self-assessment with their tutors’ assessment. Here, a paired sample t-test was used to check whether the PYP students’ and tutors’ assessments differed significantly.

Following this, the next step was to determine the relation, strength and direction of correlation between the students’ and the tutors’ assessment. Since the rating of the data was ordinal, a Spearman correlation analysis was conducted. To test the agreement between the students’ and their tutors’ assessment, weighted Kappa analysis was also used (Cohen, 1960), which gives more consideration to the closeness of agreement between ordinal data (Cohen, 1968; Fleiss & Cohen, 1973).

Then the percentage was calculated for “exact” agreement; agreement within one adjacent CEFR level (which means that, for example, a student chooses a level that is either the same, one level higher or lower than the tutor; for example, a student chooses A2+ and the tutor either chooses A2 or B1); and agreement within two adjacent CEFR levels.

To compare students’ self-assessment and tutors’ assessment with raters’ ratings, it was important to test the reliability of the raters’ ratings and the reliability of the rating scale. Inter-rater reliability analysis was carried out to make sure that the results were reliable. This is important to ensure that the process and practice of the training session was effective and the raters were well-prepared.

Inter-rater reliability for the five categories of the rating scale was measured using Cronbach's alpha (Table 3.5). It was found that the alpha value was >0.8, indicating good consistency between raters.

Table 3.6 Inter-rater reliability for the 7 raters' scales

	<i>Range</i>	<i>Coherence</i>	<i>Accuracy</i>	<i>Description</i>	<i>Overall</i>
Cronbach Alpha	0.86	0.82	0.84	0.84	0.86

Because the correlation analyses do not measure agreement, it is important to test the agreement between the seven participants using Fleiss's Kappa to test multiple inter-rater agreement (Fleiss, 1971). Fleiss's Kappa is an extension of Cohen's Kappa to allow for agreement between more than two raters to be assessed, though one of its limitations is that it does not allow for disagreements to be weighted according to how different raters' ratings are. To identify the agreement among them, the results (Table 3.6) indicate that for all five scales, the inter-rater agreement is significant based on Fleiss Kappa analysis. However, levels of agreement were low (≤ 0.11), suggesting that although raters rate in the same direction consistently, they don't always exactly agree.

Table 3.7 Generalized Fleiss Kappa for inter-rater agreement between 7 raters.

<i>Scales</i>	<i>Fleiss Kappa</i>	<i>Std. Error</i>	<i>P-value</i>	<i>95% C.I.</i>
Range	0.10	0.01	<0.001	(0.08,0.12)
Coherence	0.08	0.01	<0.001	(0.06,0.10)
Accuracy	0.08	0.01	<0.001	(0.06,0.10)
Description	0.09	0.01	<0.001	(0.07,.012)
Overall	0.11	0.01	<0.001	(0.09,0.13)

Fleiss Kappa: <0=poor agreement; 0.01-0.20=slight; 0.21-0.40=fair; 0.41-0.60=moderate; 0.61-0.80=substantial; 0.81-1.00=almost perfect

After that, I calculated the percentage of agreement between the seven raters for exact agreement, within one adjacent level, or two adjacent levels, and found (Table 3.8) the largest agreement between the two adjacent levels and almost no exact agreement among the seven raters.

Table 3.8 Percentage agreement between the 7 raters

	<i>Exact Agreement between all raters</i>	<i>All raters within one Adjacent Level</i>	<i>All raters within two Adjacent Levels</i>
Range	0%	4.9%	33.3%
Coherence	0%	3.9%	24.5%
Accuracy	0%	5.9%	26.5%
Description	0%	5.9%	24.5%
Overall	1%	5.9%	26.5%

Intra-class correlation (ICC) was also used. ICC reflects the degree of correlation and agreement between measurements and hence gives stronger results in terms of reliability. Interrater reliability among all the seven raters for each element in the rating scale was assessed using intraclass correlation coefficients (ICC). ICCs allows for an assessment of reliability among more than two raters. In this study, a two-way mixed –effect model with absolute agreement is used as the raters, here, are not randomly selected from a wider community of raters. Since we are testing the reliability of the seven raters, the average results of the raters and not the ‘single rater’ type is what is used. Cronbach alpha calculated the reliability (consistency) and therefore, the ‘absolute agreement’ is what I used for the current analysis of ICC.

Table 3.9 The Seven Raters Intraclass Correlation (ICC)

	Intraclass Correlation ^b	95% Confidence Interval		
		Lower Bound	Upper Bound	
Range	Average Measures	.833 ^c	0.771	0.881

Coherence	Average Measures	.789 ^c	0.712	0.849
Accuracy	Average Measures	.817 ^c	0.750	0.870
Description	Average Measures	.801 ^c	0.720	0.861
Overall	Average Measures	.833 ^c	0.769	0.882
Average	Average Measures	.829 ^c	0.762	0.880

Two-way mixed effects model where people effects are random and measures effects are fixed.

b. Type A intraclass correlation coefficients using an absolute agreement definition.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

Table (3.9) shows the Intraclass Correlation of absolute agreement of the seven raters of the rating scale's five categories including the average of those five categories: *Range, Coherence, Accuracy, Description, Overall, and Average.*

Looking at the results, the reliability of the ICC measures ranged from .789 to .833 with a 95% Confidence interval ranging from 0.712 to 0.882, indicating moderate to good agreement (Cicchetti, 1994).

From the above findings, I concluded that the results from the raters' ratings are reliable with high Cronbach's alpha indicating that in general the raters are rating the same underlying construct. As exact agreement was low but reliability high, mean scores across the seven raters were calculated to compare raters' scores to other groups and across PYP levels.

I used ANOVA to compare the means between self-, tutors' and raters' ratings (of the same students in these three groups) and tested the normality of these assessments in order to decide which test to use. To test the direction and the association between the three groups' assessment (self-, tutors, raters), correlation coefficients were calculated. Then, the means were compared across PYP levels to see if students differed based on their PYP proficiency level. Because the group means were used, the score was a continuous number rather than an ordinal scale (as in the individual rating scales). Kappa statistics look at agreement on a categorical scale (Cohen, 1960, Cohen, 1968) and so, as it was

unlikely average scores would agree exactly, a Kappa analysis was not appropriate.

3.7.1.3. Phase II analysis

In Phase II questionnaire (section 1 in Part Two), participants listed the writing skills required in their first year in college and evaluated how well they were prepared by the PYP in those skills. The frequencies of participants' responses on the Likert scales were analysed.

The second part in Part Two of the questionnaire contains the CEFR grid where participants chose one level which represents their minimum requirements on each scale. Descriptive analysis of students and staff was conducted. An independent sample t-test was performed to compare scores between the staff and the students' perceptions of the CEFR levels required. One-way ANOVA was then used to compare the results across the five colleges to examine whether the requirements differed between colleges. Following this, the data were compared via ANOVA across the PYP levels (elementary, intermediate and advanced) to see if the perceptions of the required CEFR levels differed based on students' proficiency levels.

3.7.1.4. Phase III analysis

In this phase, quantitative data (CEFR levels achieved/required) from students participating in both Phases I and II ($n=252$) were compared to identify any differences (Allen, Crossley & McNamara, 2015). Because the students were the same in the two phases, the samples were comparable (Menard, 2002) and the paired samples t-test (Field, 2009) was used. For both phases, CEFR levels were coded numerically and differences calculated by simply subtracting Phase II from Phase I. Negative differences indicate that the CEFR level perceived as achieved

in Phase I was higher than the one perceived as required in Phase II, while positive differences indicate that the CEFR level required was higher than that achieved. Using a paired t-test, the null hypothesis was that Phase I achieved=Phase II required for perfect alignment.

In order to triangulate the results and to look at misalignments from different perspectives, various data from Phase I were compared with data from Phase II. For example, the same procedure mentioned above was followed to compare PYP tutors' data in Phase I with the students' data in Phase II (to examine the alignment between the perspectives of these two groups). Only those students ($n=252$) who were assessed by their tutors in Phase I and participated in Phase II were considered in this analysis (as required for the paired t-test). The same applies to the raters' data in Phase I; only the students whose texts were rated in Phase I and who also participated in Phase II ($n=48$) were included in the paired t-test analysis (see Figure 6.1 in Chapter 6 for a diagram summarising the data analysis).

Then, PYP students', tutors' and raters ratings data were compared with MHCC staff data to identify misalignments. As these are independent samples, an independent samples t-test was used to compare the means of the two groups (e.g. students' self-assessment (in Phase I) with MHCC staff data (in Phase II)) (Field 2009; Prescott 2014.). In these particular groups, as the sample sizes were quite low, and in order to use an independent sample t-test, it is important to verify the assumption of normality in the two groups (Yurdugül, 2008). I used Levene's F test to test for equality of variances between the two groups (Field, 2009, p. 788).

When comparing the means (e.g. of the tutors' assessment of the students ($n=252$) in Phase I versus Phase II staff ($n=19$) perceptions of the levels required), the two independent unequal groups required the independent two-sample-t-test and the assumption of equal variances was tested using Levene's F test.

3.7.2. Qualitative analyses

3.7.2.1. Analysing students' letters

The purpose of the letter was to gain a general evaluation of the students' writing skills at the end of PYP from their own perspective. They were also used to check how students articulated in their own words (some in Arabic and others in English) what they *can do* and *cannot do* regarding their English writing compared with the writing skills the CEFR descriptors and levels identified quantitatively (See Appendix A16 for a sample of those letters).

The main pre-defined categories used to analyse the letters were as follows: (1) the writing skills offered at the PYP (in each PYP level), (2) the skills the students perceived they *can do*, (3) the skills they perceived themselves to still have difficulty with (*cannot do*), and (4) students' satisfaction with their writing (identified from the questions about how confident they feel regarding their writing and the advice they offered to their friends). Thematic analysis (Braun & Clarke, 2006) was used to identify the themes and patterns in the data. This allowed for a more in-depth analysis of what students could do.

The letters were hand-written on A4 paper. Each letter was assigned an ID number to refer to when quoting and to maintain anonymity of the participants. To analyse these letters, they were scanned and uploaded to the MAXQDA software as scanned photos. Then, codes and themes were identified on the relevant parts on the letters (Braun & Clarke, 2006).

Due to the exploratory nature of the study, and beside the categories identified prior to analysis, I also followed inductive coding methods, allowing new themes to emerge (Dörnyei, 2007).

3.7.2.2. Analysing FGs and interviews

The main purposes of analysing the FG meetings and staff interviews were:

(1) To explore students' writing proficiency (in CEFR scales or additional) achieved (Phase I) and required (Phase II) based on the perception of the study participants.

(2) To explore students' experience with writing in the PYP and after joining the college.

(3) To identify differences between what had been achieved and what was required based on participants' views and perceptions of writing (Phase III); that is to identify misalignments (if any) between the present and the target situation. The identification of misalignment would, then, indicate the gaps in the writing course at the PYP and their causes.

Thematic analysis was used to identify patterns and themes in the analysis of the FGs and interviews. An 'initial set of codes' (Knodel, 1993, p.46) was developed based on the 'questioning route' of the FG (Krueger & Casey, 2000), known as the 'focus group guidelines' (Knodel, 1993, p.46). There were some predetermined categories prior to conducting the interviews and the FGs. I was also open to including additional emerging themes from the data (Alderson, Brunfaut, & Harding, 2014). All FGs and interviews were transcribed verbatim. The qualitative software MAXQDA was employed to code the data and analyse the underlying themes of the FGs and interviews.

There were some activities (in both phases) on which participants worked together (in pairs or threes) to identify the CEFR skills they believed they had reached or should have reached and to justify their choices. They also identified different skills they *can do*, *want to do* or *were doing*. Participants wrote and brainstormed what was required (see Appendix A10 for examples). Participants' responses to these activities were also entered and analysed using MAXQDA. After listening to the data several times and reading the transcripts many times to establish familiarity with the data, I identified the common themes. I had also followed the One Sheet Of Paper (OSOP) method (Ziebland & McPherson, 2006, p.407) as one of the ways to help interpret the data (see Appendix A17 for an example). In this method, the researcher makes a summary of each code on one sheet of paper. This "involves reading through each section of data and noting on [a] single sheet of paper, all the different issues that are raised by the coded extracts, along with the relevant respondent IDs" (*ibid.*, 2006, p.409).

In Phase III (gap analysis), the identified themes from both phases were compared to look for the gaps between the two phases. This procedure of

comparing the data is not always a straightforward one, and like many qualitative analysis procedures, some categories do not fit and others overlap (Radnor, 2001).

3.8. Limitations

In carrying out the two phases of this study, I encountered several issues, some of which were the result of the policy of segregation in the university, and other factors concerning the time-frame for data collection as well as logistical considerations.

Due to the segregation policy in Saudi Arabia (Smith and Abouammoh, 2013), it is impossible for female researchers to conduct studies on male campuses (or vice versa). The only way this can be done is by having someone on the other campus taking responsibility for helping conduct the research; however, this option was not available to me for this project. As a result, both phases of the study were conducted only on the female campus of the PYP and the female MHCCs at the university. As such, no data from male students was collected for either phase.

Time and logistical constraints caused me to focus data collection in Phase I on only the MT at the PYP. This subsequently led to the collection of data from only the MHCCs in Phase II.

These time and logistical issues also impacted on the way in which the FGs were conducted as well as the participants in each group. The initial plan was for me to conduct more FGs among students and tutors, with randomly selected students from different classes in each level. However, these FGs did not go according to plan as the study was conducted towards the end of the academic year. At this time, many of the students were busy preparing for their final examinations and it consequently became logistically challenging to arrange for FGs with students from different classes. As a result, only three FGs were conducted with students. All the students were from the intermediate level and each group consisted of students from the same class. Few focus groups were

conducted in phase one especially with the tutors. This could limit the in-depth interpretations of the findings.

Getting the exact same students to participate in both phases of the study proved challenging. Although I obtained approval for data collection from the five MHCCs at the university, a small proportion of the students who participated in Phase I did not participate in Phase II for various reasons, including lack of availability or unwillingness. Similarly, about 70 students who participated in Phase II did not actually participate in Phase I (19% of total participants) for reasons such as changing speciality or coming from different universities.

Another limitation concerned the number of academic staff participating in Phase II of the study. Even though >300 students participated in Phase II, very few staff participated, potentially limiting the analysis, particularly when comparing results between students and their teaching staff. However, the results from the staff should still provide a general overview of their perception of the students' needs and required skills in writing, along with the qualitative data collected.

Another limitation of the study arose from the restricted training that both students and tutors received in using the CEFR assessment grid, and the fact that participants may not have been familiar with the CEFR descriptors. So, potentially, students, in particular, may not have been able to accurately evaluate their language proficiency levels. However, as stated by Little (2006, p.185), "although learners may not always be able to identify formal deficiencies in their use of target language, they generally know which communicative tasks they can and cannot perform, and with what degree of assurance". Moreover, based on experience, "adult learners are capable of making such qualitative judgements about their competence" (CoE, 2001, p.192). In addition, detailed instructions were given just before students used the scales to ensure they were consistent and aware of the process of self-assessment in Phase I or choosing the levels required in Phase II.

To conduct the remainder of the analysis and to compare the data from raters with the students' and tutors' data, I chose samples from the students' actual

texts they wrote for the end of the PYP exam randomly from among the students who had also participated in the self-assessment and had been assessed by their tutors in Phase I. I numbered the texts based on the students' number in the SPSS file to maintain anonymity and not to indicate any reference of the students' actual proficiency level at the PYP or their assessment (either for self-, tutors or raters'). This process of data collection (raters' ratings) has its own limitations. First, the scale that was used is not ideal, but was the only available option at the time of data collection due to time constraints. When piloting the scales, participants mentioned that the scales have very general descriptors that are not suitable to one single task (i.e. the task they are assessing in this study). However, this does not mean that the scales are invalid as they prove to be reliable in the context of this study with high Cronbach's alpha indicating that in general the raters are all rating the same underlying construct. However, the raters rarely all gave the exact same rating, which is likely to be a result of these issues.

Another limitation is that the training session (two-hour benchmarking training session) was short and potentially insufficient. However, only expert raters were recruited, who had experience in assessing L2 written texts using the CEFR, to participate in this study. Also, participants were motivated to participate as they received an incentive of vouchers upon participation and completion of the ratings. Enough time (two to three weeks) was given to the participants to submit their ratings. During the training session, and for each sample rated and benchmarked, time was given to discuss ratings and justifications for choosing a certain level. However, I could not have all seven raters in one training session and had to split them into two groups on different days. The scale used to rate the text data were not the perfect choice, but since I only managed to get approval to use those texts at a very late stage in my project and because of time limitations, it was impossible to adapt the scale to be more suitable to the task in hand.

Chapter 4

A present-situation analysis

Phase I

4.1. Introduction

Following a present-situation analysis approach (Brown, 2016), this chapter presents the findings from the Phase I quantitative and qualitative data. The aim of this phase is to explore perceptions of students' CEFR writing proficiency levels achieved by the end of their studies in the MT of one of the PYPs in Saudi Arabia.

Towards the end of the intensive year at the PYP, the students, their English language tutors and independent raters were asked to use the CEFR scales to quantitatively identify the levels reached by students at the end of this year.

Qualitative data collected from the students' *letter to a friend* and the FGs with the students, tutors and coordinators were then considered as part of an in-depth exploration of student writing and to triangulate and validate the findings from the quantitative data (Creswell, 2005). Thematic analysis was used to identify the themes emerging from the data after it was coded (Harding, 2013; Saldaña, 2015). Finally, to bring quantitative and qualitative findings of this phase together, they are discussed in relation to the relevant research questions at the end of this chapter.

4.2. Students' CEFR levels at the end of the PYP

This part aims to answer the study's RQ1. In order to answer this question, the following analyses were followed:

First, a descriptive analysis of the data was undertaken. It was also important to explore how participants used the CEFR scales and whether their assessment was providing a sound, reliable judgement. Therefore, additional analyses were employed, including comparisons of self-, tutor and raters' assessments. With these analyses, I not only explore how the different groups used the CEFR scales, but also provide an overview of the students' CEFR levels perceived as achieved.

More specifically, to answer the first question, the following analyses were performed and visualized in Figure 4.1:

- a. A descriptive analysis of the self, tutor and raters' assessment.
- b. A One-Way analysis of variance (ANOVA) to compare participants' assessment across PYP levels.
- c. Paired samples t-tests were used to compare the students' and their tutors' assessments.
- d. Correlation and weighted kappa analyses were conducted between the students' and their tutors' assessments to identify the strength and the direction of the relationship and the degree of agreement between them.
- e. ANOVA and correlation analyses were conducted between the self-, tutor and raters' assessments.

The findings from each of these analyses are presented below.

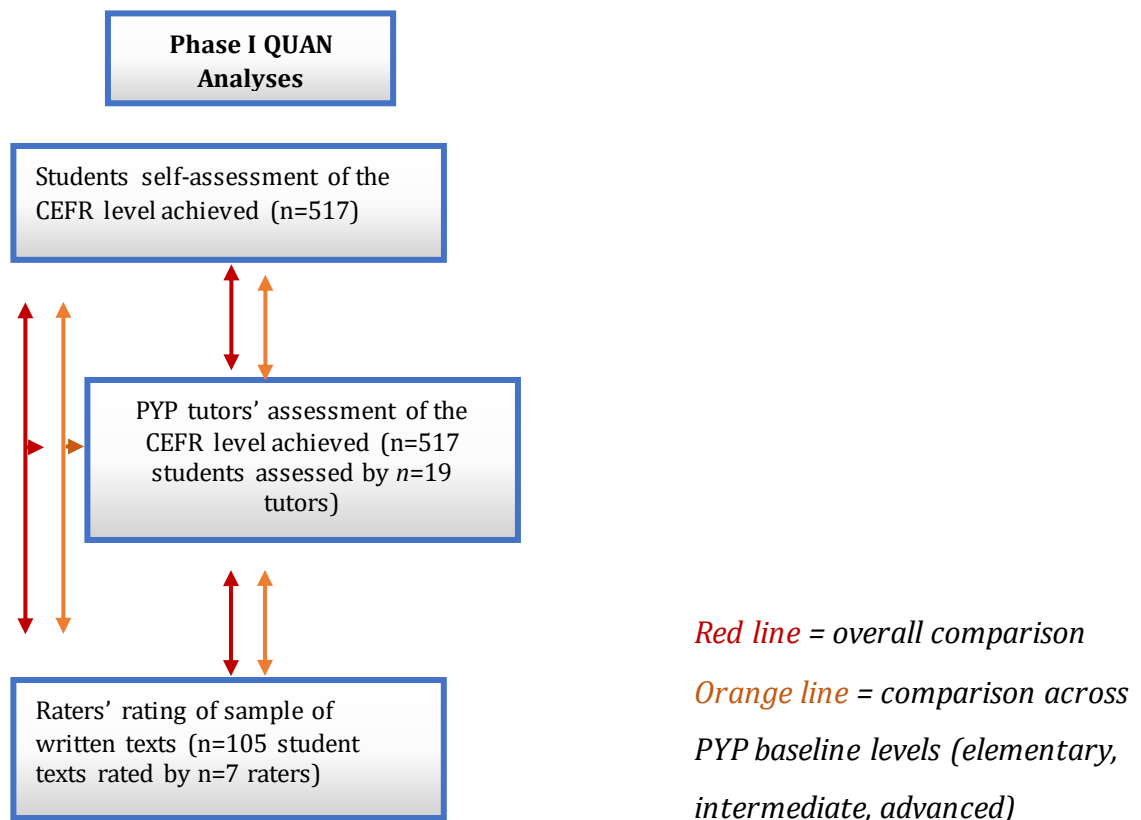


Figure 4. 1 Visualisation of the quantitative analyses conducted in Phase I

4.2.1. Descriptive analyses

Descriptive analysis, as stated by Green (2013, p.41), helps to understand and indicate “the spread of ability” of the students to compare what we have found against our expectations. This analysis is necessary in this study, given the different levels (*elementary, intermediate, and advanced*) at the PYP. It is important to examine the students’ and their tutors’ perceptions of the students’ proficiency levels, to identify whether participants’ assessments vary across the PYP levels, especially given the ceiling effect identified from the PYP final exam (reported in Chapter 1). This ceiling effect made it difficult to explore PYP students’ proficiency across the PYP levels. A more robust analysis of the students’ proficiency levels at the end of the PYP used results from the raters’ data on random samples of texts.

Tables 4.1, 4.2 and 4.3 summarize the means and standard deviations [SD] of the students’ self-assessment, tutors’ assessment and raters’ ratings, respectively. The high reliability of $\alpha=0.88$ and $\alpha=0.95$ for students and tutors’ assessment, respectively, allowed the possibility of using average scores from the ten CEFR scales (Bland & Altman, 1997) (shown as the last line of these tables). High reliability ($\alpha=0.84$) was also obtained from the raters’ rating scale. This is based on the CEFR scale taken from the assessment grid for writing in the *Manual* (CoE, 2009, p.187) and has five simplified categories compared to the ten CEFR scales used for the self- and tutors’ assessment.

Table 4.1 Descriptive analysis of PYP students’ self-assessment across the PYP levels

CEFR Scales	Elementary n=73		Intermediate n=268		Advanced n=176	
	M	SD	M	SD	M	SD
Overall Written Production	5.57	2.35	6.24	2.17	7.91	1.66
Overall Written Interaction	3.93	2.10	4.22	2.28	6.67	2.56
Type of Texts	3.94	2.05	4.28	2.23	6.27	2.48
What Can They Write	4.40	2.24	4.87	2.25	6.80	1.97
Vocabulary Range & Control	3.55	2.00	3.95	1.97	5.85	2.37
Grammatical Accuracy	4.32	2.68	5.08	2.39	6.16	2.84

Orthographic Control	5.05	2.77	5.41	2.67	7.00	2.14
Processing Texts	3.81	1.54	4.39	1.76	6.13	2.23
Reports and Essays	4.14	2.44	4.50	2.41	6.75	2.04
Note Taking	5.22	2.48	5.44	2.30	6.94	2.17
Average of Scales	4.48	1.58	4.92	1.53	6.73	1.43

M=Mean, SD=Standard deviation
Coding scheme for CEFR Scales: 1 (A1); 2 (A2); 3 (A2+); 4 (B1), 5 (B1+); 6 (B2); 7 (B2+); 8 (C1); 9 (C2)

Table 4. 2 Descriptive analysis of PYP tutors' assessment across PYP levels

CEFR Scales	Elementary n=73		Intermediate n=268		Advanced n=176	
	M	SD	M	SD	M	SD
Overall Written Production	4.38	1.86	5.99	2.05	7.56	1.89
Overall Written Interaction	4.12	2.03	5.63	2.16	6.88	1.82
Type of Texts	4.46	2.24	5.80	2.21	7.28	1.80
What Can They Write	3.52	1.85	4.98	1.85	6.52	2.04
Vocabulary Range & Control	3.80	1.59	4.96	1.82	6.31	2.19
Grammatical Accuracy	3.88	1.89	4.98	1.74	6.16	2.24
Orthographic Control	4.22	2.52	4.89	1.83	6.97	1.88
Processing Texts	3.05	1.16	4.06	1.42	6.13	2.38
Reports and Essays	4.03	2.08	5.25	2.05	6.24	2.29
Note Taking	3.75	2.40	4.84	2.17	5.89	2.52
Average of Scales	3.79	1.45	5.12	1.60	6.65	1.54

M=Mean, SD=Standard deviation
Coding scheme for CEFR Scales: 1 (A1); 2 (A2); 3 (A2+); 4 (B1), 5 (B1+); 6 (B2); 7 (B2+); 8 (C1); 9 (C2)

Table 4. 3 Descriptive analysis of the raters' assessment of sample students' texts across the PYP levels

Rating Categories	Elementary n=14		Intermediate n=55		Advanced n=36	
	M	SD	M	SD	M	SD
Range	3.57	1.21	3.90	1.32	5.05	1.28
Coherence	3.50	1.07	3.92	1.35	4.79	1.38
Accuracy	3.47	1.09	3.67	1.26	4.83	1.37
Description	3.55	1.22	3.82	1.28	4.86	1.36
Overall	3.56	1.13	3.87	1.29	4.96	1.28
Average score	3.53	1.14	3.83	1.30	4.88	1.33

M=Mean, SD=Standard deviation
Coding scheme for CEFR rating Scale: 1 (A1); 2 (A2); 3 (A2+); 4 (B1), 5 (B1+); 6 (B2); 7 (B2+); 8 (C1); 9 (C2)

Students, tutors and raters give increasing scores across PYP levels (*elementary, intermediate and advanced*) for each scale. The raters' ratings are lower and less variable than the students' and the tutors' assessments. However, because the data collected from the raters is based on lower numbers compared to students and tutors' data, this needs to be treated with caution.

The discrepancy between students' and tutors' average scores is greatest at the elementary level (students: $M=4.48$ (B1); tutors: $M=3.79$ (A2+)); however, tutors' assessment is closer to the raters' ($M=3.53$ (A2+)). At the intermediate level, students ($M=4.92$ (B1)) and tutors ($M=5.12$ (B1+)) both scored higher than raters ($M=3.83$ (A2+)). The pattern was similar at the advanced level (students: $M=6.73$ (B2), tutors: $M=6.65$ (B2), and raters: $M=4.88$ (B1)). Using results from all three participating groups (*students, tutors and raters*), the average students' perceived CEFR levels at the elementary level range between A2+ and B1; at the intermediate level between A2+ and B1+; and at the advanced level between B1 and B2.

4.2.2. Exploring and comparing participants' assessments across and within PYP levels

One-way ANOVA, correlation analysis, paired sample t-test and kappa analysis were performed to explore the following:

- Comparing PYP students' self-assessment across PYP levels.
- Comparing PYP tutors' assessment across PYP levels.
- Comparing Raters' ratings of the written samples across PYP levels.
- Comparing PYP students' and tutors' assessment within PYP levels.
- Comparing students', tutors' and raters' assessment across PYP levels.

4.2.2.1. PYP students' self-assessment across PYP Levels

This section considers whether students' self-assessment of their CEFR levels varies significantly across PYP levels (*elementary, intermediate and advanced*), to

establish an empirical basis upon which to ‘make claims’ (Green, 2013, p.107) about the findings regarding the differences in assessment across PYP levels. The one-way between-groups ANOVA test was used to determine the statistical significance of the resulting differences among the three levels as ANOVA ‘looks for differences between groups which are not due to chance’ (Green, 2013, p.107). The P-value shows whether there are statistically significant differences in the mean scores of the three levels (Pallant, 2013). The ANOVA relies on the assumption of homogeneity of variances across groups (which can be tested using Levene’s test, Levene, 1960); Table 4.4 shows the scales where the assumption of homogeneity was met and ANOVA was used. Where the assumption of homogeneity was violated, equality of means across groups was tested using the Welch and Brown-Forsythe tests which do not require homogeneity of variances (Table 4.5) (Reed & Stark, 1998).

Table 4. 4 One-Way Analysis of variance of students’ self-assessment of CEFR levels across PYP levels

CEFR Scales	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	η^2
What Students Can Write						
Between Groups	488.83	2	244.42	52.58	<0.001	0.16
Within Group	2393.82	515	4.65			
Total	2882.65	517				
Reports and Essays						
Between Groups	634.05	2	317.02	60.31	<0.001	0.19
Within Group	2686.11	511	5.26			
Total	3320.16	513				
Note Taking						
Between Groups	279.96	2	139.98	26.89	<0.001	0.095
Within Group	2665.77	512	5.21			
Total	2945.74	514				
Average CEFR Levels						
Between Groups	430.63	2	215.32	95.23	<0.001	0.26
Within Group	1164.49	515	2.26			
Total	1595.12	517				

SS=Sum of squares, df=degrees of freedom, MS=mean square, F=F ratio, M=Mean, SD=Standard deviation, η^2 =Effect size: 0.02=small; 0.13=medium; 0.26=large.

Table 4.5 Robust test of equality of mean of students' self-assessment of their CEFR levels across the three PYP levels

CEFR Scales	Statistic	df1	df2	P-value
Overall Written Production				
Welch	56.05	2	186.89	<0.001
Brown-Forsythe	46.07	2	219.18	<0.001
Overall Written Interaction				
Welch	61.47	2	199.63	<0.001
Brown-Forsythe	69.48	2	338.76	<0.001
Type of Texts				
Welch	44.49	2	199.82	<0.001
Brown-Forsythe	49.86	2	338.40	<0.001
Vocabulary Range & Control				
Welch	46.06	2	194.25	<0.001
Brown-Forsythe	51.53	2	316.85	<0.001
Grammatical Accuracy				
Welch	13.99	2	188.51	<0.001
Brown-Forsythe	14.90	2	282.66	<0.001
Orthographic Control				
Welch	29.50	2	191.96	<0.001
Brown-Forsythe	25.11	2	242.60	<0.001
Processing Texts				
Welch	52.33	2	205.86	<0.001
Brown-Forsythe	62.06	2	362.55	<0.001

df=degrees of freedom

In both Tables, for all CEFR scales, the P-value was <.001, indicating that ‘we can be 99 per cent confident’ that these differences among the PYP levels were not ‘due to chance’ (Green, 2013, p.115). In addition, the ANOVA results report η^2 , which is a measure of effect size (larger effect sizes reflecting larger differences; Miles & Shevlin, 2001): values around 0.02 indicate “small”, 0.13 “medium” and 0.26 “large” effect sizes (Cohen, 1988). The effect sizes were .095 to 0.26 indicating medium-to-large effect sizes (differences between groups). The largest effect size was observed for *the Average* of all items ($\eta^2=0.26$). Because these tests do not establish whether differences are found among all three groups, *post-hoc* pairwise tests were used (Appendix B, Tables B1 and B2). These tests allow for multiple comparisons to be made between each pair of levels

(Salkind, 2010). As with the ANOVA, it relies on the assumption of homogeneity of variances. Therefore, only items for which this assumption was met are reported in Table B1. Tamhane's T2 test is an appropriate alternative for cases where variances are not assumed to be equal, and as with Tukey's HSD, multiple comparisons are adjusted for (Cramer & Howitt, 2004). Therefore, the remaining items are reported in Table B2 with comparisons made using Tamhane's test.

Significant differences are evident between the advanced and intermediate levels and the advanced and elementary levels. There were no significant differences between the elementary and intermediate levels, except in the *Processing Texts* scale where all the three levels were significantly different.

4.2.2.2. PYP tutors' assessment across PYP levels

The mean scores in Table 4.2 above indicate differences between tutors' assessment for all the CEFR scales across the three PYP levels. Tables 4.6 and 4.7, below, confirm these differences. η^2 indicates a large effect except for *Note Taking* where the effect is small.

Table 4.6 One-Way Analysis of variance of tutors' assessment across PYP levels

CEFR Scales	SS	df	MS	F	P-value	η^2
Overall written Production						
Between Groups	654.09	2	327.05	84.91	<0.001	0.24
Within Group	2006.79	521	3.85			
Total	2660.88	523				
What Students Can Write						
Between Groups	590.42	2	295.21	80.05	<0.001	0.23
Within Group	1928.81	523	3.69			
Total	2519.22	525				
Reports and Essays						
Between Groups	253.37	2	126.69	27.38	<0.001	0.10
Within Group	2221.15	480	4.63			
Total	2474.52	482				

Note Taking							
Between Groups	250.54	2	125.27	22.78	<0.001	0.08	
Within Group	2640.20	480	5.50				
Total	2890.74	482					
Average CEFR Levels							
Between Groups	547.07	2	273.54	113.51	<0.001	0.30	
Within Group	1267.55	526	2.41				
Total	1814.63	528		84.91	<0.001		

SS=Sum of squares, df=degrees of freedom, MS=mean square, F=F ratio, M=Mean, SD=Standard deviation, η^2 =Effect size: 0.02=small; 0.13=medium; 0.26=large.

Table 4. 7 Robust test of equality of mean of tutors' assessment across PYP levels

CEFR Scales	Statistic	df1	df2	P-value
Overall Written Interaction				
Welch	63.84	2	242.69	<0.001
Brown-Forsythe	60.26	2	357.31	<0.001
Type of Texts				
Welch	64.67	2	235.86	<0.001
Brown-Forsythe	59.86	2	317.83	<0.001
Vocabulary Range & Control				
Welch	59.00	2	253.36	<0.001
Brown-Forsythe	60.82	2	426.13	<0.001
Grammatical Accuracy				
Welch	40.63	2	233.06	<0.001
Brown-Forsythe	44.31	2	366.37	<0.001
Orthographic Control				
Welch	77.17	2	159.67	<0.001
Brown-Forsythe	63.16	2	166.48	<0.001
Processing Texts				
Welch	94.79	2	194.58	<0.001
Brown-Forsythe	116.16	2	357.74	<0.001

df=degrees of freedom

The results of the *post-hoc* tests (Appendix B, Tables B3 and B4 for scales where the homogeneity assumption was met or violated, respectively) show significant differences in tutors' assessments between all three PYP levels.

4.2.2.3. Raters' assessment of students' written texts across PYP levels

Seven raters (experienced in assessing English L2 texts and using CEFR scales) rated a random sample of 105 PYP student texts (10% of the total number). Using the average ratings across the seven raters (Table 4.8), the ANOVA shows significant differences across the three PYP levels, with large effect sizes.

Table 4.8 One-Way ANOVA of raters' assessment across PYP levels

		<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	η^2
Range	Between Groups	37.28	2	18.64	27.823	<i>p</i> <0.001	0.36
	Within Groups	66.32	99	0.67			
	Total	103.59	101				
Coherence	Between Groups	24.64	2	12.32	18.76	<i>p</i> <0.001	0.28
	Within Groups	65.04	99	0.66			
	Total	89.7	101				
Accuracy	Between Groups	35.33	2	17.66	28.99	<i>p</i> <0.001	0.37
	Within Groups	60.32	99	0.61			
	Total	95.65	101				
Description	Between Groups	29.93	2	14.97	24.28	<i>p</i> <0.001	0.33
	Within Groups	61.04	99	0.62			
	Total	90.97	101				
Overall	Between Groups	33.23	2	16.61	25.66	<i>p</i> <0.001	0.34
	Within Groups	64.11	99	0.65			
	Total	97.34	101				

SS=Sum of squares, df=degrees of freedom, MS=mean square, F=F ratio, M=Mean, SD=Standard deviation, η^2 =Effect size: 0.02=small; 0.13=medium; 0.26=large.

The *post-hoc* analysis (see Appendix B, Table B5) showed significant differences in the raters' ratings of students at the advanced versus intermediate or elementary levels for all categories (*range, coherence, accuracy, description* and *overall*), but not between the intermediate and elementary levels in any category.

4.2.2.4. PYP self- and tutors' assessment within PYP levels

A comparison between students' and tutors' assessment within PYP levels was conducted using a paired sample t-test, Spearman's correlation coefficient and weighted Kappa analysis.

4.2.2.4.1. Paired sample t-test

The descriptive analyses (Tables 4.1 and 4.2 above) show close similarity between the students' and their tutors' assessment in some, but not all, CEFR scales and PYP levels. These findings are supported by the results of the paired t-test (Table 4.9). *Cohen's d_z* provides an estimate of the effect size (Pallant, 2013); $d=0.2$ is considered "small", 0.5 "medium" and 0.8 "large" (Cohen, 1988).

At the elementary level, the largest effect sizes were observed for *Overall written production* and *Processing Texts*, followed by *Note Taking*, with students rating themselves significantly higher than their tutors. At the intermediate level, the largest (medium size) differences were for *Type of texts*, *Overall Written Interaction*, and *Vocabulary Range & Control*; in each case the students rated themselves lower than the tutors.

With the advanced-level students, most of the CEFR scales show very similar means (with non-significant P-values and small effect sizes), indicating that students and their tutors have similar perceptions of the CEFR levels students have reached in those scales. However, this was not true for all scales, with tutors scoring significantly higher for *Type of Texts* and significantly lower for *Note Taking* and *Reports and Essays* (small effect size). While this analysis gives some indication of differences between students and tutors, discrepancies can be missed by simply comparing the means; therefore, the next section looks at how closely scores align.

Table 4. 9 Paired differences between self-and tutors' assessment in each PYP level

<i>CEFR Scales</i>	PYP Students		PYP Tutors		<i>T</i>	<i>df</i>	<i>P</i>	<i>Cohen's d_z</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>				
<u>Elementary (n=72)</u>								
Overall Written Production	5.62	2.33	4.41	1.92	3.72	70	<0.001	0.44
Overall Written Interaction	3.96	2.10	4.11	2.12	-0.56	70	0.576	-0.07
Type of Texts	3.94	2.06	4.46	2.38	-1.54	70	0.128	-0.18
What Can They Write	4.40	2.26	3.47	1.85	3.07	71	0.003	0.36
Vocabulary Range & Control	3.59	2.00	3.86	1.66	-0.91	70	0.367	-0.11
Grammatical Accuracy	4.34	2.70	3.85	1.95	1.41	70	0.164	0.17
Orthographic Control	4.78	2.92	4.24	2.59	1.04	50	0.304	0.15
Processing Texts	3.80	1.61	3.00	1.22	2.84	50	0.006	0.40
Reports and Essays	4.00	2.62	4.10	2.05	-0.22	49	0.826	-0.03
Note Taking	5.04	2.69	3.80	2.46	2.70	50	0.009	0.38
Average Scales	4.49	1.59	3.97	1.65	2.24	71	0.028	0.26
<u>Intermediate (=232)</u>								
Overall Written Production	6.26	2.17	5.97	2.08	1.52	226	0.129	0.10
Overall Written Interaction	4.23	2.33	5.60	2.17	-6.61	226	<0.001	-0.44
Type of Texts	4.28	2.26	5.79	2.25	-7.77	228	<0.001	-0.51
What Can They Write	4.78	2.25	4.94	1.86	-0.85	228	0.394	-0.06
Vocabulary Range & Control	3.87	1.95	4.94	1.86	-6.64	230	<0.001	-0.44
Grammatical Accuracy	5.05	2.37	4.95	1.75	0.57	230	0.570	0.04

Orthographic Control	5.47	2.70	4.87	1.86	2.88	217	0.004	0.19
Processing Texts	4.36	1.70	4.01	1.41	2.48	217	2.014	0.17
Reports and Essays	4.55	2.31	5.20	2.10	-3.09	210	0.002	-0.21
Note Taking	5.43	2.18	4.81	2.20	3.00	211	0.003	0.21
Average Scales	4.89	1.51	5.18	1.67	-2.15	230	0.032	-0.14
<u>Advanced (n=170)</u>								
Overall Written Production	7.96	1.65	7.62	1.82	1.87	168	0.064	0.14
Overall Written Interaction	6.74	2.56	6.90	1.83	-0.66	168	0.510	-0.05
Type of Texts	6.35	2.47	7.24	1.81	-4.01	169	<0.001	-0.31
What Can They Write	6.86	1.95	6.56	2.05	1.45	169	0.150	0.11
Vocabulary Range & Control	5.86	2.40	6.31	2.19	-1.75	168	0.082	-0.13
Grammatical Accuracy	6.14	2.89	6.19	2.19	-0.19	168	0.847	-0.01
Orthographic Control	6.99	2.16	7.05	1.78	-0.30	168	0.762	-0.02
Processing Texts	6.12	2.22	6.24	2.34	-0.51	168	0.613	-0.04
Reports and Essays	6.78	2.07	6.26	2.26	2.18	169	0.030	0.17
Note Taking	6.90	2.19	5.96	2.43	3.84	168	<0.001	0.30
Average Scales	6.75	1.44	6.70	1.56	0.34	169	0.734	0.03

M= Mean, SD=Standard deviation

Coding scheme for CERF Scales: 1 (A1); 2 (A2); 3 (A2+); 4 (B1), 5 (B1+); 6 (B2); 7 (B2+); 8 (C1); 9 (C2)

Cohen's-d_z calculated as Mean difference / SD difference. Cohen's-d calculated as $2 \times t / \sqrt{t}$, 0.2=small effect; 0.5=medium; 0.8=large

4.2.2.4.2. Exploring the relationship between self- and tutors' assessment

This section focuses on investigating the strength and direction of the relationship between students' and tutors' assessments, using Spearman's correlation coefficient (r). Values of r of 0.00-0.19 indicate "very weak" correlation; 0.20-0.39 "weak"; 0.40-0.59 "moderate"; 0.60-0.79 "strong" and 0.80-1.0 "very strong" correlation. Additionally, the weighted Cohen's Kappa coefficient was used to measure the degree of exact agreement between students and tutors, which takes into account the exact agreement that can be attributed to chance (Smeeton, 1985). It thus provides additional information beyond the correlation coefficient, which does not take exact agreement into account. Hence, even if a correlation is high, it does not mean that the two raters (students and tutors) ever assigned exactly the same rating (CEFR level). For example, if students always rated one grade lower than the tutor, the correlation would be high despite the differences, and therefore it is important to test the agreement between them. As the ratings used in this study are ordinal data, a weighted Kappa statistic was used (Cohen, 1968). This treats small differences less harshly than large differences. For example, if the student scores A1 and the tutor A2, this disagreement is given less weight than if the student scores A1 and the tutor C2. In this analysis, weights were calculated using the squared differences between ratings (Fleiss & Cohen, 1973). Kappa values of 0–0.2 indicate "slight" agreement, 0.21–0.4 "fair", 0.41–0.6 "moderate", 0.61–0.8 "substantial", 0.81–1 "almost perfect" and 1 "perfect" agreement (Landis & Koch, 1977). Percentages were calculated of student-tutor pairs who exactly agreed, those within one adjacent CEFR level and those within two adjacent CEFR levels. Analyses were conducted on all responses (see Table 4.10).

Table 4. 10 Correlation and agreement between ratings of self- and tutors' assessment

CEFR Scales	Correlation (r) ($n=517$)	Weighted Kappa ($n=517$)	% exact agreement	% within one adjacent CEFR level	% within two adjacent CEFR levels
Overall Written Production	0.29 $P < 0.001$	0.27	31.5	38.9	65.5
Overall Written Interaction	0.22	0.22	23.3	33.2	62.7

	P<0.001				
Types of Texts the Students can write	0.29 P<0.001	0.25	23.6	31.5	60.4
What Students can write	0.28 P<0.001	0.28	25.7	31.6	67.9
Vocabulary Range and Control	0.25 P<0.001	0.25	21.7	35.2	61.6
Grammatical Accuracy	0.23 P<0.001	0.19	15.9	40.8	61.8
Orthographic Control	0.26 P<0.001	0.26	21.5	31.3	68.0
Processing Texts	0.30 P<0.001	0.32	29.9	48.4	73.7
Reports and Essays	0.23 P<0.001	0.15	20.2	45.9	65.0
Note Taking	0.18 P<0.001	0.15	22.7	39.4	59.5
Average	0.39 P<0.001	0.39	19.0	52.4	79.9

There is a significant positive correlation between the ratings of students and tutors for all items (CEFR scales), though the strength of the relation is weak to moderate (all $r < 0.30$ for individual items; $r = 0.39$ for overall average).

Weighted Kappa was low (max=0.39), indicating only weak to moderate agreement in student and tutors' assessment. Overall, 19.0% of pairs agreed exactly; 52.4% agreed within one level and 79.9% within two levels.

4.2.2.5. Comparing samples of self-, tutors and raters' assessment

This section compares students', tutors' and raters' assessments across the three PYP levels, using only the sample ($n=105$) where data exist from self-assessment, tutor assessment and raters, and the average score on the rating scales for each.

Initially, to simplify the analysis and to enable comparison between assessors (*Students, Tutors and Raters*), I calculated the average for each variable and tested the normality of these distributions to decide which test to use. The Kolmogorov-Smirnov (K-S) test of normality was used (Field, 2009; see Appendix B Table B6 for the results). The K-S test for normality showed that the data were normally distributed for students, tutors, and raters. As mentioned in Field (2009), $P < .05$ 'indicates a deviation from normality' (p. 146). Here, the K-S test is not significant,

with $P > .05$ indicating normal distribution. Therefore, parametric statistics were used.

4.2.2.5.1. One-way analysis of variances (ANOVA)

To compare how the three groups differ, a one-way ANOVA was used (Table 4.11).

Table 4. 11 One-Way ANOVA between self-, tutors and raters

	SS	df	MS	F	P-value	η^2
Between Groups	113.74	2	56.87	25.99	$p < .0001$	0.20
Within Groups	662.9	303	2.19			
Total	776.65	305				

SS=Sum of squares, df=degrees of freedom, MS=mean square, F=F ratio, η^2 =Effect size: 0.02=small; 0.13=medium; 0.26=large.

Table 4.11 shows that there are significant differences in the ratings between *self-*, *tutors* and *raters*, with a large effect size. To identify where the differences are located, Tukeys Post Hoc analysis was conducted (Table B7 in Appendix B) and showed that the raters gave significantly lower ratings than both the students and tutors.

Next, I explored the differences using ANOVA across the PYP levels to check how raters, students and tutors differ (Table 4.12).

Table 4. 12 One-Way ANOVA for ratings between self-, tutors and raters across PYP levels

<i>PYP Levels</i>		<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	η^2
	<i>Between Groups</i>	18.26	2	9.13			
<i>Elementary</i>	<i>Within Groups</i>	41.48	36	1.15	7.93	$p < 0.01$	0.31
	<i>Total</i>	59.74	38				

	<i>Between Groups</i>	39.57	2	19.78			
<i>Intermediate</i>	<i>Within Groups</i>	288.67	159	1.82	1.90	<i>p</i> <0.01	0.12
	<i>Total</i>	328.24	161				
	<i>Between Groups</i>	63.83	2	31.92			
<i>Advanced</i>	<i>Within Groups</i>	167.74	102	1.64	19.41	<i>p</i> <0.01	0.28
	<i>Total</i>	231.57	104				

SS=Sum of squares, df=degrees of freedom, MS=mean square, F=F ratio, η^2 =Effect size:0.02=small; 0.13=medium; 0.26=large.

There were significant differences in the average ratings between students', tutors' and raters' assessment for each of the three PYP levels. The effect size indicates large differences except for the intermediate level where the effect is medium. Table B8 in Appendix B explores where the differences are located, showing that for the average CEFR level, the raters gave significantly lower scores versus students and tutors at all the PYP levels.

4.2.2.5.2. Correlation analysis between students', tutors' and raters' assessment

Correlation analysis was carried out to explore the relations between the three assessments (students', tutors' and raters') overall (Table 4.13) and stratified by PYP level (Table 4.14).

Table 4. 13 Overall correlation analysis between self, tutors' and raters' assessment

	Raters n=105		Students n=105		Tutors n=105	
	Pearson Correlation	P-value	Pearson Correlation	P-value	Pearson Correlation	P-value
Raters	1		0.44**	<0.001	-0.11	0.27
Students	0.44**	<0.001	1		-0.065	0.51
Tutors	-0.11	0.27	-0.065	0.51	1	

** Correlation is significant at the .01 level (2-tailed).

Table 4.13 shows a significant positive correlation between the raters' ratings and the students' self-assessment.

Table 4. 14 Correlation analysis between self, tutors and raters assessment across PYP levels

PYP Levels		Raters		Students		Tutors	
		Pearson Correlation	P-value	Pearson Correlation	P-value	Pearson Correlation	P-value
Elementary n=14	Raters	1		0.33	0.28	-0.20	0.51
	Students	0.33	0.28	1		0.20	0.52
	Tutors	-0.20	0.51	0.20	0.52	1	
Intermediate n=55	Raters	1		0.13	0.35	0.09	0.502
	Students	0.13	0.35	1		-0.05	0.72
	Tutors	0.09	0.50	-0.05	0.72	1	
Advanced n=36	Raters	1		0.24	0.16	0.46**	0.005
	Students	0.24	0.16	1		0.33	0.056
	Tutors	0.46**	0.005	0.33	0.056	1	

** Correlation is significant at the .01 level (2-tailed).

In Table 4.14, the correlations between the three groups (students, tutors and raters) were conducted separately for each of the three PYP levels. There was no significant correlation between the three groups at any PYP level, except between raters and tutors in the advanced level. This correlation was found to be significantly positive ($r=0.46$). However, it is important to treat the results with caution especially given the small sample sizes across the PYP levels, and because the tests used might not have required students to show their full range of abilities (discussed below in Section 4.4.1).

In this section, average student, tutors and rater's ratings have been used which means the 'score' for each of the three groups is no longer an ordinal scale and so it is not possible to calculate Kappa to assess agreement. Instead Cronbach's alpha was calculated to assess the intra-rater agreement between self, tutors and raters' ratings of the students. Cronbach's alpha of ≥ 0.9 is described as 'excellent' internal consistency; ≥ 0.8 'good'; ≥ 0.7 'acceptable'; ≥ 0.6 'questionable'; ≥ 0.5 'poor'

and <0.5 as 'unacceptable'. It was found to be 0.651 indicating 'questionable' agreement between the three groups (Bland & Altman, 1997).

To summarize, average ratings of students, tutors and raters were compared overall and by PYP level. While the three groups rarely agreed exactly, ratings were generally pointing to the same direction. Tutors' and students' ratings were more similar to each other than to those of the raters, who tended to rate students' texts lower, particularly for those in the advanced level. However, there were also differences between students and tutors, whose average ratings differed more at the elementary level than the advanced level.

4.2.3. Summary of the CEFR levels as perceived by PYP FG participants

This section summarises the CEFR levels that the students had achieved at the end of the PYP (see table 4.15), as agreed by participants in the FGs. As explained in Chapter 3, participants were given the writing descriptors from the CEFR self-assessment grid (CoE, 2001, pp.26-27). These descriptors were presented in isolation/pieces and participants were asked to classify the ones the students *could do* from the ones they *could not do*. This was done except for the FGs with coordinators, as they have no direct contact with the students to judge their proficiency; coordinators discussed the CEFR scales in relation to their responsibilities in their ELSD unit. For example, the curriculum team discussed the use of the CEFR in PYP curriculum, and the assessment team in relation to testing and assessment. In the table below, the key points are reported.

Table 4. 15 CEFR level perceived achieved by the end of the PYP based on FG data

Participants	CEFR levels perceived students <i>can do</i>
<u>PYP Tutors</u>	
Elementary	B1
Intermediate	B1
Advanced	B2/C1
<u>Intermediate-level students</u>	
B1	5 FGs
B2	3 FGs
<u>PYP coordinators: Key themes emerging from the FG data</u>	
Assessment	Designing test tasks based on A2 CEFR levels
Continuous assessment	Level-based tasks but still easier than their actual level over 4-6 weeks on one task per term.
Curriculum	The team reported to set higher curriculum exit CEFR levels versus actual levels achieved by the end of PYP*
Professional development	Not much training for the teachers on teaching writing as writing is not ' <i>called on</i> ' by students and tutors at the PYP.

*Referring to the PYP curriculum framework in Chapter 1.

Tutors identified that some students at the elementary and intermediate levels can write up to B1 level requirements. Students at the advanced level can perform up to C1 requirements on condition that they are asked to write '*in general context only*' [TUT5_PHI_FG4_187]. Only intermediate-level students participated in FGs in this phase, with most selecting B1 (and the rest B2) as the levels to which students perform by the end of the PYP. Some of these findings support those from the quantitative data, where most students at the intermediate level perceive themselves and were perceived to have achieved B1.

4.3. Writing at the PYP

This section focuses on the analysis and results from the qualitative data of Phase I. FGs with PYP stakeholders (students, tutors and coordinators) and the *letters to a friend* written by the students were analysed with the aim of answering RQ2. Table 4.16 summarises the main themes identified, which are discussed below.

Table 4.16 Phase I themes: Writing at the PYP

-
- PYP students: *perceived proficiency, satisfaction and preparedness*
 - Writing in the PYP Curriculum: *“on the back burner”*
 - Vocabulary: *“the Core”*
 - Medical writing: *“getting blood out of a stone”*
 - Academic writing: *“combating plagiarism”*
 - *“One-Size-Fits-All”*: The assessment of writing at the PYP
-

4.3.1. PYP Students: perceived proficiency, satisfaction and preparedness

4.3.1.1. Proficiency

PYP tutors and coordinators were not expecting students at the elementary level to *“make it to medical school”* [CO09_PH1_FG6, 1094] and should they enrol, *“they will suffer”* [TUT2_PH1_FG5_209]. This is because they (both tutors and coordinators) have noticed a low proficiency of students’ English including writing in the elementary level. Those students use poor grammar, especially *“in subject-verb agreement, they have mixed tenses in their writing, and they know the rules but find it hard to apply them in writing”* [TUT1_ PH1_FG5_224, 225]. The students make *“systematic mistakes”* and have difficulty with the use of *‘cohesive devices and the use of connectors’* [CO011_PH1_FG6_783]. These findings support those from the quantitative data. The elementary level students were found to be between A2+ and B1. The CEFR descriptors in these levels reflect participants’ descriptions of students’ proficiency. For example, A2 descriptors in the CEFR *grammatical accuracy* scale state that the learner *“uses some simple structures correctly, but still systematically makes basic mistakes – for example tends to mix up tenses...”* (CoE, 2001, p.114).

Elementary-level students have *“horrible handwriting”* [TUT3_PH1_FG5_249] with *“disastrous spelling mistakes”* [TUT1_PH1_FG5_253]. All students have problems with *“capitalization”, “starting a new paragraph”* and *“where to put the comma”* and *“it is difficult for them to write complete correct sentences”* [COO11_PH1_FG6_783]. Elementary-level students, at the same time, identified forming grammatically and syntactically correct comprehensive sentences and spelling as most difficult. They saw their limited vocabulary as an obstacle to their ability to write different types of texts. Many expressed concerns that they can write only *“very short paragraphs”* [S32_PH1_LETTER32]. This means that many students in this level are working at sentence level with some basic orthographic problems. These support the quantitative findings as the *Overall written production A2* descriptor stated that students at this level *“can write a series of simple phrases and sentences linked with simple connectors like ‘and’, ‘but’ and ‘because’... ”* (CoE, 2001, p.61). The students at this level *“can write with reasonable phonetic accuracy (but not necessarily fully standard spelling)”* (CoE, 2001, p.118).

Tutors and coordinators expected students in the intermediate level to face a challenging experience in the MHCCs, but not as severe as elementary-level students. One of the tutors described the intermediate-level students’ writing as *“not good and not bad; they have a lot of problems with punctuations and grammar”* [TUT2_PH1_FG5_28]. They have good paragraph structure where they can *“identify the introduction, body and conclusion”* [TUT1_PH1_FG5_51]. However, some of the students have a low proficiency in writing *“similar to elementary level”* [TUT2_PH1_FG5_197] students. Tutors mentioned that students have *“improved from writing a sentence to writing a complete paragraph”* [TUT2_PH1_FG5_136]. The students themselves found it difficult to *“write on a specific topic with specific points”* [ST7_PH1_FG2_57]. Another student said, *“I can write about the school, the university, but I cannot write about general life topics”* [ST16_PH1_FG3_85]. This means that the intermediate students can work above sentence level, at paragraph level. B1 descriptors (where most intermediate students were located) confirmed this, stating that students at this level *“can write simple connected texts on topics which are familiar or of personal interest”* (CoE, 2001, p.26). Students at the intermediate level agreed that *‘spelling mistakes’, ‘having the right*

punctuations, *correct use of words*, *writing with correct grammar and structure* and *the use of medical terminology* are the most difficult aspects of English writing.

Tutors and coordinators believed that most students at the advanced level have an acceptable writing proficiency and are ready to quickly learn and improve their proficiency. Some tutors were concerned that students might “*hit a big wall*” [TUT5_PH1_FG4_430] after the PYP. They nevertheless believe that the students are “*fast enough to pick up...what is lacking*” [TUT5_PH1_FG4_472-473]. Students at this level are expected to “*go to medical colleges*,” unlike students at the elementary level [TUT5_PH1_FG4_475].

In terms of the students’ proficiency, advanced-level students “*can write nicely-planned paragraph[s]*” on a “*familiar general topic*” [TUT4_PH1_FG4_160]. Whereas many students at this level “*can do more than descriptive writing*” [COO15_PH1_FG7_319], other students at this level are “*still dealing with basics*.” This variability found in the qualitative data reflects the range of identified CEFR levels for advanced students in the quantitative data, varying between B1+ and B2+.

Students at the advanced level have *excellent [general] vocabulary*, and can use an acceptable range of phrases and connectors, but they “*lack medical terminology*” [TUT5_PH1_FG4_473]. These statements reflect the B2 descriptors for vocabulary range where learners have “*...a good range of vocabulary for matters connected to his/her field and most general topics*” (CoE, 2001, p.112).

Advanced-level students have an above-average level of orthographic control, although they make some common mistakes. They have “*comprehensible spelling with acceptable common mistakes*” [TUT5_PH1_FG4_366] or “*slip-of-the-hands spelling*” [TUT4_PH1_FG4_372]. This again mirrors B2 descriptors on orthographic control: “*[s]pelling and punctuation are reasonably accurate but may show signs of mother tongue influence*” (CoE, 2001, p.118). The students “*generally... have good grammar control with noticeable mother tongue influence*” [COO12_PH1_FG6_795]. The students are still having difficulties “*expressing their own views in writing*” [ST479_PH1_LETTER_412]. The most salient difficulty

advanced students identify is medical writing due to a limited medical vocabulary and insufficient practice in research writing: “*I think I can convey my ideas through writing. However, I still have a lot to work on terminology ... and the medical skills*” [ST436_PH1_LETTER_436]. Their tutors commented that writing on a medicine-related topic is like “*a high mountain to climb*” [TUT4_PH1_FG4_77] and students “*go blank*” when they are asked to write on medical topics [TUT5_PH1_FG4_454].

Summary of proficiency

The perception of proficiency and what students can do in terms of their writing differed between and within levels. Elementary-level students are most likely to have reached A2+ by the end of the PYP. The students in this level are still dealing with the basics of sentence-level writing. Intermediate-level students have, seemingly, reached B1, while advanced-level students are likely to have reached B1+/B2 by the end of the PYP programme.

4.3.1.2. Satisfaction

Based on analysis of data from the FGs and the letters, most elementary-level students (84.6%) were satisfied with the PYP writing curriculum. Satisfied students wrote that they had improved while attending the programme:

At the end of the preparatory year you will find yourself better than before joining the preparatory year [ST79_PH1_LETTER_79_TR].

They felt they improved their English skills in general, and their writing skills in particular: “*... there is a huge differences between my writing now and before*” [ST49_PH1_LETTER_49_TR]. In contrast, students who were not satisfied (15.4%) felt they had made no or very limited progress in their writing proficiency.

I will be honest and tell you that I have noticed no progress in my English writing...this is because they did not care about improving writing that much” [ST14_PH1_LETTER_14_TR].

Fewer *intermediate* level students were satisfied with the writing programme at the PYP (63%) versus elementary-level students. Unsurprisingly, those who were satisfied believed they had improved, and sometimes specifically mentioned how it improved the “*basic skills of writing*” especially if they had low proficiency. Indeed, many students mentioned that “*unless if we come from a very low level*” [ST8_PH1_FG2_175], the course will not be of benefit.

Those intermediate students who were dissatisfied with the writing programme (35%) not only wrote that they had made ‘*little*’ or “*slight improvement*” but also criticized the PYP curriculum itself for having ‘*less focus on writing*’ compared to the other skills and ‘*totally ignor[ing] writing especially spelling*’ [ST318_PH1_LETTER_318_TR]. The remaining 2% did not refer in the letters to their satisfaction. Many intermediate students complained that the PYP course content is “*similar to what we learned in high school*” [ST104_PH1_LETTER_104]. Indeed, one student went so far as to declare that “*any skill I possess is actually a sum of what I have learned in high school or personal efforts*” [ST282_PH1_LETTER_282]. Other students were not happy with the ‘*simple vocabulary*’ taught in lieu of a ‘*specialised terminology*’ related to their field. They also disliked the fact that the programme only exposed them to ‘*one type of writing*’; that is, ‘*writing descriptive essay*’.

Just over half of the students at the advanced level (57.7%) were satisfied with the PYP’s writing programme. Those who expressed satisfaction saw their teachers’ role as an important factor:

[...] the teacher gave us a lot of instructions such as how an essay should look like in means of the structure and how we should connect and give examples after every new point we introduce. The thing that helped me the most was the feedback both negative and positive...” [ST497_PH1_LETTER_497].

Some advanced-level students were effusive and general in their praise: “*so this programme no matter what level you are, it will help you improve, or maintain your level in English*”. Others specifically appreciated their progress in writing ‘*well-structured essays*’ and a ‘*good thesis statement*.’ For example,

I am glad to say that I've learned a bunch of useful skills that helped me in writing well-structured essays. Now, I learned how to write a good thesis statement that will give the reader an idea of what my essay will be about..." [ST548_PH1_LETTER_548]

Dissatisfied students had many different complaints, but the two most frequent were the programme's omission of academic and medical writing and its lack of a challenge for already skilled advanced writers: "[T]here was not enough focus on scientific research"; "there was not enough practice of medical writing"; "I wish we concentrated more on writing topics that are related to the medical field, regarding medical research" [ST461_PH1_LETTER_461].

Many advanced level students described their dissatisfaction with the PYP programme in extreme terms, as "a waste of time". They described the curriculum as too general and below their level: "English classes are quite boring and we usually spend time talking about irrelevant topics" [ST501_PH1_LETTER_501]. Another bemoaned the "little that you will learn compared to the long hours spent in the class." Like their intermediate classmates, advanced students often found similarities between the PYP curriculum and their previous language instruction: "I personally didn't learn new skills that I didn't know before I got into PY..." [ST422_PH1_LETTER_422]. Another added: "[P]ersonally, I think lower-level students benefited more from it than the highest level because they worked hard to learn, unlike level C [referring to the advanced level] students who saw it easy that they didn't practice much" [ST454_PH1_LETTER_454]

Summary of Satisfaction

Overall, satisfaction rates with the PYP writing course varied greatly between levels, with satisfaction less likely among the more proficient respondents, and more likely with perceived progress in their writing compared to when they first joined the PYP. Many students showed dissatisfaction with the course, especially lack of improvement that could have helped them when they joined the medical colleges.

Students at the lower levels may have overestimated their proficiency levels when using self-assessment tools because of their belief in having greatly

improved their writing skills. Students at the intermediate levels are likewise more likely to have underestimated their proficiency because of their general dissatisfaction with their progress. The data can also be partially explained by the Dunning-Kruger effect, i.e., it takes a certain level of competence to assess one's own proficiency. It is also possible that the elementary-level students may be unaware of what they do not know and hence, over-rated their skills. At the intermediate level they assessed themselves lower than their tutors, i.e. they realised what they don't yet know (i.e. skills have improved but now tend to under-value their skills). Then, at the advanced level, where students tend to be both more competent and more realistic in their assessments of their abilities, self-assessment tends to be closer to their tutors' assessment (indicating more accuracy in their self-assessment). These findings will be explored in more detail in the discussion.

In summary, the feeling of satisfaction with the PYP programme appear to be connected to feelings of progress and development. Students at the elementary level felt they had benefited from the programme and thus felt more satisfied compared to those at other levels. The other levels (intermediate and advanced) had noticed either limited or no progress in their level and hence felt less satisfied than the elementary-level students. In addition, the PYP curriculum seemed to be less beneficial for students at the advanced levels but more beneficial for lower-level students. This feeling of satisfaction due to perceived progress might have affected the students' self-assessment and caused the overestimation of proficiency evidenced in the quantitative data. In the same vein, the feeling of no improvement that caused dissatisfaction might have caused the underestimation of proficiency in the self-assessment of the intermediate-level students. These students appreciate their progress, but they are also aware of their limitations and the language demands of the medical colleges.

4.3.1.3. Preparedness

Satisfaction with the PYP programme is different from feeling prepared. Almost half (49.5%) of elementary level students did not feel well-prepared to succeed in the following year of their studies. Problems with English proficiency were paramount in their letters: *“I feel my level is weaker than what a medical students should be...I was hoping to leave PYP with much better level in English but it seems I need to work harder to improve”* [ST37_PH1_LETTER_37]. Even those elementary level students who did feel ready frequently qualified their statements with claims that more practice and writing would help: *“I’m ready for next year, but with some more practice”*.

Some students at the intermediate level (25%) felt unprepared for the coming year, citing how the limited time and practice at the PYP failed to adequately improve their writing skills, especially academic and medical writing. One student wrote, *“I am not ready for academic writing at the university yet”* [ST119_PH1_LETTER_119]. However, three-quarters of intermediate-level students felt ready to meet the requirements of the medical colleges and, like their elementary-level counterparts, cited their progress in English writing at the PYP as evidence of their preparedness:

In my last two weeks [at the PYP], I feel that I’m ready for the next year especially that my writing skills have improved because it’s [an] important thing and beneficial for next year [ST296_PH1_LETTER_296];

After finishing from the PYP, I feel I have improved my writing skills and I am able to meet all the possible requirements as a medical student [ST373_PH1_LETTER_373].

Most advanced-level students (71%) felt ready to attend medical colleges, as they *“have the basic information. Therefore, I feel I am ready for the next year but still I will learn more just from developing my skills”* [ST472_PH1_LETTER_472]. A lack of exposure to medical and academic writing was these students’ most frequent worry: *“I am not ready...I need more training in medical writing and what I have learned was not much”* [ST477_PH1_LETTER_477];

I don't think I'm ready for next year's medical writing, and it's not because I think I'm lacking. I don't think PY provided enough academic writing practice. I could probably write a book about any general topic, but when a formal, specific academic topic a research piece, PYP did not help much [ST366_PH1_LETTER_336].

Summary of preparedness

Although satisfaction levels may be relatively high among the elementary level students (because of the noticed improvement in their writing compared when they started the programme), this does not mean that they feel fully prepared and ready to meet the language demands of the medical colleges. Intermediate and advanced students felt more prepared than elementary students (mostly because of their level even before joining the PYP). However, neither students nor staff at the PYP were aware of the specific writing skills required of first year medical students when they were in PYP, so students' perceived readiness was somewhat speculative.

4.3.2. Writing in the PYP curriculum: "*on the back burner*"

This theme explains how writing is being dealt with in the PYP curriculum and the actual focus of writing at the PYP as perceived by the different stakeholders participating in the FG discussions and the students' *letters to a friend*.

The PYP curriculum is textbook-based. As noted by a member of the curriculum team, tutors are expected to strictly follow the syllabus provided and cover the relevant textbook content. This current system limits the time dedicated to writing: "*the thing is that the syllabus does not allow students or teachers to spend enough time [to work on writing]*" [CO05_PH1_FG9_496-497]. Participants described the current pacing of the syllabus as '*too condensed*' and '*not structured*' [CO04_PH1_FG9_138]. They described the curriculum as "*a plunge system*" [TUT4, PH1, FG4, 100] where they rush to cover as many different skills and sub-skills as possible. Furthermore, the curriculum limits opportunities for writing practice. One of the curriculum members commented that the curriculum does "*not*

encourage students and teachers to focus on writing... we do have writing lessons, but if we are going to take a look at reality, the writing actually takes a side burner" [CO04_PH1_FG9_131-132]. As one tutor put it succinctly, *"the focus is [on] finishing the curriculum not students learning anything"* [TUT2_PH1_FG5_383-384]. This was also supported by a student saying; *"here, I feel it is about quantity not quality"* [S3_PH1_FG1_96].

Participants mentioned that there are *'no clear objectives'* [TUT5_PH1_FG4_406] to follow in the current syllabus, which makes it difficult for both students and tutors to identify specific skills that require development. Nor is writing part of the placement test at the beginning of the course or of the midterm exam. As part of a process of continuous assessment, students are asked to write on one short topic for about four to six weeks and they get full marks for completing the work rather than for the quality of their writing. For example, referring to the continuous assessment task, one of the coordinators commented:

Writing the same draft over four weeks doesn't make sense. In those four weeks, they could have written like two essays or four different paragraphs. Also, it needs to be really graded with marks. Not giving only full marks [CO012_PH1_FG4_897_898].

In addition, only 10% of the final standardised exam at the end of term is dedicated to writing, and even then, students are asked to write on a very easy, general and frequently recycled topic.

The teaching of writing in the PYP curriculum focuses mainly on paragraph and essay structure and organization and grammar and less on developing vocabulary. Essay and report writing are not practised consistently. Some students, however, feel *"it is still very early to learn how to write reports"* [S9_PH1_FG1_96] and are confident that they will master report writing when they begin their field of study in university.

Although students and tutors acknowledge the importance of writing in general and for the students' future academic and professional paths, writing is not perceived as important on the PYP due to its limited focus both in the syllabus and assessment; consequently, students are not interested in making the effort

to improve their writing, and tutors are unwilling to consider writing for their own professional development as indicated by one of the coordinators at the PYP:

I think it's [writing] not called on because [...] it's not something that institutionalized...that students must have to develop... [COO2_PH1_FG8_100-103].

Another added,

[...] unless there is a reason for the teachers to want and come and learn about writing [...] very similar, they both [students and teachers] need the external reason to do it [COO1_PH1_FG8_139-147].

4.3.3. Vocabulary: “*the Core*”

Vocabulary was one of the persistent themes that emerged from the data. Vocabulary was perceived as important in the curriculum and for the different participants. The curriculum team described vocabulary as ‘*the core*’ [COO6_PH1_FG9_1174] of their curriculum and said that the “*main focus goes to vocabulary development including medical terminology*” [COO4_PH1_FG9_180].

However, although the curriculum is designed to develop the students’ vocabulary, including medical terminology, vocabulary is still one of the most problematic areas affecting the students’ proficiency in writing. Students find it difficult to ‘*express nuances*’ and usually have ‘*incorrect diction*’ [COO7_PH1_FG9_1199]. Although they are exposed to an extensive range of vocabulary throughout this intensive year on the PYP, new words are introduced ‘*out of context*’ [TUT2_PH1_FG5_155] and not applied in writing.

Some students feel they have improved their vocabulary whilst on the PYP, but more are not satisfied with their mastery of medical terminology – a feeling shared by their tutors. Students also find it difficult to apply their oral vocabulary to writing tasks: “*I know the meaning of the words, but I do not know how to use them in writing*” [S14_PH1_FG3_365]. A greater emphasis on context and increased time devoted to writing practice was identified by participants as a

simple, effective solution, *“one of the major mistakes that we do here is usually vocabulary is taught out of context...and it is supposed to be taught through context”* [TUT2, PH1, FG5,155-159]. Also, the data suggested very limited exposure to medical terminology in the PYP, including in assessment.

Two opposing views emerged from the data. Some participants were against the introduction of medical terminology and medical writing in the PYP, arguing that students are going to join various medical colleges, and it is impossible to focus on each college’s different needs. They pointed out that PYP’s curriculum is explicitly about improving writing skills in general so when students join the specialized college, they can learn the specific writing skills and terminology they need there:

[...] I think a PY[P] should be just a beginning of basic English skills...academic skills as well. But not particularly geared towards a certain areas because you don’t know if you’re going to get into medical school...I think the more practical vocational skills are not predictive...for the actual skills...you can transfer them when you go to the three-year or the four-year college [COO16_PH1_FG7_1274-1280].

Other participants argue that since the PYP is designed to prepare students for their colleges via its specific PYP MT, it ought to introduce medical terminology and basic medical writing skills relevant to all colleges:

[B]ut we should give them the basic and these things [basic medical terminology]... [COO15_PH1_FG7_1282-1284];

[I]f we start the process [with introducing medical terminology] I think we armed them with some good skills that they need to have [COO13_PH1_FG7_1308-1309];

The students will be ahead of the game [if they started learning basic medical terminology starting from the PYP] [COO9_PH1_FG6_849]

In summary, vocabulary was perceived as important and the core for writing, but was being used out of the context of writing and was more focused on general vocabulary than medical terminology.

4.3.4. Medical writing: “*getting blood out of a stone*”

In this part, I discuss how writing on medical-related topics and genres (EAP & EMP) has been introduced in the PYP curriculum, based on stakeholders’ perceptions. The first important issue which emerged from the data regarding writing was that the students have very limited exposure and practice to these types of writing. EAP and EMP writing are not much in evidence in the curriculum, for example, one of the tutors remarked that “*with medical writing...there is not much focus*” [TUT4_PH1_FG4_389]. Participants pointed out that although there is a specific textbook for medical writing: “*not much time [is] dedicated to this book in their syllabus... [The current curriculum] is not fully exposing the students into the textbook [referring to the medical textbook] [CO04_PH1_FG9_523] “they are covering less than 50% of the textbook” [CO010_PH1_FG6_171]. Writing tasks are often cancelled: “if we don’t finish the whole book, how do you expect the students to fully reach a certain level by the end of the course?” [CO04_PH1_FG9_525].*

In addition, EAP and EMP writing do not feature as components of PYP mid-term or final assessment at all; “*they [the medical components] are not assessed the way that we would assess a general writing task*” [CO012_PH1_FG6_546-548]. This absence from assessment was a result of students having complained about having to complete difficult tasks that can put them at risk of losing marks:

They don’t get enough writing practice in class with medical related terms...so when they get such topics...they complain, oh ‘*we don’t know how to write about diseases’...‘it’s difficult for us to write’, ‘we know the terms, but we cannot write about it’...so apparently they need more practice in class to be able to do it... in order to give them something like that,...they need to practice it first [CO010_PH1_FG6_115-124].*

In the PYP, students have limited exposure to medical texts, and the only time they are given the chance to write on medical-related topics is in the second semester, as part of an optional task that comprises only one element of

continuous assessment. This lack of familiarity was understood to be one of the root causes of these problems: “*it’s a new whole ball game*” [TUT4_PH1_FG4_81]. Students do not have enough background knowledge of the topic, and tutors, even at the advanced levels, struggle with transferring basic writing skills to medical contexts. When trying to write in medical contexts,

the students couldn’t get it right, and they found it very difficult...to transfer basic writing skills to medical context...for most of them, this is the first time they actually see something at chunk of texts in a medical book and it is quite daunting experience for most of them
[TUT4_PH1_FG4_110]

This limited focus on medical writing in the PYP curriculum discouraged PYP students. First, the students, especially in the advanced level, desired more medical writing tuition and practice. However, not enough focus was dedicated to this type of writing. One of the tutors commented that “*many students are very disappointed because they come here very keen [...] but they do not get enough input [in the medical-related English writing practice]*” [TUT4_PH1_FG4_133_134]. Second, with less focus on teaching this type of writing and with excluding it from assessment, the students and the tutors did not take it seriously. With this attitude, this type of writing remained difficult and no improvement was recorded. They are “*interested in the medical book... but if it is not part of the exam component the students will not take it seriously*” [COO6_PH1_FG9_364-365].

Medical writing, which was part of the medical textbook assigned to the students in the PYP curriculum, was not properly taught nor practiced. The students were interested in this type of writing, but because of its difficulty, they demanded not to be assessed on it for fear of losing marks. All this led to medical writing not being taken seriously and hence both students and tutors reported dissatisfaction with the students’ limited progress in this type of writing.

4.3.5. Academic writing: “*combating plagiarism*”

Teaching academic writing is perceived as important in the PYP, as one tutor noted:

It is important to teach academic writing after some introduction to general writing. It is important that the students are exposed to different genres in writing, to structure an essay according to [its] genre and to know [its] vocabulary [TUT2_PH1_FG5_41-44].

However, students are rarely exposed to academic writing (teaching and practice) despite being cited in the PYP curriculum outcomes. Students were rarely required to write research papers or assignments, and when they did, plagiarism (especially copy-and-pasting from websites) was rife.

Writing a research paper presented students with the greatest challenge. The students were not able to express their thoughts in writing, and they lacked the necessary background knowledge to support their arguments. Furthermore, they were not allowed to use outside references for fear of plagiarism.

The fear of plagiarising, it was found, was the main reason for not engaging students in academic writing tasks. At the same time, there was evidence in the FGs to suggest that students have not been adequately taught how to summarise, paraphrase, cite, and write references:

[...] we didn't teach them how to reference. There was a lot of problem with plagiarism... What happened was is then we had a problem with plagiarism because now we hadn't taught them how to paraphrase and also we didn't give them the opportunity to reference their material but you and I both know in an academic context we would have referenced our material and with that we'd avoid the plagiarism and also we would have been given the opportunity to use a resource or resources in order to research something like that [COO14_PH1_FG7_251-257].

To summarize, academic writing was perceived as important by almost all participants. However, it was not properly considered in the PYP curriculum.

Academic writing was expected to be considered for writing assignments and homework, yet it was found that academic writing was not being requested for fear of plagiarism as students were frequently found plagiarizing. Two ways were suggested of combatting plagiarism at the PYP. The first was not to involve students in assignments which requires writing from different resources for the fear of plagiarism. The second was to prevent students from the use of resources and only ask the students to write from the top of their head. Another problem was that very limited time and practice were dedicated to teaching basic academic skills such as summarizing, paraphrasing, using references and citations and how to express their ideas and thoughts in writing.

4.3.6. *"One-size-fits-all"* the assessment of writing at the PYP

The final writing exam in the PYP is a standardised proficiency test used across all levels and tracks. It is considered a high-stakes exam where the results affect the students' future enrolment into the different MHCCs. This exam is designed locally by the PYP assessment team. They have followed the same task but with different topics over years now, requiring students to write 120/150 words on a general descriptive topic. Because it is standardised across levels, the assessment team finds a topic that is familiar to students across the three PYP levels. The process of choosing the topic was described by one of the coordinators as:

We're mainly told to look for the topics, first of all, that are common to all of the books that are being used by the different courses...so that they are looking for common topics, common themes, which is not always easy because from level A to level C [elementary to advanced], there can be a big difference in what they are covering. So usually I end up pulling out four or five themes that we are able to use and these again are in general topics...we aren't told to use the medical books at all for these themes. And, then, after that we create the prompts based on [those themes]
[CO08_PH1_FG6_167-176]

The level of difficulty of the test should accommodate all the students so as “*to be fair*” and to ensure that “*even students in the elementary level can get high marks in the exam*” [CO012_PH1_FG6_179]. Because of this, the writing task is designed for A2/A2+ level. In addition, the assessment rubric “*hasn’t been changed for years and...also has not been changed from the first to the second semester*” [CO09_PH1_FG6_243-244]. Assessments include a “*fixed-form prompt*” that is structured in a certain way on general topics with tiny word counts.

The way the task is designed and assessed at the PYP, makes it “*very easy*” for most of the students and causes a ceiling effect in the results (see Chapter 1). “*[T]he final [writing] exam is very very simple in terms of the word counts, in terms of the prompts, I mean in everything [it is very easy]*” [CO02_PH1_FG8_102-104]. And because of this simple exam the students do ‘*not take writing seriously*’ [TUT3_PH1_FG5_378]. The problem of standardisation was summarised by one of the coordinators as follows:

And this is the problem. This is the inherent problem in having a standardised test across the courses, across the tracks. You’re going to have to teach to the lowest level and designing to the lowest level. At least for a good portion of the exam. There are a few questions that are a bit more difficult, that are ok but the majority of it is aimed to satisfy elementary and intermediate levels. Advanced is sort of left out there and that’s where most of the medical students are going to be [CO08_PH1_FG6_190-195].

The assessment in this way turned to be ‘*disappointing for most of the students.*’ One tutor remarked:

[the students] say, ‘ok. I have met the benchmark they need me to be at. Then why am I sitting in this English class?’ Since students do not feel challenged, teachers have hard time keeping their interest [TUT5_PH1_FG4_518-520].

The assessment team considers the current standardized proficiency exam “*not appropriate*”. All levels of students at the PYP levels receive very high scores, and there is no indication from the overall results of the students’ different writing proficiency levels. Some coordinators believe that students “*come into the PYP and leave the PYP with very similar skill levels*” [COO1_PH1_FG8_146-147]

Students across proficiency levels agreed on the ineffectiveness of the standardised exam, although their reasoning for this belief varied. Elementary-level students perceived the test as too difficult for their proficiency. One of the students at the elementary level wrote (in English), “*You are not going to be tested on what you have [studied] in textbooks because the exam is a standardised exam across the three levels*” [ST27_PH1_LETTER_27]. Another student wrote; “*you need to work very hard in order to be able to pass the exam*”. Other students mentioned that the exam does not draw from the textbook, so they need to work harder to improve their level in English. Another said “*exams are making extra pressures for us as students in the elementary level*” [ST3_PH1_LETTER_3]

Intermediate level students viewed the exam as easy, but needed some work to excel or obtain a ‘*full mark*’. They emphasised the importance of preparations before the test in order to do well in the exam: “*in the first semester with some focus you will get the full mark*” [ST51_PH1_LETTER_51].

Students in the advanced levels found the test simple and too easy. One described it as “*below your level,*” [ST399_PH1_LETTER_399] another as “*easier than IELTS tests*” [ST437_PH1_LETTER_437]. These students did not find preparation as necessary as their intermediate peers: “*you only need one day to prepare for the test*” and “*You can only revise the list of the vocabulary prepared by the university [to be able to pass the exam]*”. Many advanced students were disappointed by the assessment’s focus on general English: “*I am sorry to tell you will be basically wasting your time per week in order to take a very general exam*” [ST367_PH1_LETTER_367].

It is to be expected that a standardized test should differentiate between students of different proficiency levels; however, the quantitative data suggest that the

test was actually very easy for all three groups, and a ceiling effect is evident in the students' scores in the final exam.

4.4. Phase I Discussion

It was important in phase I of the study to understand the students' proficiency levels and what they '*can do*' in their writing. This is because, although there have been different studies conducted relating to Saudi students' writing in general (Aljumah, 2012; Alkubaidi, 2014; Hellmann, 2013; McMullen, 2009; McMullen, 2014; Obeid, 2017; Oraif, 2016) and writing for Saudi medical students in particular (Alkubaidi, 2014; Ghobain 2014; Shukri, 2008), to the best of my knowledge none has investigated medical students' proficiency in writing in relation to the CEFR. Phase I identified students' proficiency levels achieved by the end of the PYP intensive English programme (the exit level), which is also the entry level of students joining the medical colleges. The findings of this phase also contribute to our understanding of self-assessment, tutors' assessment and raters' ratings using the CEFR scales. This phase of the discussion section is divided into two parts: firstly the quantitative findings regarding the first research question and secondly the qualitative findings regarding the second research question. I will discuss both the qualitative and the quantitative findings in light of each other in order to triangulate and support the interpretation of the findings.

4.4.1. RQ1: Students' CEFR proficiency levels at the end of the PYP

At the end of the PYP, the average CEFR level of the elementary students based on their self-assessment was found to be *B1*; tutors' assessment was *A2+*; and raters *A2+*. For intermediate students the results were *B1*; *B1+*; and *A2+*, respectively. For advanced-level students, the levels were *B2*; *B2*; and *B1*, respectively. Similar results were reported from the qualitative data.

All the participant groups (students, tutors and raters) used the CEFR scales and their descriptors to identify/assign the CEFR levels in a way that aligns incrementally with the students' PYP levels, assigning higher levels to advanced

students and lower levels to intermediate and elementary students. When the participants used the CEFR scales to assign a specific level to a certain student/text, there were no other factors which may put the assessors in a situation where they intentionally or unintentionally compare between the students in the three PYP levels. For example, all participating tutors were only teaching one level (either elementary, intermediate or advanced) when the data were collected. In this way, the norm-orientation is controlled as much as possible during data collection. Nonetheless, it is important to acknowledge that some tutors have previous teaching experience in teaching the other levels. Similarly, raters were given mixed random texts without reference to anything that might indicate to which level the texts belonged. This indicates that the CEFR can potentially be used to gain a criterion-referenced general overview of the students' proficiency levels as a starting point in a context outside of Europe and with participants having no or little experience with using the CEFR scales (Abdulhaleem & Harsch, 2018). It also suggests that the participants have comprehended the CEFR descriptors and related them to the students' proficiency.

Considering the CEFR levels identified across the three participant groups (students, tutors and raters), generally the three PYP levels differed significantly from each other. This is unsurprising as the PYP curriculum framework assumes that the CEFR exit levels will differ between different baseline PYP levels: elementary students should have achieved B1; intermediate B1/B2 and advanced C1 (see Figure 1.1). The results of this study identified similar exit levels: elementary students had achieved A2+/B1; intermediate B1/B2 and advanced B2/C1 (see Table 4.15). However, this finding contrasts with the ceiling effect found in the students' final writing exam scores, where most students scored full marks when rated by tutors at the end of PYP. This ceiling effect means that the writing scores were not differentiating between students with different levels of proficiency, which could cause problems (such as in admission) when the students join the medical colleges (see Chapters 5 & 6).

The final exams at the PYP are considered high-stakes exams, and the students' results affect their GPAs, and consequently their enrolment into the different

colleges. To get a guaranteed place in the more competitive medical colleges (especially the CM), students need to score a 'highly competitive GPA 4.5/5.0' (Al Makoshi, 2014, p.8) as well as high grades from High Schools, in the Achievement Test and the Aptitude Test (Al Alwan *et al.*, 2013). Even with these high admission requirements, this study (see Chapter 5) and the literature report that students are joining medical colleges with proficiency lower than expected (Nazim & Hazarika, 2017). This could be because administrators and decision-makers in the MHCCs rely on the students' GPA at the end of the PYP when admitting students (Al Alwan *et al.*, 2013; Al Makoshi, 2014; Al-Omar, 2014; Al-Shehri *et al.*, 2013). It offers a possible explanation for why students are still joining medical colleges with levels of proficiency that do not meet the expectations of their receiving colleges. My study shows that students are leaving the PYP with significant variations in their proficiency, yet they join the same colleges with similar GPAs. This was obvious from the ceiling effect identified where most students achieved high grades at the end of the PYP, which partially explains the dissatisfaction reported in these colleges. In other words, the PYP final assessment is not an appropriate indication of students' proficiency levels in writing.

Regarding the students' self-assessment results, elementary-level students tend to overestimate their proficiency (CEFR levels) compared to other assessors (tutors and raters). This was expected as it has been widely found in the literature that low-proficiency students tend to overestimate their proficiency (Babaii *et al.*, 2016; Leach, 2012; Ünalı, 2016; Blue, 1988). This has been described as the 'metacognitive deficits' of the 'Dunning-Kruger effect', i.e., it takes a certain level of competency to be able to assess one's own proficiency (Kruger & Dunning, 1999), since "the same knowledge that underlies the ability to produce correct judgment is also the knowledge that underlies the ability to recognize correct judgment. To lack the former is to be deficient in the latter" (Kruger & Dunning, 1999, pp.1121-1122). The inaccurate assessment may also be due to the students' lack of experience in self-assessment (Babaii *et al.*, 2016; Engelhardt & Pflingstorn, 2013). Another possible explanation, which was pronounced in my data, is elementary-level students' satisfaction with the programme (see section 4.3.1.2.) due to the progress they have observed – which

might not necessarily be sufficient to meet the expectations of their future colleges.

Some students at the intermediate-level over-estimated their levels while others under-estimated them. For those who overestimated their levels, the same rationale discussed for elementary-level students could also apply, because some students had similar proficiency to elementary-level students (supported by the qualitative data). In addition, intermediate-level students in the PYP make up the largest group ($n=268$, 46%) with a wide range of scores on the baseline placement test (46–85 out of 100). Some intermediate students will be more similar to elementary students, while others are more similar to advanced level students. Some tutors and coordinators suggested that some students at the elementary level have purposely put themselves in lower levels than their actual ones to get into the easiest course and thereby achieve higher marks:

...she [referring to a student] probably tested low on purpose or something like that on her placement test so that's why she's in category A [elementary level] [COO10_PH1_FG6_1097].

This was similar to Alobaid's findings (2016) where “one-third of [his] study students intentionally underperformed on [English as a second language] ESL placement test” (p.13). This indicates that the PYP needs to revise its placement test, admissions policy and its procedure for placing students into different levels. Similarly, Hughes (2002, p. 17) argued against buying commercial placement tests and instead called for ‘in-house’ and ‘tailor-made’ tests which were proposed as likely to be more accurate.

In contrast to the elementary level students, some intermediate-level students were found to underestimate their proficiency compared to their tutors' assessment. This could be due to those students becoming more aware of their abilities and having higher proficiency compared to the elementary level. Similar results were also found in the literature, where high proficiency students show a tendency to underestimate their proficiency level when they assess themselves (Boud & Falchikov 1989; Kruger & Dunning 1999; Kun, 2016; Hodges, Regehr, & Martin 2001; Lejk & Wyvill 2001; Tejeiro *et al.*, 2012). This tendency to

underestimate in self-assessment could be due to students being over-modest (Kun, 2016). Furthermore, the current study suggested that the underestimation of levels could also be attributed to the students' dissatisfaction with their progress during the PYP.

Regarding advanced-level students' self-assessment, generally, this was closer to that of their tutors and showed less variance than at the other levels, indicating more accurate self-assessment (Sahragard & Mallahi, 2014). This was found in other studies that described more similarities between the students and their teachers' marks/assessment and therefore considered the assessment as more accurate when students came from higher-proficient levels (Falchikov & Boud, 1989; Kun, 2016; Ünalı, 2016). This increase in accuracy could also be attributed to the Dunning-Kruger effect, where students at higher proficiency levels have the cognitive ability to assess and judge their proficiency more accurately.

One way to determine the accuracy of self-assessment is to compare it with other methods, such as tutors' judgments or other test scores (Abdulhaleem & Harsch, 2018; Ashton, 2014; Babaii *et al.*, 2016; Boud, 1991). Generally, high correlations between self-assessment and other measures of performance are unlikely (Dunning *et al.*, 2004). In this study, a significant moderate correlation between the students' self-assessments and tutors' assessments was found ($r=0.38$). This is similar to the average correlation identified by Falchikov and Boud (1989), in their meta-analysis of studies comparing self-assessment with teachers' marks, which reported an average correlation of $r=0.39$. Correlation between self-assessment and students' 'actual performance' (e.g. scores in a test) was very low with no more than $r=0.21$ (Falchikov & Boud, 1989). The correlation between students' self-assessment and the raters' ratings of the students' written texts was found to be higher than in the literature ($r=0.44$). The higher correlations found in this study compared to those reported in the literature may indicate the suitability of the CEFR to self-assessment, although more training and experience might be needed to improve the results.

Even if results correlate significantly, this does not necessarily demonstrate agreement (Fleiss & Cohen, 1973, Cohen, 1968). To the best of my knowledge,

few studies investigating self-assessment – especially language proficiency-related studies – have compared agreement between students’ self-assessment and their tutors’ assessment. In my study, I used a weighted kappa to test the significance and percentage agreement between the two assessments. Exact agreement between students’ and tutors’ assessment was low (19%) but was higher between one (52.4%) and two (79.9%) adjacent CEFR levels. The two adjacent levels in the study means that the agreement is equal to “one and a half levels, e.g. A2+ to B1+”, which is considered sufficient agreement according to the *CEFR Manual* (CoE, 2009, p.37). This means that the students were not too far away in their perceptions of their CEFR levels from those of their tutors, suggesting the value of using the CEFR scales as exemplified in this study.

The raters’ assessments were significantly lower than the students’ and tutors’ assessments across all PYP levels. These findings are consistent with Fleckenstein *et al.*, (2018), who found a correlation of ($r=0.41$), and noted that teachers overestimated their students’ overall levels when compared to the students’ actual performance in their achievement test. This overestimation in tutors’ assessment in comparison to the rating scores “led to a large, systematic discrepancy between CEFR-based judgements of teachers and the standardised test scores” (Fleckenstein *et al.*, 2018, p.9). This overestimation was similarly evident in Běrešová’s (2011) study, where teachers tended to overestimate students’ vocabulary, grammar and language use in comparison with actual test results.

Regarding the relationship between the tutors’ and the raters’ assessment, the only moderate but significant correlation ($r=0.46$) was found between the tutors and the raters in the advanced-level students, which is similar to Fleckenstein *et al.* (2018). However, it was lower than was found in a meta-analysis by Hoge and Coladarci (1989) conducted to compare teacher-based assessments with students’ actual achievement in tests, which reported that correlations ranged from $r=0.28$ to $r=0.92$, with an overall mean of $r=0.66$. Similarly, Südkamp, Kaiser and Möller (2012) found that the overall mean of correlations between teachers’ assessment and students’ achievement was $r=0.63$.

Different explanations can be given for the discrepancies between the tutors' assessment and the raters' ratings. One explanation is that though the tutors are following criterion-referenced assessment as it is usually the case when using the CEFR scales (Fleckenstein *et al.* 2018; Hughes, 2002), there is still the possibility that the tutors tended to compare the students within their classes (norm-referenced assessment) (Fleckenstein, *et al.* 2018; Lok, McNaught, & Young, 2016).

Also, the raters were focusing on specific aspects (specific texts) which may be easier to judge than students' proficiency in general (as for students and tutors using the CEFR scales) (Fleckenstein *et al.*, 2018; Südkamp *et al.*, 2012). In addition, the task that elicited the texts rated by the raters has its own limitations. As mentioned previously, it was designed based on a very low benchmark (A2). It required students to write on an easy, general, descriptive topic which did not allow students to show their actual ability and proficiency.

Also, the number of the texts rated by raters was very small (10% of the total samples). Additionally, the CEFR rating grid used by raters had limitations: due to time constraints, the grid used was exactly as it appeared in the *Manual* (CoE, 2009) and it was impossible to modify and adapt it to make it more suitable for rating the task, as recommended by Harsch and Martin (2012). Although there was a two-hour benchmarking session with the raters, who already had experience of rating L2 texts using similar scales, more time dedicated to training could have increased the reliability of their ratings. In addition, as mentioned by Huhta *et al.*, (2014), "task performance can be assessed according to its communicative adequacy, i.e. on how well the learner is able to use language to accomplish task requirements". The task rated in this study, however, was not suited to the students' levels (especially the ones in the advanced levels) and thus might not have elicited the students' actual ability.

Based on the findings, and despite the way in which the CEFR scales were used by participants to assess students' proficiency levels, and the limitations identified, there are indications enough to argue for the usefulness of the CEFR employed in the manner reported here to identify students' proficiency levels. Considering that the participating students and tutors had not been extensively

trained in using the CEFR scales to identify students' proficiency levels in writing, the findings are similar to those found in the literature (especially the ones in relation to underestimation and overestimation of self-assessment). As mentioned in Moonen *et al.* (2013), many people have little experience and exposure to the use of the CEFR scales and as suggested by Davis (2015), Fahim & Bijani (2011), Fleckenstein *et al.* (2018), and Weigle (1994), with proper instruction and training, the tutors (and even the students) would be more accurate in their assessment.

Students were motivated to assess themselves; they were reassured that their assessment would not affect any of their marks and would only be used for the research purposes. This mitigated against the possibility of deliberately giving inaccurate assessments of abilities. Furthermore, it was felt that the students would try their best to accurately assess their levels of proficiency in return for receiving a report summarising the average result from their proficiency based on their self-assessment (see Chapter 3). Also, the way the CEFR scales were formatted for this study (see Chapter 3) helped students in their self-assessment, and while there was no proper training conducted to improve the reliability of assessment (Harris 1997; Little 2002; Ross, 1998) nor experience in self-assessment (Engelhardt & Pflingsthor, 2013), detailed instructions were given to both students and tutors. As noted by Paris and Winograd, (1990), proper instruction improves the accuracy and reliability of self-assessment. Furthermore, using the CEFR scales, which are based on what learners 'can do' with language (CoE, 2001), may improve the reliability of the findings, as using functional language (i.e. 'can do' statements) has been found to increase the accuracy of self-assessment (Ross, 1998). More importantly, and something that strengthens the reliability of the current study, is the combination of different methods to identify the students' CEFR levels (self-assessment, test scores [raters' ratings] and tutors' assessment) (Abdulhaleem & Harsch, 2018; Ashton, 2014; Babaii *et al.* 2016; Boud, 1991). In addition, the way the tool was designed and formatted (which I called *Controlled Utilization of CEFR Descriptors*) facilitated the process of assessment. Students were instructed to read the descriptors for each level and then tick one of two options (*Yes, I can do* or *Not Sure*). Where they ticked '*Yes, I can*', they would move to the next level and do the same, and so on.

If they reached a level where they were not convinced that they actually could do what was mentioned in the descriptor, they ticked '*Not Sure*' and moved to the next scale. I believe that presenting descriptors in this way, with detailed accompanying instructions, helped participants to be more accurate in their assessment. Thus, we can assume that the CEFR scales can be used effectively in contexts where assessors have relatively little experience of using the CEFR, provided they are equipped with sufficiently detailed instructions, and the CEFR scales and descriptors are presented in such a way that the assessors' focus is guided towards analysing every single descriptor in a meaningful way, i.e. moving up the levels.

4.4.2. RQ2: Writing at the PYP: discussion of the qualitative findings

It is insufficient to consider students' writing proficiency only with reference to the CEFR scales and levels. As recommended by its authors, the CEFR scales should be used as a guide and should be adapted to suit the contexts where they are used (CoE, 2001). The second research question investigated the writing skills participants perceived students had achieved by the end of the intensive English course in the PYP.

The qualitative data showed that writing is not being taken seriously at the PYP by the participants of the study (coordinators, tutors and students). This finding is consistent with that of Alkubaidi (2014) who found that writing in the preparatory year "is [being taught as] a secondary [skill] in comparison to how the other skills are taught" (p.189).

Writing at the PYP mainly encompasses general topics. Although academic and medical writing are part of the PYP curriculum, in reality, limited attention is given to improving these two genres. Little time is given to practising writing on medical topics or genres and there was no evidence of assignments, tasks or homework activities being given to students to help improve their academic writing. In line with the findings of the current study, the literature suggests that most L2 writing tasks in foundation programmes focus on 'non-academic writing'

(see, for example, Reichelt's review, 1999), for example "essays on hobbies, family life, friends, holidays, and personal letters" (Kuiken & Vedder, 2008, p.28).

In the current study, the PYP's main emphasis is on teaching the mechanics of writing such as how to start a paragraph, the organization of a text, the use of punctuation (commas, full stops and capitalizations), spelling and grammar. These 'basic traditional skills' in teaching writing were also found to be the main focus of writing in other universities in Saudi Arabia (such as King Abdulaziz University's preparatory year) (Alkubaidi, 2014). Not much attention, however, was given to teaching students more complex skills, such as summarising, paraphrasing and synthesising different resources and writing using their own words and ideas. It seems that most traditional L2 writing classes only focus on teaching basic writing skills (Lee, 2013).

There are several possible explanations for this limited focus on writing in general and on medical and academic writing skills in particular. First, the PYP curriculum, as described by participants, is following a 'plunge system' where tutors must cover a lot of material in a very limited time. This "lack of time and the pressures of finishing the curriculum within a designated timeframe" is one of the obstacles to improving students' writing (Alkubaidi, 2017, p.243). Second, writing is *'not called on... [it is] not something institutionalized that students must have developed'* (COO2_PHI_FG8_99-100). Because there was limited practice, time and less weight given to writing in assessment, participants viewed it as less important and not a priority in terms of the time and effort they put into teaching writing. The question, then, is: why is writing not *'called on'*? [COO2_PH1_FG8_100-103]. Why it is on the *'back burner'* [COO4_PH1_FG9_131-132] as perceived by most of the participants? Before mentioning the potential reasons for this lack of focus on writing at the PYP, it is important to mention that the students have shown willingness to improve the skills related to academic and medical writing. They, additionally, showed concern for future success because of these limitations especially regarding medical writing and terminology.

However, the most likely reason, supported by the data, was that students and their tutors did not prioritise the importance of writing at the PYP. The most important outcome for the students at the PYP was to gain high grades by the end

of the programme to enable enrolment into the medical college of their choice. The same was found in Al-Wassia *et al.* (2015), where students saw achieving a high 'grade as the main priority' (p.16) and therefore they would focus on whichever skills featured in the final PYP exam.

This leads one to ask why writing does not carry more weight in assessment and thus become more of a priority. Why are medical and academic skills not included in assessment, thereby requiring students and tutors to give them more attention? One possible explanation is that the students are used to memorising texts in order to write, and "focus on test results rather than learning" (McMullen, 2014, p.137). Students may have joined the PYP with "exam-centered" behaviours already established and focused only on passing the exam and scoring high grades (Javid *et al.*, 2012, p.65). These exams were usually very easy and the students were not used to undertaking (more) challenging writing tasks. Another possible explanation is that writing for medical and academic purposes requires higher levels of proficiency (Kroll, 1990; Hyland, 2003), and the data suggests that the students, in general, have low levels of writing proficiency, the highest levels being between B1+ and B2 and the remainder reflecting a much more basic level of proficiency. This generally low level of proficiency may possibly discourage both students and tutors from attempting these types of writing. In addition, including skills which require a high level of proficiency in the exam is, from the participants' points of view, 'unfair'. According to them, including these skills in the exam would put the students at risk of losing marks, which would affect their GPA, and, so their future admission to the university colleges. They continuously complained to upper management that the exam was 'not fair' and not suited to their level; consequently, and in order 'to be fair', the exam was designed and benchmarked to lower criteria (A2) so 'even elementary-level students could pass the exam' [CO06_PH1_FG7_140]. This concept of fairness has led to the exam being 'very easy' as indicated in the analysis of the data.

The PYP uses a standardised exam system where students across the three PYP levels sit the same exam. The students' fear of losing marks, which would negatively affect their GPAs, has led to them demanding easier assessment. Demanding easier assessment could also be due to the traditional methods that

are followed in English language education in Saudi schools (Alnassar & Dow, 2013). According to Alnassar and Dow (2013), these methods trained learners to be passive where their learning is based on “memorizing information to be retrieved only to pass examinations”, hence, the students “focus on test results rather than learning” (McMullen, 2014, p.137).

Having very easy exams for all the students has not only steered the students’ attention away from improving their writing skills, but also caused a ceiling effect. Unfortunately, this easy assessment had negatively impacted the students’ progress in writing. This is a very important point that needs to be taken more seriously. As argued in Lewin and Dunne (2000), when the pressure to score high grades becomes intense, the quality of education becomes questionable. And when high-stakes tests create pressure, this can lead to negative washback and steer the focus away from meaningful learning to the mere memorisation of facts for the purpose of passing exams (Lewin & Dunne, 2000).

Academic writing, which can be assessed through other measures (e.g. as part of an assignment or as a class/homework task), was still not considered by PYP stakeholders sufficiently in the PYP curriculum for the same reasons mentioned above. Academic writing was also considered a difficult skill. This is consistent with Fageeh’s (2003) finding that students had difficulties dealing with academic writing. Another reason for the lack of attention given to academic writing was students’ tendency towards plagiarism (El Tantawi *et al.*, 2016). It was found that whenever students were assigned a writing task, they plagiarised. Plagiarism seems to be a common problem among Middle Eastern students (see, for example, Handjani & Habibzadeh, 2013). According to Handjani & Habibzadeh (*ibid.*), plagiarism in the Middle East is “due to a lack of linguistic expertise” and the fact that “in some cultures, it is not considered misconduct”. In this current study, I found the main reason for plagiarism was that the students were not trained in how to write academically and avoid plagiarism. As Alhojailan (2015) found, the student might have a limited understanding of the meaning of academic writing and its main purposes. Strict rules concerning plagiarism need to be put in place in the PYP and clearly explained to students, along with strategies for avoiding plagiarism.

The little attention given to improving the different writing skills leads one to doubt whether the proficiency levels identified by the students, tutors and raters are achieved as a result of joining the PYP, or whether these identified levels represent the entry rather than exit levels of the students. With little focus on the teaching, learning and assessment of writing at the PYP, it is perhaps unreasonable to expect progress. Having no progress and entering different “academic majors” without the necessary language skills was also identified in McMullen (2014, p.137) who studied the effect of PYP in three different universities in Saudi Arabia.

The other concern is whether the PYP has prepared the students to meet their medical and academic writing needs as expected. I am raising these two concerns here because, first, not much focus, time and practice were dedicated to writing, which makes me doubt the students’ progress. The second is the dissatisfaction found among participants, especially the students, with their progress and development in writing that was obvious in the data. Many students (especially in the advanced level) mentioned that they feel their writing at the end of the PYP was similar to when they first joined. However, it is hard to confirm this point because writing was not a component of the placement test at the start of the PYP. Thus at present it is impossible to compare the students’ proficiency at the points of PYP entry and exit.

With the current assessment methods in the PYP, students at different entry levels appear to join the medical colleges with similar proficiency based on their marks in the exit exam, but the reality is different (as identified from the range of levels achieved in this study’s data). So it is important to have a clearer policy with specific criteria describing exactly what students need to achieve to enrol into these colleges. This problem was also identified by Javid *et al.*, (2012) who recommended “that the universities [in Saudi Arabia] should follow a strict admission policy and only those students [who meet the strict policies] should be allowed admission” (p.56). They also recommended relying on international tests for admission instead of ‘in-house examinations’ (Javid *et al.*, 2012, p.65). Though I do not completely agree with their recommendations, I agree that the admission policy to specific colleges should be modified and transparently

conveyed and explained to all stakeholders at the university and most importantly at the PYP.

Tutors and coordinators have fewer expectations of elementary-level students enrolling into the more competitive medical colleges. If they do indeed have a lower chance of being accepted, then why would the PYP curriculum and assessment cater for and accommodate the proficiency of a group of students who are not going to join the medical colleges? If this is true, then, and as questioned by some of the coordinators, why should they be enrolled from the start into the MT? Or from my point of view, and if we want 'to be fair', those elementary-level students should be challenged and supported more to meet the requirements to enrol into the medical colleges, rather than the PYP lowering the standards to accommodate their low levels of proficiency.

Chapter 5

A Target-situation Analysis

Phase II

5.1. Introduction

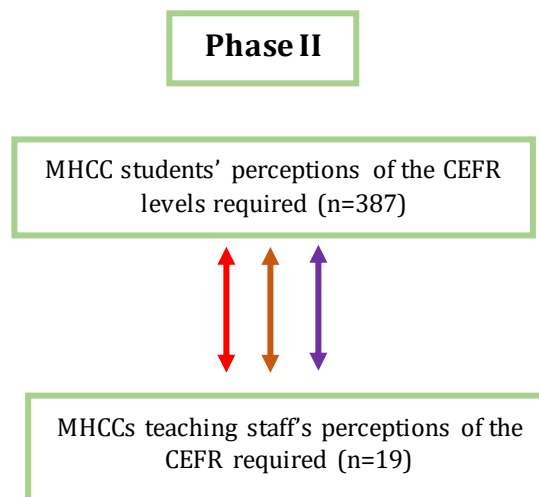
Following a target-situation analysis approach (Brown, 2016), this chapter presents the findings from the Phase II quantitative and qualitative data. The aim of this phase is to identify the students' CEFR writing levels and the writing skills required of first-year students at medical colleges in a Saudi Arabian university. Towards the end of students' first-year in college, students and teaching staff were asked to use the CEFR scales to quantitatively identify the CEFR levels minimally required by students in this year. Again, they were given detailed instructions on how to use the scales. Qualitative data, using FGs and interviews, were then analysed in depth to better understand students' writing requirements in their first year of college and to triangulate the findings from the quantitative data. Thematic analysis was used to identify the themes that emerged from the data after coding (Harding, 2013; Saldaña, 2015). The discussion at the end of this chapter unites quantitative and qualitative findings of this phase to answer its research questions 3 and 4.

5.2. The CEFR levels required of first-year MHCC students

A total of $n=387$ students from the five MHCCs (*CAMS*, *CD*, *CM*, *CN* and *CPH*) and 19 teaching staff members from three colleges (*CM*, *CN* and *CPH*) participated in the quantitative part of Phase II of the study.

RQ3 aims to identify the CEFR level(s) perceived by students and teaching staff as required in first-year MHCCs. To answer this question, the following analyses were conducted and visualised in Figure 5.1:

- a. Descriptive analysis to identify the CEFR writing levels minimally required as perceived by the students and their teaching staff in the first-year of college.
- b. Independent sample t-test to examine the differences between the students' and the teaching staff's perceptions of the CEFR levels required.
- c. One-way ANOVA to explore the differences in the CEFR levels perceived across the medical colleges.
- d. One-way ANOVA to explore the CEFR levels perceived as required across the students' three PYP proficiency levels.



Red line = overall comparison

Orange line = comparison across PYP baseline levels (elementary, intermediate, advanced)

Purple line = comparison across colleges (CM, CD, CPH, CN, CAMS)

Figure 5. 1 Visualisation of the quantitative analyses conducted in Phase II

5.2.1. Descriptive analysis of the CEFR levels perceived as required

The CEFR levels perceived as minimally required in the first-year in college in each of the ten CEFR scales and the average are summarized in table 5.1.

Table 5. 1 Summary of the CEFR levels students and teaching staff perceived as being minimally required at MHCCs

CEFR Scales	Medical Students <i>n</i> =374		Teaching staff <i>n</i> =19	
	M	SD	M	SD
Overall Written Production	6.91	1.94	5.89	1.41
Overall Written Interaction	5.14	2.08	4.95	1.68
Type of Texts	5.36	2.26	6.47	2.44
What Can They Write	5.48	2.25	6.32	1.89
Vocabulary Range & Control	5.19	1.95	5.63	1.89
Grammatical Accuracy	5.82	1.81	5.47	1.65
Orthographic Control	5.76	2.12	5.63	1.68
Processing Texts	5.18	1.97	4.47	1.35
Reports and Essays	5.92	1.34	5.84	1.50
Note Taking	5.89	1.64	5.53	1.39
Average CEFR Level	5.59	1.28	5.62	1.00

M= Mean, SD=Standard deviation

Coding scheme: 1 (A1); 2 (A2); 3 (A2+); 4 (B1), 5 (B1+); 6 (B2); 7 (B2+); 8 (C1); 9 (C2)

The average CEFR level(s) required of students in the medical colleges ranged between levels B1+ and B2.

Because of the limited number of staff participating in this phase, the results should be treated with caution. For the same reason, only an overall analysis of the staff data was conducted (see below) and there was no analysis conducted across colleges and PYP levels for the staff data.

5.2.2. Comparing students and teaching staff's perceptions of the CEFR levels required

It is important to examine how students and staff perceived the minimally required levels and whether they differ. The staff's and students' perceptions of the CEFR levels are considered as two independent samples. To compare the mean differences in these two groups for each scale, an independent two-sample *t*-test was conducted, assuming equal variances of the two groups. The sample sizes are vastly different in the two groups, (i.e. unbalanced samples). Levene's *F* test (Levene, 1960) was used to test the equality of variances between the two groups (students and staff). The variances of all the CEFR scales were not significantly different between students and staff, except for the *processing texts* scale. Therefore, a Welch's two sample *t*-test with unequal variances was performed for the *processing texts* scale. A negative difference implies that students' assessment of required CEFR levels is higher than the staff's.

Table 5. 2 Differences between students' and staff's perceptions of CEFR levels required at MHCCs

<i>CEFR Scales</i>	Staff n=19		Students n=245		<i>t</i>	<i>df</i>	<i>P</i>	<i>Cohen's d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>				
Overall Written Production	5.89	1.41	6.11	1.94	-0.48	262	0.63	-0.06
Overall Written Interaction	4.95	1.68	5.14	2.08	-0.40	262	0.69	-0.049
Type of Texts	6.47	2.44	5.36	2.26	2.07	262	0.04	0.26
What Can They Write	6.32	1.89	5.48	2.25	1.58	261	0.12	0.20
Vocabulary	5.63	1.89	5.19	1.95	0.95	261	0.34	0.12
Grammatical Control	5.47	1.65	5.82	1.81	-0.81	262	0.42	-0.10
Orthographic Control	5.63	1.68	5.76	2.12	-0.25	260	0.80	-0.03
Processing Texts*	4.47	1.35	5.18	1.97	-2.10	24	0.046	-0.86
Reports and essays	5.84	1.50	5.92	1.34	-0.25	261	0.35	-0.03

Note Taking	5.53	1.39	5.89	1.64	-0.94	261	0.35	-0.12
Average CEFR Scales	5.62	1.00	5.59	1.28	0.11	262	0.91	0.014

M=Mean, SD=Standard deviation

Coding scheme for CEFR Scales: 1 (A1); 2 (A2); 3 (A2+); 4 (B1), 5 (B1+); 6 (B2); 7 (B2+); 8 (C1); 9 (C2)

Cohen's-d calculated as $2 \times t / \sqrt{df}$; *t-test carried out without the assumption of equal variances.

Cohen's-d calculated as $2 \times t / \sqrt{df}$, 0.2=small effect; 0.5=medium; 0.8=large

There are no significant differences except for the *type of texts* and *processing texts* scales, indicating mostly close similarity between the students' and staff's perceptions of CEFR levels required, although it is important to treat the results with caution as the sample of staff is small and this may reduce the power of the study to detect small differences between students and staff (Jones, Carley & Harrison, 2003). Staff's perceptions regarding the *types of texts* are significantly higher than students' perceptions while staff perceptions regarding the *processing texts* scale are significantly lower than the students' perceptions.

5.2.3. Perceptions of the CEFR levels required across MHCCs

The levels students perceived as required were next broken down by the five colleges (Table 5.3).

Overall average mean scores ranked from the CM ($M=6.02$, the highest), through the CD ($M=5.85$), CAMS ($M=5.40$), CPH ($M=5.28$) to the CN ($M=4.69$, the lowest). Although the ranking of the top four colleges varied between the different scales, the CN scored lowest on all scales.

Table 5.3 Descriptive analysis of the CEFR levels perceived as required by students across MHCCs

College	CM n=81		CPH n=76		CN n=64		CD n=52		CAMS n=101	
	M	SD	M	SD	M	SD	M	SD	M	SD
Overall Written Production	6.59	1.93	6.37	1.88	4.73	2.55	6.29	1.63	6.16	1.98
Overall Written Interaction	5.49	1.92	4.58	1.89	4.42	2.38	5.54	2.04	5.16	2.18
Type of Texts	5.79	2.23	4.72	2.18	4.49	2.78	5.98	2.04	5.10	2.36
What Can They Write	6.21	2.16	5.12	1.99	4.23	2.36	5.65	1.88	5.30	2.60

Vocabulary Range & Control	5.92	1.80	5.03	1.97	4.56	2.36	5.17	1.63	5.15	2.07
Grammatical Accuracy	6.41	1.53	5.68	1.94	4.89	2.26	5.69	1.48	5.51	2.04
Orthographic Control	6.18	2.33	5.08	1.90	4.48	2.40	6.17	2.04	5.58	2.26
Processing Texts	5.54	2.18	5.00	1.74	4.20	1.69	5.75	2.07	4.89	1.87
Reports and Essays	6.21	1.24	5.59	1.29	5.47	1.37	6.15	1.49	5.90	1.28
Note Taking	6.39	1.67	5.61	1.54	5.33	1.50	6.08	1.69	5.51	1.52
Average CEFR Level	6.02	1.33	5.28	1.15	4.69	1.52	5.85	1.07	5.40	1.42

M=Mean, SD=Standard deviation

To test the differences by college, a one-way ANOVA was used (Table 5.4). For those variables where the homogeneity of variances assumption was violated, alternative robust tests are provided in Table 5.5. There were statistically significant differences between the colleges for all scales.

Table 5. 4 One-way ANOVA of CEFR levels perceived as required by MHCC students across MHCCs

CEFR Scales	SS	df	MS	F	P-value	η^2
Overall written interaction						
Between Groups	69.48	4	17.37	4.00	0.004	0.043
Within Group	1564.25	358	4.37			
Total	1633.73	362				
Orthographic Control						
Between Groups	137.36	4	34.34	7.08	<0.001	0.073
Within Group	1735.34	358	4.85			
Total	1872.69	362				
Reports and Essays						
Between Groups	28.59	4	7.15	4.10	0.003	0.044
Within Group	623.96	358	1.74			
Total	652.55	362				
Note Taking						
Between Groups	54.72	4	13.68	5.50	<0.001	0.057
Within Group	898.64	361	2.49			
Total	953.36	365				
Average CEFR Levels						
Between Groups	73.03	4	18.26	10.42	<0.001	0.102
Within Group	641.30	366	1.75			
Total	714.33	370				

SS=Sum of squares, df=degrees of freedom, MS=mean square, F=F ratio, η^2 =Effect size: 0.02=small; 0.13=medium; 0.26=large.

Table 5. 5 Robust test of mean of CEFR levels perceived as required by MHCC students across MHCCs

CEFR Scales	Statistic	df1	df2	P-value
Overall Written Production				
Welch	6.08	4	166.87	<0.001
Brown-Forsythe	8.76	4	301.18	<0.001
Type of Texts				
Welch	5.08	4	168.55	0.001
Brown-Forsythe	4.94	4	323.79	0001
What Students Can Write				
Welch	7.00	4	170.51	<0.001
Brown-Forsythe	7.35	4	347.89	<0.001
Vocabulary Range & Control				
Welch	4.11	4	168.50	0.003
Brown-Forsythe	4.18	4	323.66	0.003
Grammatical Accuracy				
Welch	6.07	4	170.92	<0.001
Brown-Forsythe	5.86	4	321.57	<0.001
Processing Texts				
Welch	6.44	4	166.31	<0.001
Brown-Forsythe	6.27	4	312.98	<0.001

df=degrees of freedom

Further analyses were used to locate those pairs of colleges that showed significant differences (*post-hoc* pairwise comparison tests in Appendix C Table C1 using Tukey's HSD test where the assumption of homogeneity was met and Table C2 using Tamhane's test where the assumption was violated). CN appeared to be the most different from the other colleges, with significant differences in all ten scales and the overall average compared with MC; on six scales and the overall average versus CD; on two scales and the overall average versus CAMS and on one scale versus CPH. Other pairs were more similar; for example, no scales were significantly different between MC and CD, or CD and CAMS, or CAMS and CPH.

As mentioned previously, comparing the staff perceptions of the CEFR levels required across colleges was not possible due to the limited number of participants ($n=19$) from only three colleges out of five.

5.2.4. Perceptions of the CEFR levels required across students' PYP levels

The same analyses were repeated comparing mean CEFR levels perceived as being required across the PYP levels (*elementary, intermediate and advanced*) (Table 5.6). Students from the advanced level perceived that higher CEFR levels were required for all scales and the overall average.

Table 5.6 Descriptive analysis of the CEFR levels perceived as required by MHCC students across the PYP levels

CEFR Scales	Elementary, n=48		Intermediate, n=186		Advanced, n=141	
	M	SD	M	SD	M	SD
Overall Written Production	5.15	2.51	5.81	2.17	6.71	1.60
Overall Written Interaction	3.91	1.95	4.89	2.09	5.60	2.04
Type of Texts	4.11	2.24	4.95	2.36	5.86	2.28
What Can They Write	4.06	2.13	5.19	2.38	5.90	2.16
Vocabulary Range & Control	4.57	2.14	5.06	2.06	5.53	1.88
Grammatical Accuracy	5.15	1.86	5.34	2.06	6.21	1.68
Orthographic Control	4.57	2.03	5.24	2.22	6.14	2.25
Processing Texts	4.20	1.63	4.80	1.87	5.64	2.01
Reports and Essays	5.36	1.03	5.68	1.34	6.28	1.34
Note Taking	5.27	1.24	5.57	1.50	6.22	1.72
Overall Average	4.73	1.42	5.23	1.31	5.97	1.30

M=Mean, SD=Standard deviation

One-way ANOVAs and robust tests were used to compare the means across the three PYP levels (Table 5.7 and 5.8, respectively).

Table 5.7 One-Way ANOVA of CEFR levels perceived as required by MHCC students across PYP levels

CEFR Scales	SS	df	MS	F	P-value	η^2
Type of Texts						
Between Groups	125.86	2	62.93	11.721	<0.001	0.061
Within Group	1927.37	359	5.37			
Total	2053.23	361				

Vocabulary Range and Control						
Between Groups	36.46	2	18.23	4.53	0.011	0.025
Within Group	1428.80	355	4.03			
Total	1465.26	357				
Orthographic Control						
Between Groups	108.70	2	54.35	11.11	<0.001	0.059
Within Group	1741.05	356	4.89			
Total	1849.75	358				
Note Taking						
Between Groups	46.74	2	23.37	9.383	<0.001	0.050
Within Group	894.14	359	2.49			
Total	940.88	361				
Average CEFR Levels						
Between Groups	71.49	2	35.74	20.43	<0.001	0.101
Within Group	636.98	364	1.75			
Total	708.47	366				

SS= Sum of squares, df= degrees of freedom, MS= mean square, F=F ratio, η^2 =Effect size: 0.02=small; 0.13=medium; 0.26=large.

Table 5.8 Robust test of equality of mean of CEFR levels perceived as required by MHCC students across PYP levels

CEFR Scales	<i>Statistic</i>	<i>df1</i>	<i>df2</i>	<i>P-value</i>
Overall Written Production				
Welch	13.60	2	117.16	<0.001
Brown-Forsythe	11.26	2	124.07	<0.001
Overall Written Interaction				
Welch	13.21	2	128.43	<0.001
Brown-Forsythe	12.98	2	204.17	<0.001
What Students Can Write				
Welch	13.33	2	132.20	<0.001
Brown-Forsythe	12.67	2	211.25	<0.001
Grammatical Accuracy				
Welch	11.00	2	129.73	<0.001
Brown-Forsythe	10.37	2	194.30	<0.001
Processing Texts				
Welch	13.54	2	133.41	<0.001
Brown-Forsythe	14.09	2	238.15	<0.001
Reports and Essays				
Welch	13.46	2	143.02	<.0001
Brown-Forsythe	13.83	2	269.55	<0.001

df=degrees of freedom

There were significant differences between the three PYP levels, so *post-hoc* pairwise comparisons were conducted using Tukey’s HSD or Tamhane’s test, as appropriate (Appendix C Tables C3 and C4, respectively).

There was a significant difference in the students’ perceptions of the CEFR levels required when comparing the advanced level versus the intermediate level (all but one scale) or versus the elementary level (all scales). However, differences were only significant between the perceptions of students at the intermediate and elementary levels for two scales.

The lowest perception of CEFR levels required was found in the CN and among the elementary level students. To examine possible connections between students at the elementary level and those enrolled at the CN, a cross-tabulation between PYP levels and medical colleges was performed (Table 5.9). Around 39% of the students in the advanced levels enrolled at the CM, with only 2% in the CN. On the other hand, most of the students in the elementary levels enrolled either at the CN (24%) or CAMS (48%), which potentially explains the differences across the three PYP levels. Visual representations of participants in this phase can be seen in figure 5.2 below. Few elementary level students participated in this phase and results; therefore, results need to be treated with caution.

Table 5.9 Students PYP Levels * MHCCs Crosstabulation

	CM	CPH	CN	CD	CAMS	Total
Elementary level	1	4	5	1	10	21
Intermediate Level	18	34	13	18	45	128
Advanced Level	40	17	2	24	20	103
Total	59	55	20	43	75	252

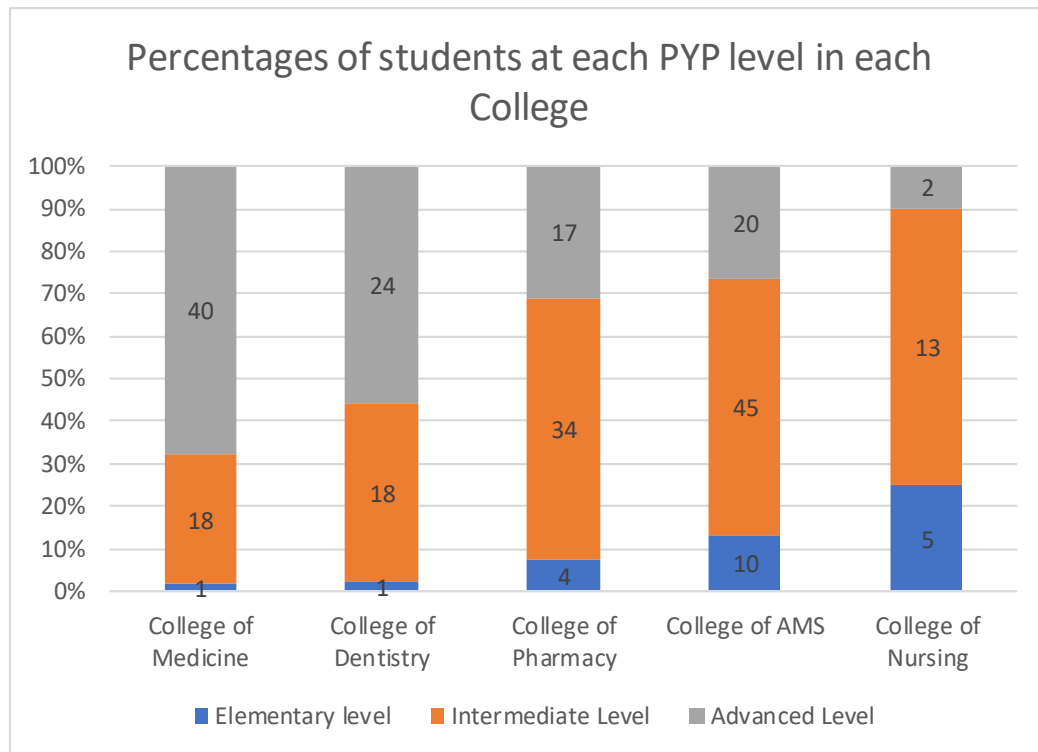


Figure 5. 2 Percentage of students participating in Phase II from each college across the students' PYP levels

In summary, the first part of this chapter reported the CEFR level (i.e. B2) perceived as being the minimum required in the first-year of college by both MHCC students and their teaching staff. Apart from the CN, where students were significantly perceived as having lower CEFR levels compared to the other colleges, there were few differences between them, indicating similarities in the levels required across colleges. When the data were compared across PYP levels, there were significant differences mainly between the advanced-level students and elementary and intermediate levels. Students from the elementary level comprised 25% of students who joined the CN but only around 2% of those joining the CM or CD. Conversely, advanced-level students made up 68% of those in CM and 56% of those in CD, but only 10% of those in CN. This might explain the significant differences identified in the data.

5.2.5. The CEFR levels perceived as required by FG/interview participants

Medical students and staff during the FGs and the interviews were given the CEFR writing descriptors from the self-assessment grid (CoE, 2001, pp.26-27). For each level, the individual descriptors were presented in small boxes and participants were asked to agree on the one that represented the minimal level required of first-year students in the medical colleges. Participants were also requested to justify their choices (Table 5.10 summarises the findings). The quotations in the table are the consensus statements that participants agreed and wrote as justifications of their choices of required level.

Table 5. 10 Summary of the CEFR levels selected by participants during the FGs/interviews with justifications of choices

Students	CEFR Levels	Justifications	Staff	CEFR Levels	Justifications
<u>College of Medicine</u>					
FG1 GP1	B1	'We were required to write two essays at the beginning of the year but then there was no more writing' [TR]	Staff 1	B1	"Because the students can do this. The other ones are difficult. The students for example cannot write complete long reports" [TR]
FG1 GP2	B2	'We have needed what mentioned in here (B2 descriptors) for the learning skills subject' [TR]	Staff 2	B2	
FG2	B2	'because we need to write medical research, we need to have professional communication like sending and responding to emails, we need to write in the exam expressing our own understanding of the subject, we need to collect data and write reports and write summaries' [TR]			"It [the descriptors in B2] has reasonable representations of minimum expectations of the level [required]. The others are quite low or quite high" [TR]
FG3	B2	'because we need to convey the information in very clear and understandable way especially when talking about a specific disease, to write clearly especially when answering SAQ and to prove OUR point of view for choosing a case with medical evidences'			
<u>College of Pharmacy</u>					
FG1	B2	'In the college of pharmacy, we need to write information in form of evidences in a simple way that make it clear for the others to understand. It is also important to be able to write research and to give enough information stating our own opinion. For example, at the end of the poster' [TR]	Staff 3	C1	"These are the requirements. We hope they come and do analysis and critical appraisal" [TR]
FG2 GP1	B2	'We need these descriptors when answering exam questions. We need to convey our thoughts and be able to write our opinions' [TR]	Staff 4	C1	"We wish they come [to medical colleges] with this level. This is what we want them to do" [TR]
FG2 GP1	B2	"We need to be able to convey our opinions' [TR]	Staff 5	B1	"This one...I think is reasonable. The rest are very advanced for first-year students. They will learn these later" [TR]

College of Applied Medical Sciences

FG1 GP1	B2	“because these skills are needed in writing research” [TR]
FG1 GP2	C1	“Because we need to be able to write reports. We need to be able to state our opinion and to make our writing style suitable to different types of readers” [TR]
FG2	C1	“Because we need to write stating our opinions especially when writing medical reports. In first-year if college we needed to write research and to summarise articles and to write presentations” [TR]

College of Nursing

FG1	B1	“Because these are important for nursing students to prepare them for what they need” [TR]	Staff 6	C1	“Because we are letting them to analyse cases and we want them to write reports and essays”
FG1	B1	“We need basic terminology to be able to write coherent text” [TR]	Staff 7	B2	“Because when they write assignments, they need to make some analysis...the rest are too high or too low”
FG1	B1	“Because these are important for nursing students to prepare them for what they need” [TR]	Staff 8	B2	“Because we want students to have communications in writing...they need to convey messages through reports that could be critical for the patients’ care. The students need to form ideas and put them in writing to other to read” [TR]
			Staff 9	C1	“the students need to write reports and make an argument”

FG=Focus Group, GP=Group (pairs or threes) in the FG when working on activities; TR=Translated into English; SAQ=short answer question

Most participants chose B2 as the level minimally required for their first year in these colleges. The students needed to write clearly and provide evidence, particularly in answering exam questions. The justifications for those who chose B1 as the minimum required level was that the B1 descriptors reflect what students can actually do/write at this stage. B1 descriptors cover the basic needs, which many felt was all that was necessary given that writing is not heavily required in the first-year of college. On the other hand, those who selected C1 as the required level stressed their wish that students could meet this level to be able to write research papers and reports related to their subjects. It was also noticed that the students in the CN seemed to choose lower levels than the staff (similar to the quantitative findings), which differed from the other colleges where students and staff were more similar.

To conclude, the average CEFR level(s) perceived to be the minimum required in the first-year medical colleges was B2. Student and staff perceptions of the CEFR levels required were similar. Across the five colleges, the highest level of proficiency required was in the CM, whereas the lowest was in the CN. Interestingly, when the data were compared across PYP levels, it was found that the higher the students' language proficiency, the higher the perceived required level of CEFR proficiency. Through performing a cross-tabulation analysis between the PYP levels and the colleges, it was found that most of the advanced level students joined the CM and most of the students from the elementary level enrolled at the CN and the CAMS. However, not many students from the elementary level joined the MHCCs compared to students from the other levels and, thus, the results need to be interpreted with caution.

5.3. Writing at the MHCCs

This section aims to answer the study's RQ4. This question reflects my desire to understand in more depth the writing skills required of students in their first-year in college and to triangulate this with findings from the quantitative data. The findings (from Phase II) will then be compared in Chapter 6 with those from the previous chapter (in Phase I) to answer the main overarching research

question of the study. Therefore, in this part, I explored both students' and staff's views on what writing proficiency they considered to be necessary for first-year medical students, using data from the FGs, including students' mind maps (see Appendix A9(a), in the appendices), the interviews, and the open-ended questions from the Phase II questionnaire. Table 5.11, below, summarises the themes that emerged from the analysis of the data, discussed below.

Table 5. 11 Phase II themes: writing in first-year MHCCs

<p>Writing in the medical colleges: three points of view</p> <ul style="list-style-type: none"> - Writing for exam purposes: <i>highly required</i> - Writing for course tasks and assignments: <i>moderately/partially required</i> - Writing for future use: <i>marginally required</i>
--

5.3.1. Writing in the MHCCs: three points of view

For almost all participating teaching staff and students, writing was seen as an important skill for students in the medical field, including in their first year of college:

[W]riting is essential for the students ... especially if they undertake the medical courses, it is important to be skilful in writing [SF8_PH2_CN_67-77].

Teaching staff had high expectations in terms of the students' proficiency in writing after the PYP: "*we expect that they have this higher level [in writing]...because they're already in college*" [SF7_PH2_CN_177-186]. At the same time, staff in these colleges showed concern that students were failing to meet those expectations at entry to the MHCCs. These expectations and concerns were also confirmed by the students' themselves:

When you join the college, they don't consider what level of English background you had [at the PYP], and if, for example, you ask what 'abdomen' means, an instructor will laugh and say 'Are you sure you were in the medical track?' [ST7_PH2_CAMS_FG2_119-120].

In contrast, however, some students considered themselves as already having met the first-year writing requirements. For some of them, not much was required in terms of writing in the first-year: "[w]e do not need to write that much...mostly short answer questions" [ST713_PH2_CM_FG9_173-174]. For example, although students enrolled in the CPH identified writing skills as essential to both successful studies and their future practice, they did not feel that writing proficiency was emphasised or especially needed in the first year: "*some students feel writing is not part of what they do [in first-year] so they do not care about it*" [ST4_PH2_CPH_FG5_254].

Three clearly distinguishable purposes for writing were found in the data, for which different levels of expectation emerged, which could explain these two opposing views. Based on their perceived importance, these writing abilities were classified based on their perceived importance as: *highly required*, *moderately/partially required* and *marginally required*.

Regarding those writing abilities that are highly required, students need particular writing skills to answer written exam questions and to pass exams related to their subjects. For this reason, writing for this purpose was viewed as very important as it would affect their scores at the end of the year. The students were sometimes required to submit written assignments, reports and mini-research papers. However, although these tasks were considered important, less weight was dedicated to their assessment and they were infrequently collected; consequently, this type of writing was perceived as only moderately necessary. The third point was that writing is an important '*lifelong*' skill [SF2_PH2_CM_99] which will be required as part of the students' professional future needs. These abilities included writing for academic purposes and professional communication. However, because these skills were not immediately required by the students in the first year of college, they were perceived as marginally required for first-year students.

These three points of view on writing in the first-year in college and how they are perceived are further explained below.

5.3.1.1. Writing for Exam Purpose: *highly required*

Writing was important for the students in the first year of college to be able to answer mid-term and final exam '*short answer questions*' (SAQs) as part of their assessment in their medical subjects:

I need writing more than before because I need to write for the short answer questions [...] we didn't have this before. I need to be able to write in order to convey the information I have [ST11_PH2_CM_FG4_90-91].

According to the staff, the students minimally need

to write direct, clear and neat English [especially] when answering the SAQs and they need to write in a systematic way. Whatever they are writing... the students need to have an introduction, [a] body and a conclusion [SF2_PH2_CM 244-246].

The staff specifically identified the students' ability to "*form their ideas and put them into writing [that is] very simple and very clear*" as key to students' success [SF7_PH2_CN_100-101]. This meant that the students needed to be able to convey knowledge of their subject in writing. The students were minimally required to have the '*basic skills of writing*', the basic '*academic vocabulary*' and '*the basic medical terminology*' to answer exam questions. They needed to be able to express their opinions, argue points and express their thoughts in writing. Students agreed that they needed to be able to prove opinions about a specific point during exams, especially "to show why I choose to explain certain diseases and provide medical evidences" [ST3_PH2_CM_FG3_350].

Basic grammatical structure was considered necessary to write "*comprehensive, simple and clear structur*" [SF1_PH2_CM_43] when answering exam questions. The staff expected students to come to the college knowing "*how to form a*

complete sentence" [SF5_PH2_CPH_45]. Vocabulary, including medical terminology, was widely regarded as the most important writing skill for the first year of college. About 40% of the respondents to the open-ended questions singled out this item for special attention, stressing its role as a basic building-block for all other writing skills. One student explained, "*Vocabulary is the most important element and the other skills come after it*" [ST322_PH2_OQ]. Therefore, familiarity with medical terms and vocabulary was a pressing necessity for students in all the colleges. One student made a direct link between vocabulary and language mastery: "*if we have the medical vocabulary, I will be able to write any sentence.*" However, most students were struggling with their limited medical terminology and felt they were unprepared by the PYP curriculum. They claimed that academic staff expected them "*to come knowing at least the basic medical terminologies*" [ST200_PH2_OQ], yet the PYP vocabulary was "*irrelevant to their field of study.*" As one of the students in the CN remarked, "*all the words are new. We are only memorising new words more than studying for the subject*" [ST6_PH2_CN_FG7_125].

The staff also emphasised the need for exposure to a range of different general and medical vocabulary in the PYP:

The students need a mixture of different terminology...ones related to the field of pharmacy, for example, tablet, capsule, injection, pharmacology, blood...etc., and they also need medical terminology like [medical] names of diseases, pathology, histology, physiology, anatomy [SF5_PH2_CPH_146-150].

The students confirmed their need for vocabulary to be able to write in exams, stating for example that "*we need medical terminology a lot ... [W]e need them to answer short notes quiz, short answer questions and give-reasons questions.*" [ST8_PH2_CN_FG8_432]. The students in the CM likewise emphasised the expectation of breadth and high-level vocabulary:

We were supposed to be prepared with the basics and to have basic academic vocabulary, which will help us when we write [ST2_PH2_CM_FG3_363].

It is important for the students to have enough vocabulary regardless of their levels to be able to express their thoughts using different words [ST94_PH2_OQ].

Some students were more specific in their references to medical terminology: *“medical terms are very important in the first year of college as they make the students’ life here a lot easier”* [ST112_PH2_OQ].

These again were reflected in B2 descriptors where students have *“a good range of vocabulary for matters connected to his/her field”* (CoE, 2001, p.112).

Though these SAQs required short answers written in a few sentences, they still required a high level of proficiency. As seen from the analysis of the mind-maps, the interviews and the FGs, the answers to these questions required some advanced rhetorical modes ranging from listing and matching to analysing, summarizing, comparing, reflecting, and arguing on topics related to their medical fields. One participant described their need to answer exam questions as follows:

[w]e need to be able to present the information with clear evidences and to convey our ideas in a simple and understandable way that can be suitable...It is also important to be able to...state our opinions [ST9_PH2_CPH_FG6_75-76].

The students in the first year of college are generally asked to write on medical-related topics (see Appendix C Table C5 for a summary of related topics students wrote about in this year). At the same time, there were no common topics that could be identified among the colleges. This means that the students required both basic medical terminology and terms related to their field of study in each college. They also needed sound reasoning and good argumentative and descriptive skills that helped them respond to their writing tasks.

This reflects B2 descriptors in the CEFR’s overall written production scale, where a learner is expected to be able to *“write clear, detailed texts on a variety of subjects related to his/her field of interest, synthesising and evaluating information and arguments from a number of sources”*.

Students' grammar in most of the colleges was manageable. However, the students in the CN seemed to have more problems with their grammar. The students have a *"very weak sentence structure and sometimes they write the answers in Arabic ... because they know I can understand Arabic"* [SF6_PH2_CN_121-122]. This reflects the findings from the quantitative data where the lower-level students more often joined the CN than the other colleges. Staff agreed that a focus on mastery of the rules of basic grammar helps develop clear writing. A CPH staff member said, *"students need to write using correct grammar, but not to an expert level"* [SF5_PH2_CPH_138-139]. Few staff members at the CPH disagreed about the importance of grammar in their field. Some claimed that grammar was not necessary in their field, so they don't focus on grammar in the students' writing. Others were in complete disagreement: *"[T]he students should come to us knowing how to write in correct grammar"* [SF3_PH2_CPH_144-145]. Students and staff in the CM also acknowledged the importance of having *"clear and basic correct grammar"* in students' writing and felt that there was no need for *"complicated grammatical rules"*.

5.3.1.2. Writing for course tasks and assignments: *moderately required*

Some writing types and skills (e.g. academic writing, reflections, reports and poster presentations) were expected of the students in their first-year of college for course tasks and assignments. However, these skills, while important, were perceived as less necessary than those required for exam purposes. Almost all participating staff expected students to be able to *"write academically"* because it was an important skill for all university students. One staff member at the CPH said:

now we are in a world that requires every single university student be able to write correct scientific research. All our work in this speciality is based on research [SF4_PH2_CPH_142-143].

Another added that students in the medical fields need *"to be able to write academically; they need to know how to paraphrase, summarise...and write a*

literature under a course called academic writing” [SF2_PH2_CM276-277]. They specifically mentioned three academic skills necessary for student development: *‘paraphrasing’*, *‘summarising’*, and *‘citation’* as the main skills to be able to write *‘research papers’*. The students also need to be able *“to state their opinion about the different research and to support their arguments”* [SF4_PH2_CPH_253].

Though writing for academic purposes was perceived as important, the students were infrequently required to submit written academic assignments and subject-related reports, or to write short research papers. One staff member commented: *“We do not ask for too much writing here...It is more practical here...the students need very basic skills of writing”* [SF5_PH2_CPH_130-135].

It is unusual for students to be asked to write essays or reports in some of the first-year colleges. Interestingly, even with the limited involvement of first-year students in academic research writing, the students themselves, especially in CM, emphasized the overall importance of this writing purpose – especially *“advanced writing for research”* – and showed interest in developing academic skills such as *‘summarising’* and *‘paraphrasing’* and acquiring familiarity with the different genres related to their professional fields of study.

Although very little attention was paid to research and academic writing in the first-year MHCCs’ curriculum, almost all participants (students and staff alike) emphasised the importance of *“high level academic English”* writing.

5.3.1.3. Writing for future use: *marginally required*

The importance of being able to write for professional communication was considered by the participants in Phase II. For example, it was expected of students in the CM and CPH to be able to write *‘scientific research’* and *‘medical/clinical reports’* related to their subjects. In the CN, staff expected the students to be able to *“keep clear documentation and submit clinical reports”* [SF8_PH2_CN_134-136] and to write *‘reflective essays’* on specific situations. Students also need to acquire general writing skills to *“write about the patients*

and...to give scenarios and stories" [SF8_PH2_CN_34-35]. However, none of the above were required as part of the students' progress or assessment in the first-year of college.

Students need writing for their profession...they are going to write [patient's] history and write about the case [SF2_PH2_CM_104-105].

[W]riting is very important later on when students need to write a proposal for an application or write a personal statement...they will later need to write research as well [SF4_PH2_CPH252-253].

To summarize, in general, students in their first-year of college were required to write complete and clear sentences to convey their knowledge of the subject, especially when answering exam questions. They were expected to use a varied range of medical terminology (basic and related to their fields) and to be able to express their opinions and put their acquired information in writing, particularly when answering short answer questions and short essays, or when submitting assignments related to the medical topics of their subjects. A range of vocabulary and basic medical terminology was highly required and expected of students in the first year of college. Though not much is required in terms of written assignments and tasks in students' first-year in college, the students still needed some academic writing skills when completing short answer questions (SAQs) in exams. Looking at the rhetorical modes identified from the students' responses to the brainstorming activity collected from the FGs, students from all participating colleges needed to be able to write an analysis – for example, *"analysing patients' disease"* – and to write an *"analysis of a situation"*. They needed to write a *"description"*, *"give reasoning"*, *"summarising"* and *"paraphrasing"* as part of their answer to exam questions or as part of the written assignments (although these were rarely requested).

5.4. Phase II Discussion

Unlike the previous phase, in which participants identified the perceived CEFR levels of students at the end of the PYP, this phase enabled participants to select the levels that they believed were minimally required for the first year of college. This is an important step, as it identifies the minimum required CEFR levels and skills in the 'target situation' and allows comparison with the 'present situation' at the PYP, a process that will be discussed in the next chapter. The discussion of this phase's results follows the order of the research questions.

5.4.1. RQ3: The minimum CEFR levels required of MHCCs

Looking at the average results from all ten CEFR scales, most first-year MHCC students and staff identified the CEFR level B2 as the minimally required levels.

Interestingly, there were few significant differences between students and staff regarding CEFR levels required; the *Types of texts students should be able to write* and *Processing texts* scales. Staff perceptions on the *Types of texts* scale were significantly higher than students' and significantly lower on the *Processing texts* scale. Connecting these findings to the qualitative data, I found that the staff were not satisfied with the students' writing proficiency. They indicated their willingness to ask the students to compose different types of writing (e.g. research, assignments, reports, etc.), but they often avoid doing so to accommodate the students' generally low proficiency. On the other hand, because the students had not been asked to write different types of texts, they perceived lower requirement levels than their teaching staff. The difference in expectations between the students and staff is also documented in the literature. For example, Üstünlüoğlu *et al.* (2012) found that students considered some of the writing skills (e.g. paraphrasing and synthesising information when writing) as only complementary, whereas staff members considered these skills as necessary for the students. Some of the discrepancies found between students and staff could be due to the fact that staff (and policymakers at the colleges) were not explicit enough concerning language requirements. In addition, the staff's

accommodation to the students' low proficiency to a certain extent gave students the wrong (i.e. lower) perception of requirements.

Identifying B2 as the minimum required level is in line with Harsch's (2018) observation that B2 is used as the standard for entry to most universities, although usually without any empirical basis (Deygers, Zeidler, Vilcu, and Carlsen, 2018). This finding is also consistent with the findings of Carlsen (2018), who concluded that B2 could be used as a minimum entry level to universities. She found that a B2 level of proficiency enabled foreign students to succeed academically, whereas those whose proficiency was lower than B2 struggled with their academic subjects. B2 was also found by Harsch, Ushioda and Ladroue (2017) to be the perceived appropriate level for entrance to UK universities. So in this study, students' and staff's perception of the minimally required CEFR level was similar to that found in other contexts. This is important, especially as universities in Saudi Arabia need to meet international standards, specifically in colleges where English is the main medium of instruction (Al-Shehri *et al.*, 2013).

This is particularly critical as none of the medical colleges in Saudi universities have specified the English proficiency required for admission to medical colleges other than high scores in English courses at High School and at the PYP (Al Alwan *et al.*, 2013; Albishri, Aly, & Alnemary, 2012; Al Makoshi, 2014; Alshehri, 2001). This is important because the admission criteria are similar for different medical colleges in Saudi Arabia. Relying only upon students' high school and PYP scores is insufficient, as High school English courses were found to have little impact on improving the students' proficiency, especially writing (Al Makoshi, 2014). In addition, in the current study, the ceiling effect found in the students' results at the end of the PYP made it unhelpful in discriminating between students for the purpose of admission into these MHCCs.

To the best of my knowledge, no study has empirically identified the CEFR levels required of first-year college students in the Saudi context, although the CEFR has been introduced into their curriculum (see Chapter 1). Almost all studies that have investigated medical students' English requirements have discussed the need for '*higher proficiency*' than the students' current proficiency in order for students to be able to cope with university demands (Alfehaid, 2014; Alshehri,

2001; El Tantawi, Al-Ansarir, Sadafiand, & Al Humaid, 2016; Al-Eisa & Smith, 2013; Ghobain, 2014; Shukri, 2008). However, there is no clear definition offered of what a “*higher level*” of proficiency means; therefore, identifying B2 as the minimally required level not only maps students’ proficiency at the end of the PYP, but could also be used as “a key reference” (Harsch & Rupp, 2011, p.3) for what students should be able to accomplish at these MHCCs and those at other universities in Saudi Arabia.

Consideration of B2 as the minimum required level for admission to MHCCs would help PYP policymakers and curriculum designers design and plan their curriculum and assessments to prepare students to achieve this level. Because of the ceiling effect in the students’ current PYP-exit exam results, students join the MHCCs with similar exam results but with disparate proficiency levels. This was confirmed in the analysis of this phase’s data as well. The staff have noticed these discrepancies in the students’ proficiency, which did not align with what they expected.

Looking back at the identified CEFR levels of the students in Phase I, most students (especially elementary and intermediate level students) would join the medical colleges with proficiency lower than that minimally required. The students’ struggle with their writing in Phase II was also more pronounced in this phase’s qualitative data. Jiménez-Muñoz (2014) observed that students who join college after high school with levels between A2 and B1 “will find it impossible to cope with the linguistic demands of academic tasks” (p.30). This also explains most of the students’ dissatisfaction with their levels in this phase (Phase II), especially in their answers to the questionnaire’s open-ended questions. This insufficient level of proficiency of students joining medical colleges has been reported in several studies (e.g. Alfehaid, 2014; Alhossaynee, 2006; Ghobain, 2014; Shukri, 2008). Most of the students expressed difficulty, most notably with medical terminology and with their writing in general. They also expressed their need for additional support during their time at the college.

Grouping the students by colleges also yielded interesting results. There were few significant differences among students’ perceptions of the required CEFR levels between colleges, except the CN. A cross-tabulation between the PYP levels

and the colleges indicated that most of the advanced level students joined the CM, while most of the elementary level students enrolled at the CN and the CAMS. At the CN, the perceived requirements were significantly lower compared to the other four colleges. Interestingly, when the data were compared across PYP levels (to check if the students' proficiency has an effect on the students' choices of the required CEFR levels), the highest level of proficiency required was in the CM (where advanced-level students were more common), whereas the lowest CEFR level required was in the CN (where elementary-level students were more common).

One possible explanation for these differences (between the CN and the other colleges) could be the CN's lack of stringent writing requirements in comparison to the other colleges. However, from the analysis of the qualitative data, it was found that the students in the CN are still required to write assignments similar to or even more writing tasks than those required of students in the other colleges. The students at this college are expected to analyse and write about a case during the exam; they need to reflect on a situation and are asked to submit more written assignments compared to the students in the CM and CD. But looking at the data, students in the CN were mostly from the elementary level and their level was obviously weaker compared to that of the students in the other colleges.

There are three possible explanations for the lower CEFR level required in the CN in comparison to the other colleges. First, because of the students' low proficiency (most of the students having come from the elementary level), the staff did not focus much on the linguistic areas of students' writing (as was also evident from the analysis of the qualitative data). In other words, the staff members were trying to accommodate the students' proficiency by exercising more leniency in their requirements. This was clearly indicated, both in the CN and elsewhere in the qualitative data, where staff tended to lower the writing requirements; hence the students underestimated the required language CEFR level. This also became clear when, for example, one of the staff in the CN mentioned that she even accepted students' answers that were written in a mixture of English and Arabic. The second possibility is that the students,

because of their low proficiency, had difficulty accurately judging the levels required as explained by the Dunning-Kruger effect (Kruger & Dunning, 1999). However, students were requested to assess the abilities required of them and not their own abilities. Students in the lower levels tend to perceive lower levels required compared to students with higher proficiency. This suggests that the lower the students' language proficiency, the lower the perceived abilities required of them and the higher the students' language proficiency, the higher the perceived required level of proficiency. A third interpretation is that, because of the students' low proficiency level, they viewed reaching certain levels as satisfactory, even if they were lower than those required; consequently, they chose a lower level as minimally required. This suggests that those responsible for conducting NA studies need to exercise caution and take the students' proficiency into consideration, because even when expressing their needs or in their perceptions of requirements, the students' proficiency might affect the way the students perceive the skills required. Students with low proficiency might not give accurate representations of their needs, compared to higher level students who gave closer results to their teaching staff perceptions of required levels.

5.4.2. RQ4: Writing skills required in the MHCCs

It is important to emphasise that the CEFR levels identified in Phase II as required only refer to the students' general English language proficiency and this needs to be differentiated from the specific skills (Athanasidou *et al.*, 2016), i.e. English for specific academic purposes (ESAP), required as part of the students' discipline; that is the particular academic literacies needed to navigate the content of their discipline (Lea & Street 1998; Murray, 2016a). Therefore, it is not enough to depend solely on the CEFR descriptors to get an in-depth overview of the students' writing requirements. According to Sahinkarakas and Arifi (2007), "there are a good number of descriptors [in the CEFR] to meet most but not all the ESP learners' needs" (p.91). This being the case, an exploratory in-depth analysis of the qualitative data collected in Phase II was deemed necessary to gain a better understanding of the writing skills required of first-year MHCC students.

Previous studies investigating language needs in similar contexts (Alblowi, 2016; Ghobain, 2014; Shukri, 2014) have found that staff members usually have high expectations of the students' English proficiency when they join the colleges, because of their attendance at the PYP. The same finding arose in the current study, where most staff expected students to join the college with high proficiency after an intensive year of English at the PYP. As suggested by the interview data, staff were disappointed with students' generally lower level compared to their expectations, particularly in writing.

Broadly speaking, both staff and students viewed writing as a very important skill required of students who joined MHCCs. However, the perceived writing requirements, especially for first-year students, differed between staff and students. According to the students, writing in the first year was only important for answering exam questions and occasionally submitting short essays or assignments. Beyond those instances, they did not view writing as an important skill for first-year students. But, at the same time, they acknowledged the importance of writing in their future years in medical college, where they would be required to write academically about medical topics. They were also aware of the importance of writing for their future profession. The students thus showed concern over their lack of academic and medical writing skills for their future needs and recommended that they receive continuous language support while they were studying in college.

Unlike the students' perception of the writing skills required for the first year, staff mentioned that writing requirements should extend far beyond answering short exam questions. According to most staff, students should be required and able to submit reports, assignments and mini-research papers in their first-year. However, most of them avoided requesting these tasks from the students, for reasons that will be outlined in the next chapter. Staff had similar concerns as students regarding the students' future academic, medical, and professional writing skills, noting that they would not have additional language support after the PYP to help with their academic and medical writing, not least because it was found in this study that academic staff do not consider checking and modifying students' language to be part of their responsibility.

The discrepancies between students' and staff's perceptions of the required skills of first-year students are understandable. Liu *et al.* (2011) highlighted "the importance of understanding needs as a complex, multiple, and conflicting concept" (p.277). When students view and assess their needs, they usually look at these needs in relation to "necessities and goals" (Liu *et al.*, 2011, p.271). In other words, what was actually required and asked of the students in their first year of college was what was perceived as needed/required by the students. Therefore, it is important to make explicit these required skills early on, both in the curriculum and assessment.

In relation to the writing skills currently required of first-year students in MHCCs, students must possess certain basic writing skills to answer exam questions, and be able to write clearly using simple, well-structured sentences. Their writing should clearly convey their understanding of the information needed to answer the given question. In their answers, there is no need for complex grammar, and they need to be able to use basic academic vocabulary, general medical terminology, and specific terms related to their discipline. Different skills and rhetorical modes are required, ranging from simple to more complex, when answering the exam questions. The students need to recite information by listing facts (which requires rote memorisation) and also need other skills to answer exam questions. Looking at the rhetorical modes identified by the students' responses to the brainstorming activity conducted during the FGs, students from all participating colleges must be able to "analyse patients' diseases", write an "analysis of a situation", write a "description", "give reasoning", "summarise", and "paraphrase" as part of their exam answers or written assignments (which are occasionally requested). The students needed to analyse and synthesise information from different sources and put them together using their own words. They also needed to state their own opinions in writing. These more complex skills are the areas in which many students experience difficulty, as stated by the participants in this study and also reported more widely in the literature (Swales & Feak, 2004).

Broadly speaking, these required skills reflect the range of levels B1-B2 of the CEFR writing scales. To answer the exam questions that require rote memorisation of information, the students need to meet B1, where they can “write straightforward connected texts on a range of familiar subjects within [their] field of interest” (CoE, 2001, p.61). At the same time, they are required higher levels (reflecting B2) to answer some of these questions, such as those that require “synthesising and evaluating information and arguments from a number of sources” (CoE, 2001, p.61).

Having said that, there was some leniency observed on the part of staff, mainly toward the linguistic aspects of students’ answers. Because of the continuous errors and staff feeling that corrections are not their responsibility, most tended to overlook language-related mistakes. Some accepted incorrect spelling “as long as the word is recognised” [SF10_PH2_CN_155]. Others accepted incorrect sentence structure provided they managed to understand what the students were trying to say. In essence, the staff were allowing B1-level capability, in terms of grammar, i.e., “[e]rrors occur, but it is clear what he/she is trying to express” (CoE, 2001, p.114), but at the same time, the staff also ideally required a minimum level of B2, where students “[s]how [...] a relatively high degree of grammatical control. Does not make mistakes which lead to misunderstanding” (CoE, 2001, p.114)..

In their general use of vocabulary (as identified from both MHCC students and staff data), students needed to exhibit “a good range of vocabulary for matters connected to [their] field and most general topics” (CoE, 2001, p.113), which again reflects B2 (scale *general use of vocabulary*). In addition, the students needed to have basic academic vocabulary and medical terminology, requiring a higher level of proficiency than B2 but not stated explicitly in the CEFR vocabulary scales.

The three perceived views of writing required at the MHCCs (*highly, partially* and *marginally* required) were found to reflect Murray’s (2013) tripartite articulation of proficiency: *general proficiency, academic literacy* and *professional communication*. According to Murray (2013), *general proficiency*, which “comprises a set of generic skills and abilities...reflected in learning that includes a focus on grammar...vocabulary development, reading and writing skills...”

(p.303), is what is *highly* required by all MHCC students. As recognized by Elder (1993), Johnson (1988) and Murray (2013), these skills are of high importance as they can affect students' academic success as students need them to answer exam questions and complete other tasks and assignments. In the current study, many students were found to lack the *general proficiency* required and, hence, writing is a very difficult task for them.

The need to write academic medical assignments, reports, and/or research was perceived as *partially* required by the students. One possible cause for this is that these types of tasks require proficiency beyond the students' capability and many students, therefore, have difficulty with them. Students had insufficient opportunities to practice more advanced skills of writing such as academic writing skills in medical-related topics while at the PYP. In addition, in these MHCCs, academic and medical writing skills were not systematically required, meaning that some staff members considered these tasks as part of their course and others did not. Also, these tasks did not play a significant role in the students' assessments, in which case the linguistic elements were excluded, and so these skills were perceived by the students as less important compared to *general proficiency* skills. For the students, these skills were only perceived as important for future use, while in contrast, the staff emphasized their importance to first-year medical students and advised that they should be required.

There is a fine line between the academic and medical skills required while the students are still studying in college and those required later in their professional lives. These are also different from the basic general writing *proficiency* that has to be developed prior to joining these colleges. According to Murray (2013), those two classifications (for college and for professional use) are differentiated in the concepts of *academic literacy* and *professional communication skills*, although he notes that there is clearly a degree of overlap. In the former, the students need *academic literacy* (which was perceived as *partially* required in this study) that is specific to their discipline and only relevant to a particular college/context (Sebolai, 2016). The latter (*marginally* required) is the skill set that students need to develop later in college as they prepare for their profession. These two concepts were also reflected in Maher (1986), who distinguished

between English for medical education purposes (EM-EP), where English needs to be taught as “part of primary medical studies” (p.115), and English for medical occupational purposes (EM-OP), which is relevant to the specialised skills needed for their profession. There are some shared characteristics between the two, but the level and the way they should be taught differ for students and for professionals (Maher, 1986).

Academic literacy “requires expansion of vocabulary, grammatical and discourse knowledge beyond what is required for social communication” (Cummins, 2009, p.22). In academic literacy, as stated by Murray (2013), students need “the specialised vocabularies, concepts and knowledges associated with particular disciplines, each of which has its own distinctive patterns of meaning-making activity (genres, rhetorical structures, argument formulations, narrative devices, etc.)” (p.303). For example, students in the CAMS were asked to write lab reports and design posters on topics related to their discipline. These requirements differed from requirements of other colleges, such as the CN or the CPH. This difference reflects the fact that each discipline has its own social meanings and identities (Lea & Street, 1998; Lea & Stierer, 2000). This was also clear in this study based on the analysis of the students’ mind-maps during the FGs, for which I asked them to list all the topics and the tasks they had been involved in during their first year in college. Although there were some common skills across all colleges, each seemed also to have specific literacies related to their discipline.

Because of the particular nature of the academic literacies required in the colleges, and also considering that the PYP has to prepare all students one year prior to joining the MHCCs, I would argue that improving students’ academic literacies should not be the PYP’s responsibility. The staff, and even the students, should not expect the PYP to be the place that tutors students in the academic literacies and discipline-specific terminology related to their colleges. Students should join these colleges with general writing skills, and possibly with general basic academic literacies (such as the basic medical terminology that is common across all the MHCCs), though this requirement is obviously currently lacking in the PYP. But to expect the students to join having the academic literacies relevant to their colleges (e.g. writing a clinical report) seems impossible as the PYP is only

one intensive year aiming to prepare all the students in the medical track (MT) who are expected to join different colleges. Staff and students need to be aware that, with each type of writing (which differs between colleges), there are some linguistic elements that must be taught alongside the main content. This is important because according to Gunn, Hearne, and Sibthorpe (2011), most university staff expect that students who have secured a place at a university should come with the academic literacies required of their discipline. But what was found in the current study is that the students have not been exposed to such literacies prior to joining these colleges and they need to develop them while studying in their colleges. This is similar to El Tantawi *et al.* (2016) conclusion. In their evaluation of first-year Saudi dental students' scientific writing skills, they found that the preparatory year might not be enough "to develop adequate writing skills among undergraduate dental students" (p.148). Therefore, conversancy in these literacies needs to be developed either by content teachers (college teaching staff) or by content teachers in collaboration with language teachers (Crocker, 1981; Jacobs, 2007) while studying at their colleges.

The *marginally* required skills are only marginal for first-year students but not for students in subsequent years, and particularly the final year in college. These skills refer to the *professional communication* skills that relate to the students' future professions, such as writing patient histories and handover notes, relevant documentation, writing published research, etc. Participants (both staff and students) were aware that these skills were not required for first-year students; however, because of the students' generally low level of proficiency and because there would be no language support in the future, this was a worry for them. The MHCC staff expressed their concern at the very low level of students' writing proficiency, and yet additional language support is unavailable during students' academic years at the university. In addition, some staff do not consider themselves responsible for dealing with language-related matters in students' writing. Furthermore, preparing students in these skills may not be the PYP's responsibility because it is only one intensive year prior to the start of the different colleges. Students need continuous language support while they are in college to improve these skills related to their profession.

The question remains, then, as to what the PYP should teach in order to prepare students for the MHCCs? If I consider Maher's (1986) definition of what English for medical purposes (EMP) means and implies, EMP in the PYP curriculum should be designed to cover only the basic general *proficiency* and general *academic literacies* (which are common across the MHCCs). According to Maher (1986), EMP is designed to “meet the specific English language needs of the medical learners (e.g. nurse, GP, dentist, etc.)” (p.112). It should focus only on topics, themes, and a “restricted range of skills which medical students in these domains might require” in the first year of college. As found in the analysis of the current study, there are some general and specific skills that are shared among the different colleges. For example, all the students must be able to write a summary and know certain basic medical terminology (related to academic literacy) that is used across the different colleges. They must possess the kinds of basic writing skills (basic general proficiency) that many lack. It is important, therefore, for the PYP to balance their curriculum and assessment to cater to the different writing skills required. The students need to be well prepared and to practice writing more often to develop strong basic skills proficiency, in order to more easily learn the different discipline-specific *academic literacies* and the *professional communication* skills later on.

In the future, more studies need to be conducted to help identify the general proficiency skills and the general academic literacies (which are shared across MHCCs) that should be introduced and developed in the PYP curriculum and the discipline/college-specific academic literacies and professional communication skills that should be learned and acquired alongside the content in the MHCCs. Those skills that are specific and unique to each discipline (i.e. college-specific *academic literacies*) should not be the responsibility of the PYP. Essentially, the PYP, as a prepratory one-year programme, needs to have a more general focus. Policymakers, therefore, need to reconsider integrating language support in each college curriculum to improve the students' conversancy in the academic literacies required for their field of study, and this needs to be a collaborative effort between the content lecturers and the English language teaching staff (Jacobs, 2007; Murray and Nallaya, 2016; Shukri, 2008). The students need basic writing skills, but they also still need extra language support while they are studying in

college. This was explicitly requested by both the students and staff members. The reasons for offering this recommendation are two-fold. First, and as indicated by the staff, looking at the students' language and modifying the students' writing is time-consuming, and they do not feel it is their responsibility. Second, the staff themselves might not be qualified enough to help the students improve their academic literacies. The same argument applies to teaching the marginally-required skills related to their profession (*professional communication skills*). Additional language support should be considered to help students acquire these important skills for the future.

To summarise, the students in the first year of MHCCs are required to join the colleges with general English language skills, including writing *proficiency*. These skills were perceived as *highly required* by both the students and their teaching staff. Their importance is not only based on the students' need to pass the exams but, as mentioned by Murray (2013) and documented in different studies (Elder, 1993; Johnson, 1988; Tonkyn, 1995), these basic skills are "prerequisites to developing academic literacy and professional communication skills" (p. 303), which will be the skills required in the future but also partially required in their first year of college.

Both students and staff have shown frustration over students' lack of preparedness to meet the colleges' requirements at the PYP. As is obvious from the previous phase, the curriculum did not prepare students sufficiently to meet their writing needs. This seems to be a common problem in different preparatory programmes in Saudi Arabia (Alghamdi, 2015; McMullen, 2014; Alkubaidi, 2017). It was also obvious that the students are only focusing on the skills that allow them to pass exams. According to Alkubaidi (2014), the materials provided in their foundation programme are "not designed for their specialization in university. It is general everyday English, and therefore, being in a country where English is a foreign language, it appears only logical that their goals are solely placed on passing the exam as there is no immediate use for learning to write" (p.211). My recommendation is similar to that of Murray and Nallaya (2014), who argue that "it is crucial that institutions equip their students with the academic literacies relevant to their disciplines if they are to ensure that they

both thrive academically during their studies and exit their programmes suitably equipped as graduates ready for the world of work” (p.2). As first suggested by Lea and Street (1998), there should be a distinction between academic literacies and study skills required of students in each field. From my point of view, the students need the *general proficiency* skills prior to joining the medical colleges. On the other hand, for the *academic literacies* required for each discipline, each college should be responsible for providing and maintaining their own language support and ensuring the students have access to the literacies required of them. Therefore, and as advised by Murray and Nallaya (2014), it is not only the staff and policymakers who need to be aware of this issue, it is also the students who “need to develop a working understanding of ... the genres to which they should strive to conform in their written work effectively...to allow them to stake a claim to membership in those communities [of practice]” (p.3).

On the one hand, although students acknowledge the importance of writing in general and for their academic and professional lives, unlike the academic staff, they do not consider writing demanding or very necessary in the first year of medical college. This is because the importance of any language skill is linked to the extent to which it is required in exams or as part of their assessment. This means that if curriculum designers, policymakers and academic staff want writing to be taken seriously, it needs to be considered as an integral part of students’ assessment and to be given enough weight in their exams and assignments. Simply put, to be important, it should be a required element of assessment.

Chapter 6

'What students can do' versus 'what students need to do':

Mind the Gap!

Phase III

6.1. Introduction

This chapter (Phase III) uses a gap analysis approach (Brown, 2016) to identify the misalignments between the writing achieved at the end of the PYP (*present-situation analysis*) in Phase I and the writing required (*target-situation analysis*) in Phase II. It aims to find out whether the PYP has adequately prepared students to meet the writing requirements in the university MHCCs and to identify any potential gaps which could help the PYP to better prepare students to meet the MHCC writing requirements. The first part of this chapter presents the misalignments between the CEFR levels acquired (in Phase I) and those required (in Phase II) to answer the study RQ5 by comparing the quantitative findings. To answer RQ6, the second part compares the qualitative results. The last part, which answers RQ7, discusses possible causes for misalignment along with implications for development. Answering Phase III research questions will lead to answering the study's main overarching question, namely whether the PYP has adequately prepared students to meet the writing requirements of the first-year in MHCCs.

6.2. Misalignments between the CEFR levels achieved (Phase I) and required (Phase II)

This section aims to answer RQ5 and, thus, the CEFR levels identified (in Phase II) as required and those perceived as achieved at the end of the PYP (in Phase I) were compared to identify any misalignments between them. CEFR levels were expressed as a numerical score: 1 (A1); 2 (A2); 3 (A2+); 4 (B1); 5 (B1+); 6 (B2);

7 (B2+); 8 (C1); 9 (C2). First, using paired t-tests, misalignments were identified across all students who participated in Phase I and Phase II by comparing their scores in each phase. Misalignments were calculated as Phase II minus Phase I score. Negative results indicate that the students achieved a CEFR level in Phase I higher than that required in Phase II. The differences between Phase I (achieved) and Phase II (required) were then compared across colleges and PYP levels using one-way ANOVA to establish whether misalignments differed by college or PYP levels, with appropriate *post-hoc* tests.

The analysis was then repeated using data provided by PYP students, PYP tutors and raters in Phase I, and the staff and students' choices of required levels in Phase II as part of the triangulation method which can be used as an additional determination of where the gap(s) might be located. Figure 6.1 visualizes the quantitative analysis of misalignments carried out in this chapter.

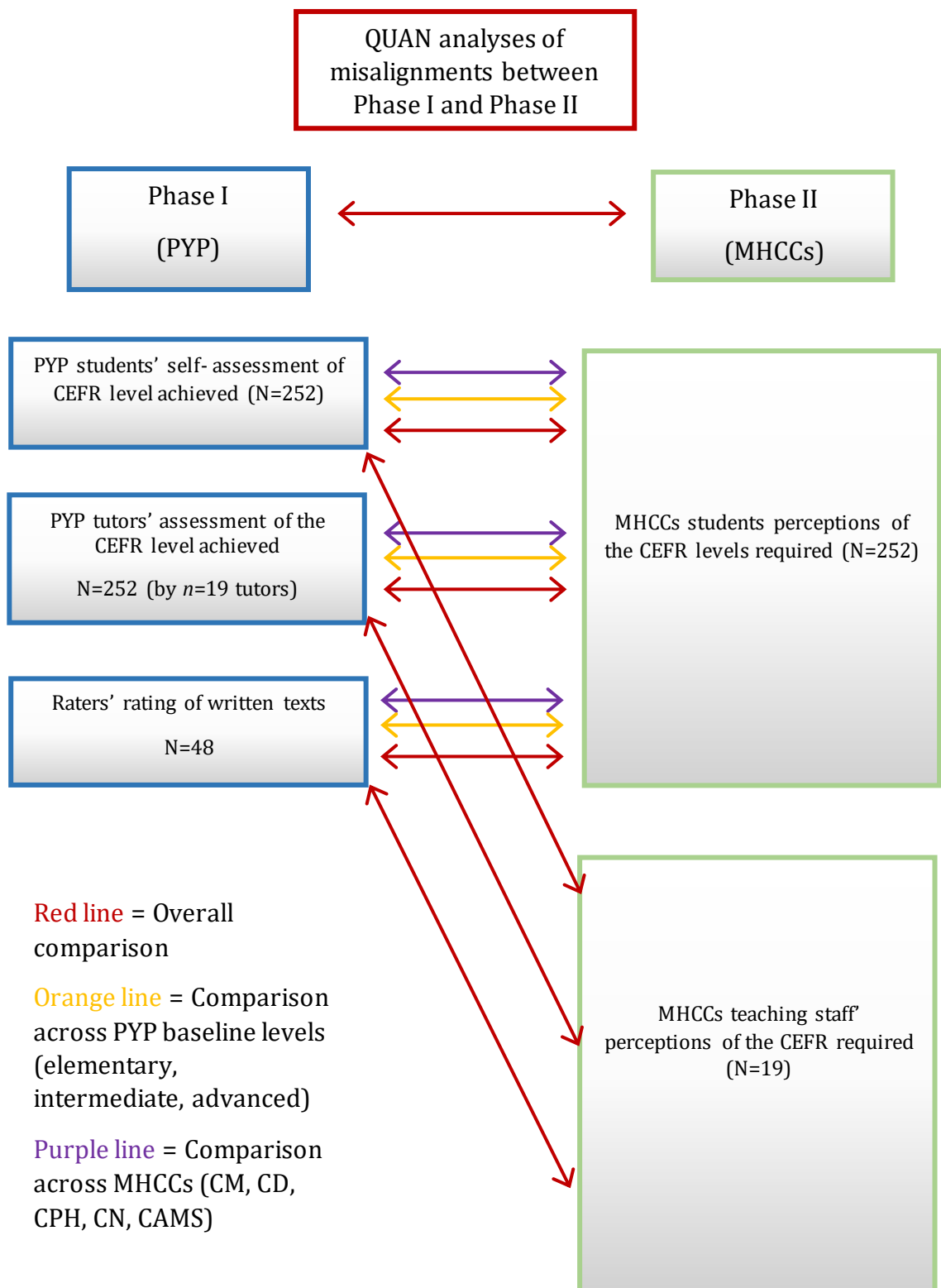


Figure 6. 1 Visualisation of the quantitative analyses of misalignments between CEFR levels achieved in PYP and those perceived as required in first year MHCCs.

6.2.1. Misalignments between CEFR levels achieved and required as perceived by the same students

In this section, I compared the students' self-assessment of their CEFR levels at the end of the PYP programme with the same students' perceptions of the required CEFR levels in college (only the students who participated in the two phases were considered for the analysis). The misalignments are investigated in general, across the MHCCs and based on the students' proficiency level at the PYP.

6.2.1.1. Overall misalignments of the students' CEFR levels

To explore possible misalignments between what CEFR levels students perceived as required (Phase II) and the levels they felt had been achieved (Phase I), paired sample t-tests were used, with the results summarised in Table 6.1. Positive differences (in difference/misalignment column) indicate that the required CEFR levels are higher than those achieved, while negative differences indicate that achieved levels are higher than those required.

Table 6.1 Overall misalignments in students' CEFR levels as perceived by student in Phases I and II

CEFR Scales	Phase I (Levels Achieved)		Phase II (Levels required)		Difference (Misalignment)		<i>t</i>	<i>df</i>	<i>P-value</i>	<i>Cohen's d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>				
Overall										
Written Production	6.85	2.23	6.13	1.94	-0.71	2.83	-4.01	251	<0.001	-0.25
Overall										
Written Interaction	5.16	2.68	5.08	2.06	-0.08	3.22	-0.37	251	0.71	-0.031
Type of Texts	5.08	2.56	5.30	2.25	0.21	3.19	1.07	251	0.29	0.063
What Can They Write	5.64	2.40	5.43	2.36	-0.21	3.07	-1.07	249	0.29	-0.065
Vocabulary Range & Control	4.73	2.39	5.16	1.93	0.43	2.71	2.50	249	0.013	0.15

Grammatical Accuracy	5.56	2.60	5.77	1.83	0.21	2.83	1.18	250	0.24	0.071
Orthographic Control	6.17	2.56	5.76	2.14	-0.41	2.94	-2.20	248	0.029	-0.14
Processing Texts	5.02	2.19	5.17	2.00	0.15	2.77	0.85	249	0.40	0.036
Reports and Essays	5.42	2.50	5.88	1.30	0.47	2.60	2.85	249	0.005	0.19
Note Taking	6.12	2.41	5.86	1.63	-0.26	2.65	-1.57	249	0.12	-0.11
Average Scales	5.67	1.76	5.55	1.28	-0.11	1.80	-0.99	251	0.32	-0.056

M=Mean, SD=Standard deviation, df=degrees of freedom

Coding scheme for CEFR Scales: 1 (A1); 2 (A2); 3 (A2+); 4 (B1); 5 (B1+); 6 (B2); 7 (B2+); 8 (C1); 9 (C2)

Misalignment score calculated as phase II level MINUS phase I level.

Cohen's d_z calculated as Mean misalignment/SD of misalignment. Cohen's d calculated as $2 \times t/\sqrt{t}$, 0.2=small effect; 0.5=medium; 0.8=large

Students report significantly higher levels achieved in Phase I on the *Overall written production* and *Orthographic control* scales than required in Phase II.

Conversely, students report significantly lower levels achieved in Phase I on the *Vocabulary range and control* and *Reports and essays* scales. The effect sizes (Cohen's d) were small for all four scales that were significantly different. Other scales were not significantly different.

6.2.1.2. Misalignments of the students' CEFR levels across MHCCs

The misalignments were then compared across colleges (Table 6.2). In some of the CEFR scales and colleges, students perceived higher CEFR levels were required than they had achieved (positive differences); e.g. in the *Reports and essays* scale for four colleges. However, the magnitude of differences varied between colleges.

Table 6.2 Misalignments in students' CEFR levels across MHCCs as perceived by the students

College	CM n=59		CPH n=55		CN n=20		CD n=43		CAMS n=75	
	M	SD	M	SD	M	SD	M	SD	M	SD
Overall Written Production	-1.02	2.92	-0.07	3.25	-1.45	3.00	-0.79	2.23	-0.71	2.66
Overall Written Interaction	-0.46	3.42	-0.20	2.80	0.30	2.99	0.28	2.83	0.01	3.63
Type of Texts	-0.08	3.30	0.02	3.31	0.85	3.82	0.88	2.52	0.04	3.16
What Can They Write	0.27	3.03	-0.54	2.63	0.30	3.51	-0.28	2.19	-0.45	3.68
Vocabulary Range & Control	0.61	2.79	0.28	2.72	0.40	2.41	0.21	2.12	0.53	3.06
Grammatical Accuracy	-0.02	2.74	0.73	2.92	0.05	1.88	-0.23	2.67	0.31	3.11
Orthographic Control	-0.36	2.77	-0.53	3.16	-0.65	2.76	-0.02	3.22	-0.53	2.83
Processing Texts	-0.25	2.91	0.29	2.58	0.20	2.09	0.49	3.15	0.15	2.75
Reports and Essays	-0.34	2.15	0.36	2.85	1.65	3.08	0.37	2.08	0.93	2.71
Note Taking	0.03	2.78	-0.15	2.60	0.55	2.42	-0.58	2.93	-0.63	2.46
Average Scales	-0.24	1.78	-0.08	1.79	0.08	1.78	-0.04	1.47	-0.13	2.03

M=Mean, SD=Standard deviation

Coding scheme for CEFR Scales: 1 (A1); 2 (A2); 3 (A2+); 4 (B1); 5 (B1+); 6 (B2); 7 (B2+); 8 (C1); 9 (C2)

Misalignment score calculated as phase II level MINUS phase I level.

To test for significant differences, ANOVA and alternative robust tests were used. Table 6.3 indicates the significant results; the complete findings are reported in Tables D1 and D2 in Appendix D.

Table 6.3 One-Way Analysis of Variance ANOVA of misalignments across MHCCs as perceived by students

CEFR Scales	SS	df	MS	F	P-value	η^2
Reports and Essays						
Between Groups	83.04	4	20.76	3.18	.014	.049
Within Group	1599.20	245	6.53			
Total	1682.24	249				

SS=Sum of squares, df=degrees of freedom, MS=mean square, F=F ratio, η^2 =Effect size: 0.02=small; 0.13=medium; 0.26=large.

There were no significant differences across the colleges in most of the CEFR scales, indicating similar misalignments across the five colleges. Scores differed significantly across colleges only for the *Reports and essays* scale ($P=.014$) with a medium size effect, which was then analysed using *post-hoc* tests (Table D3 in

Appendix D). For the CN and CAMS results, misalignments were significantly higher than in the CM for the *Reports and essays* scale. For this scale at the CM, the achieved level was slightly higher than the required level. No other misalignment significantly differed by college. These findings, interestingly, support the qualitative findings (in Chapter 5) where it was found that the students in the CAMS and the CN are being required to write essays and reports more than in the other colleges. At the same time, as found in Chapter 5, most of the lower level students have enrolled in these two colleges, which again explains this gap.

6.2.1.3. Misalignments of the students' CEFR levels across PYP levels

In this section, sizes of misalignments perceived by students between CEFR levels attained and required were explored across PYP levels (elementary, intermediate and advanced) (Table 6.4).

Table 6.4 Misalignments in students' CEFR levels across PYP levels as perceived by students

Level	Elementary n=21		Intermediate n=128		Advanced n=103	
	M	SD	M	SD	M	SD
CEFR Scales						
Overall Written Production	-0.43	2.64	-0.28	3.30	-1.31	2.02
Overall Written Interaction	-0.38	2.04	0.60	3.18	-0.85	3.29
Type of Texts	0.05	2.67	0.63	3.27	-0.26	3.14
What Can They Write	-0.24	2.64	0.24	3.39	-0.75	2.65
Vocabulary Range & Control	0.76	2.14	0.96	2.80	-0.30	2.56
Grammatical Accuracy	0.19	2.48	0.24	2.71	0.17	3.05
Orthographic Control	-0.35	3.17	-0.25	3.19	-0.62	2.56
Processing Texts	0.71	2.08	0.48	2.69	-0.37	2.92
Reports and Essays	0.95	2.25	1.26	2.73	-0.60	2.09
Note Taking	-0.48	2.80	0.14	2.70	-0.72	2.51
Average Scales	0.04	1.80	0.28	1.92	-0.63	1.52

M=Mean, SD=Standard deviation

Coding scheme for CEFR Scales: 1 (A1); 2 (A2); 3 (A2+); 4 (B1); 5 (B1+); 6 (B2); 7 (B2+); 8 (C1); 9 (C2)

Misalignment score calculated as phase II level MINUS phase I level.

For some scales (e.g. *Reports and essays*), the advanced students achieved levels higher than required (negative misalignment value), while elementary and intermediate students required higher levels than they had achieved (positive values).

To investigate whether the differences among the three PYP levels were significant, ANOVA and alternative robust tests were used (significant results in Tables 6.5 and 6.6 below; complete results in Tables D4 and D5 in Appendix D).

Table 6.5 One-Way Analysis of Variance ANOVA of misalignments in CEFR levels across PYP levels as perceived by students

CEFR Scales	SS	df	MS	F	P-value	η^2
Vocabulary Range and Control						
Between Groups	93.01	2	46.51	6.61	0.002	0.051
Within Group	1738.19	247	7.04			
Total	1831.20	249				
Processing Texts						
Between Groups	47.83	2	23.92	3.17	0.044	0.025
Within Group	1863.70	247	7.55			
Total	1911.52	249				
Note Taking						
Between Groups	43.07	2	21.54	3.11	0.046	0.025
Within Group	1709.50	247	6.92			
Total	1752.58	249				
Average CEFR Levels						
Between Groups	47.95	2	23.97	7.79	0.001	0.059
Within Group	766.20	249	3.08			
Total	814.14	251				

SS=Sum of squares, df=degrees of freedom, MS=mean square, F=F ratio, η^2 =Effect size: 0.02=small; 0.13=medium; 0.26=large.

Table 6.6 Robust test of equality of mean of misalignments in CEFR levels across PYP levels as perceived by students

CEFR Scales	<i>Statistic</i>	<i>df1</i>	<i>df2</i>	<i>P-value</i>
Overall Written Production				
Welch	4.61	2	55.59	0.014
Brown-Forsythe	4.40	2	86.62	0.015
Overall Written Interaction				
Welch	5.96	2	67.46	0.004
Brown-Forsythe	8.10	2	174.76	<0.001
Reports and Essays				
Welch	18.00	2	57.50	<.001
Brown-Forsythe	18.56	2	95.52	<.001

df=degrees of freedom

Where significant differences by PYP level were observed, *post-hoc* pairwise comparisons were conducted (Tables D6 and D7 in Appendix D). In the *Reports and essays* scale, a significant difference was found between the advanced and both the intermediate and elementary levels but not between the elementary and intermediate levels. In some other scales, significant differences were seen between the advanced and intermediate levels such as in *Vocabulary range and control* and the *Average scale*. The significant differences evident between the advanced and intermediate levels but not at the elementary level are expected as the elementary-level students over-assessed and intermediate-level students under-assessed their abilities in Phase I, which explains why the gap appears bigger for the intermediate level students than the elementary ones.

6.2.2. Misalignments in CEFR levels between PYP tutors and MHCC students

This part compares PYP tutors' perceptions of the CEFR levels achieved by students at the end of the PYP with students' perceptions of the CEFR levels required in the MHCCs. The same students who were assessed by PYP tutors in Phase I and then participated in Phase II were considered for the analysis (using the same analyses as above).

6.2.2.1. Overall misalignments in CEFR levels between PYP tutors' and MHCC students

PYP tutors' perceptions of the CEFR level achieved were compared with the levels required in the MHCCs as perceived by the students (Table 6.7).

Table 6.7 Overall misalignments in CEFR levels between PYP tutors and MHCC students

<i>CEFR Scales</i>	Phase I PYP tutors		Phase II MHCCs students		Differences (Misalignments)			<i>t</i>	<i>df</i>	<i>P-value</i>	<i>Cohen's d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>					
Overall Written Production	6.64	2.17	6.11	1.94	-0.52	2.69	-3.04	244	0.003	-0.19	
Overall Written Interaction	6.11	2.11	5.14	2.08	-0.96	2.77	-5.44	244	<0.001	-0.36	
Type of Texts	6.36	2.24	5.36	2.26	-1.01	2.78	-5.67	244	<0.001	-0.36	
What Can They Write	5.47	2.12	5.28	2.25	0.01	2.71	0.071	243	0.94	-0.000	
Vocabulary Range & Control	5.20	2.07	5.54	1.95	0.34	2.46	-2.17	242	0.031	0.12	
Grammatical Accuracy	5.44	2.08	5.83	1.81	0.39	2.55	2.38	243	0.018	0.15	
Orthographic Control	5.87	2.06	5.85	2.09	-0.03	2.768	-0.14	229	0.89	-0.000	
Processing Texts	4.91	2.16	5.23	1.98	0.31	2.56	1.85	230	0.065	0.12	
Reports and Essays	5.61	2.23	5.95	1.35	0.34	2.40	2.14	229	0.033	0.13	
Note Taking	5.37	2.34	5.89	1.67	0.53	2.76	2.88	228	0.004	0.18	
Average Scales	5.77	1.76	5.64	1.26	-0.13	1.82	-1.12	228	0.26	-0.056	

M=Mean, SD=Standard deviation, df=degrees of freedom

Coding scheme for CEFR Scales: 1 (A1); 2 (A2); 3 (A2+); 4 (B1); 5 (B1+); 6 (B2); 7 (B2+); 8 (C1); 9 (C2)

Misalignment score calculated as phase II level MINUS phase I level.

Cohen's d calculated as Mean misalignment/SD of misalignment. Cohen's-d calculated as $2 \times t / \sqrt{df}$, 0.2=small effect; 0.5=medium; 0.8=large

There are some small but significant positive misalignments between the CEFR levels required and achieved (i.e. the required level is higher than attained), for the *Vocabulary range and control*, *Grammatical accuracy*, *Reports and essays* and *Note-taking* scales. This indicates that there are some gaps, though small, in the students' proficiency required for medical colleges writing, and as such, students need to be more prepared in those skills prior to joining the colleges.

The negative differences, on the other hand, indicate that the level attained is higher than that required. For example, the students seem to have achieved a slightly higher level in the *overall written productions* and *interactions* scales than the levels required. This may be explained by the limited writing required of students in the medical colleges (see Chapter 5).

6.2.2.2. Misalignments in CEFR between PYP tutors and MHCC students across MHCCs

These results were compared across colleges (Table 6.8); significance was tested using ANOVAs (Table D8 in Appendix D). There were no significant differences observed across the colleges for any of the scales, indicating that the misalignment observed by the comparison between the PYP tutors (levels attained) and MHCCs students (level required) is similar across colleges.

Table 6.8 Misalignments in CEFR levels between PYP tutors and MHCC students across MHCCs

College	CM		CPH		CN		CD		CAMS	
	M	SD	M	SD	M	SD	M	SD	M	SD
Overall Written Production	-0.96	2.83	-0.39	2.59	-0.30	3.25	-0.51	2.31	-0.35	2.75
Overall Written Interaction	-1.31	2.31	-1.39	2.84	-0.10	3.65	-0.93	2.81	-0.63	2.73
Type of Texts	-1.27	2.74	-1.28	2.89	-0.10	3.23	-0.71	2.50	-1.04	2.78
What Can They Write	-0.16	2.66	-0.43	2.66	0.50	2.76	-0.13	2.68	0.44	2.79

Vocabulary Range & Control	-0.44	2.31	-0.32	2.69	-0.42	2.63	-0.58	2.46	-0.11	2.38
Grammatical Accuracy	0.22	2.29	0.46	2.81	0.58	2.87	0.20	2.56	0.54	2.51
Orthographic Control	-0.27	2.84	-0.53	2.65	0.31	3.53	0.21	2.93	0.33	2.45
Processing Texts	-0.62	2.67	0.45	2.61	0.69	2.27	0.77	2.64	0.54	2.30
Reports and Essays	0.25	2.26	0.19	2.87	0.13	2.70	0.33	2.30	0.58	2.15
Note Taking	0.69	3.01	0.46	2.70	0.63	3.16	0.23	2.82	0.61	2.52
Average Scales	-0.41	1.77	-0.20	2.05	0.16	2.25	-0.13	1.70	0.07	1.64

M=Mean,SD=Standard deviation

Coding scheme for CEFR Scales: 1 (A1); 2 (A2); 3 (A2+); 4 (B1); 5 (B1+); 6 (B2); 7 (B2+); 8 (C1); 9 (C2)

Misalignment score calculated as phase II level MINUS phase I level.

6.2.2.3. Misalignment in CEFR levels between PYP tutors and MHCCs students across PYP levels

Misalignments in tutors' data and MHCCs students' data were then compared across the three PYP levels (Table 6.9).

Table 6.9 Misalignments in CEFR levels between PYP tutors and MHCC students across PYP levels

CEFR Scales	Elementary, n=25		Intermediate, n=115		Advanced, n=105	
	M	SD	M	SD	M	SD
Overall Written Production	1.08	2.75	-0.59	2.90	-0.83	2.30
Overall Written Interaction	-0.04	2.80	-0.80	3.00	-1.36	2.48
Type of Texts	-0.24	3.02	-0.89	2.91	-1.32	2.56
What Students can Write	1.04	2.26	0.19	2.90	-0.43	2.52
Vocabulary Range & Control	0.67	2.63	-0.10	2.42	-0.95	2.36
Grammatical Accuracy	0.96	2.69	0.31	2.33	0.34	2.76
Orthographic Control	1.00	2.94	0.24	2.82	-0.47	2.61
Processing Texts	1.94	2.26	0.72	2.20	-0.39	2.74
Reports and Essays	1.28	2.93	0.42	2.34	0.10	2.34
Note Taking	1.61	3.01	0.48	2.74	0.38	2.72
Average Scales	0.95	2.14	0.02	1.90	-0.47	1.82

M=Mean,SD=Standard deviation

Coding scheme for CEFR Scales: 1 (A1); 2 (A2); 3 (A2+); 4 (B1); 5 (B1+); 6 (B2); 7 (B2+); 8 (C1); 9 (C2)

Misalignment score calculated as phase II level MINUS phase I level.

For most scales, the level required as perceived by elementary students is higher than their achieved levels, according to tutors' perceptions (positive misalignment value), while the advanced students have achieved or surpassed the required level (negative misalignment values) on most scales.

To identify whether the differences in misalignments are significant across PYP levels, ANOVA analyses and robust tests were conducted (significant results reported in Table 6.10 and 6.11, respectively; complete results in Tables D9 and D10 in Appendix D).

Table 6. 10 One-Way Analysis of Variance ANOVA of misalignments between PYP tutors and MHCC students across PYP levels

CEFR Scales	SS	df	MS	F	P-value	η^2
Overall written Production						
Between Groups	74.58	2	37.29	5.33	0.005	0.042
Within Group	1694.55	242	7.00			
Total	1769.13	244				
What Students Can Write						
Between Groups	50.68	2	25.34	3.53	0.031	0.028
Within Group	1732.28	241	7.19			
Total	1782.96	243				
Vocabulary Range and Control						
Between Groups	57.83	2	28.92	4.95	0.008	0.040
Within Group	1400.82	240	5.84			
Total	1458.65	242				
Orthographic Control						
Between Groups	45.97	2	22.99	3.07	0.048	0.026
Within Group	1697.87	227	7.48			
Total	1743.84	229				
Average CEFR Levels						
Between Groups	35.77	2	17.89	5.62	0.004	0.047
Within Group	719.80	226	3.19			
Total	755.58	228				

SS=Sum of squares, df=degrees of freedom, MS=mean square, F=F ratio, η^2 =Effect size: 0.02=small; 0.13=medium; 0.26=large.

Table 6. 11 Robust test of equality of mean of misalignments between PYP tutors and MHCC students across PYP levels

CEFR Scales	<i>Statistic</i>	<i>df1</i>	<i>df2</i>	<i>P-value</i>
Processing Texts				
Welch	9.58	2	48.71	0.001
Brown-Forsythe	10.37	2	84.57	0.001

df=degrees of freedom

Not all the CEFR scales were significantly different across PYP levels. For example, the *Reports and essays* scale showed no significant differences, indicating similar gaps across all PYP levels. This suggests that, regardless of their proficiency level, most students in the MHCCs need to improve their reports and essay writing skills to meet the colleges' requirements.

For the scales that showed significant differences, a *post-hoc* test was conducted (see Tables D11 and D12 in Appendix D). For *Overall Written Production*, elementary-level students have a large deficiency in skill (positive misalignment value), which is significantly different from the intermediate and advanced who have exceeded the required levels (negative misalignment values). For the *Vocabulary range and control*, *What students can write* and *Processing texts* scales, elementary-level students again have large gaps, which are significantly different from the advanced but not the intermediate students.

In summary, when comparing PYP tutors' perceptions of the CEFR levels attained (in Phase I) with MHCCs students' perceptions of the CEFR levels required (in Phase II), the students generally show similar proficiency levels to those required, except for particular CEFR scales such as the *Grammar accuracy*, *Vocabulary range and control*, *Note-taking* and *Reports and essays* scales. The differences between the colleges were not significant, indicating similar gaps across colleges. However, when the data were compared across PYP levels, the significant differences in misalignments were mostly found between the elementary and the advanced levels, but not between elementary and intermediate levels, showing a bigger gap in the elementary level.

6.2.3. Misalignments in CEFR levels between raters' ratings and MHCC students

In this part, I compared the mean CEFR levels assigned by the raters to the students' written texts in Phase I with the mean of the students' perceptions of the CEFR levels required in Phase II (using similar analyses to the above). Only students whose texts were rated by the raters in Phase I and who participated in Phase II were considered for the analysis ($n=48$).

6.2.3.1. Overall misalignments in CEFR levels between raters' ratings and MHCC students

Table 6.12 shows the results of the paired sample t-test between Phase I raters' mean ratings and Phase II students' mean scores.

Table 6.12 Overall misalignments in CEFR levels between raters' ratings and MHCC students

	Phase I (Levels Achieved) $n=48$		Phase II (Levels required) $n=48$		Differences (Misalignment)		t	df	P -value	Cohen's d
	M	SD	M	SD	M	SD				
Average Scales	4.44	0.99	5.36	1.12	0.92	1.35	4.73	47	<0.001	0.84

M=Mean, SD=Standard deviation, df=degrees of freedom

Coding scheme for CEFR Scales: 1 (A1); 2 (A2); 3 (A2+); 4 (B1); 5 (B1+); 6 (B2); 7 (B2+); 8 (C1); 9 (C2)

Misalignment score calculated as phase II level MINUS phase I level.

Cohen's d_z calculated as Mean misalignment/SD of misalignment. Cohen's-d calculated as $2 \times t/\sqrt{df}$, 0.2=small effect; 0.5=medium; 0.8=large

A significant positive difference (significant misalignment) was identified, indicating that the CEFR level required (as perceived by the students in Phase II) was higher than that achieved by the students as rated by the raters in Phase I.

6.2.3.2. Misalignments in CEFR levels between raters and MHCC students across MHCCs

Using ANOVA, the results were then compared across colleges (Table 6.13).

Table 6.13 Misalignments in CEFR levels between raters' rating and MHCC students across the MHCCs

	CM		CPH		CN		CD		CAMS					
<i>Ratings</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>F</i>	<i>df</i>	<i>P-value</i>	η^2
Phase I	5.13	1.21	4.52	0.76	4.33	1.27	5.15	0.65	4.06	0.77	3.1	4,40	0.024	0.24
Phase II	6.25	0.82	5.16	1.11	5.95	0.21	5.79	1.25	4.91	1.19	2.4	4,40	0.061	0.20
Differences (Misalignm ent)	1.12	1.55	0.65	1.40	1.62	1.06	0.64	1.25	0.85	1.31	0.3	4,40	0.85	0.03

M=Mean, SD=Standard deviation, df=degrees of freedom, η^2 =Effect size: 0.02=small; 0.13=medium; 0.26=large.

There was no significant difference in the misalignments across colleges; for each college, the students' perceptions of CEFR required levels in Phase II were higher than their levels achieved as identified by the raters.

6.2.3.3. Misalignments in CEFR levels between raters' ratings and MHCCs students across PYP levels

A one-way ANOVA was then conducted across PYP levels to see if the misalignments differed by the students' proficiency level (Table 6.14).

Table 6.14 Misalignments in CEFR levels between raters' ratings and MHCC students across PYP levels

<i>Ratings</i>	Elementary		Intermediate		Advanced		<i>F</i>	<i>df</i>	<i>P-value</i>	η^2
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>				
Phase I	3.57	0.40	3.94	0.85	5.16	0.75	13.96	245	<.001	.38
Phase II	4.50	0.62	5.23	1.22	5.67	1.15	1.64	245	.25	.068
Differences	0.63	0.93	1.28	1.41	0.50	1.22	2.03	245	.14	.083

M=Mean, SD=Standard deviation, df=degrees of freedom, η^2 =Effect size: 0.02=small; 0.13=medium; 0.26=large.

Scores in both Phases increased with PYP level, and the students' assessment in Phase II were consistently higher than Phase I ratings at all PYP levels, indicating higher levels required than those achieved on the PYP. There was no significant difference in differences (misalignment) across PYP levels.

6.2.4. Misalignment in CEFR levels between PYP students' self-assessment and MHCC teaching staff

In this part, misalignment between the levels achieved as perceived by the students' self-assessment in Phase I were compared with the teaching staff's assessment of the levels required in Phase II. The students in Phase I and the teaching staff from Phase II are two independent samples. Additionally, the analysed data in each phase has unequal numbers ($n=252$ students in Phase I and $n=19$ teaching staff indicated their perceptions of the required levels in general in Phase II). To compare the mean differences in these two assessments for each

CEFR scale, an independent two-sample *t*-test was employed and Welch's two sample *t*-test with unequal variances was used.

Table 6. 15 Overall misalignments in CEFR levels between PYP students and MHCCs teaching staff

	Students (Phase I) N=252		Staff (Phase II) N=19		Difference (Misalignment)		<i>t</i>	<i>df</i>	<i>P</i>	Cohen's <i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>				
Overall Written Production	6.85	2.23	5.89	1.41	-0.95	0.35	2.70	25.38	0.012	0.33
Type of Texts	5.08	2.56	6.47	2.44	1.39	0.61	-2.29	269	0.023	-0.28

M= Mean, SD=Standard deviation

Coding scheme for CERF Scales: 1 (A1); 2 (A2); 3 (A2+); 4 (B1), 5 (B1+); 6 (B2); 7 (B2+); 8 (C1); 9 (C2)

Cohen's-d calculated as $2 \times t / \sqrt{df}$. Cohen's-d calculated as $2 \times t / \sqrt{df}$, 0.2=small effect; 0.5=medium; 0.8=large, **t*-test carried out without the assumption of equal variances

There was a significant difference in the assessment for *Overall written production*. The academic staff's assessment in Phase II was significantly lower than the students' assessments in Phase I. The opposite was found with the *Type of texts* scale where the staff perception of required level was significantly higher than the level perceived by students to have achieved.

6.2.5. Misalignment in CEFR levels between PYP tutors and MHCC teaching staff

In this part, misalignment between the level achieved and the level required was identified by comparing PYP tutors' perceptions of the students' proficiency achieved in Phase I with the teaching staff's perceptions of the required levels in Phase II.

The tutors in Phase I and the teaching staff from Phase II are two independent samples. Additionally, the analysed data in each phase has unequal numbers ($n=252$ students were assessed by $n=19$ PYP tutors in Phase I and $n=19$ teaching staff indicated their perceptions of the required levels in general in Phase II). To compare the mean differences in these two assessments for each CEFR scale, an independent two-sample *t*-test was employed, assuming equal variances of the

two groups using Levene's F test. Three of the CEFR scales (*Overall Written Production, Processing texts* and *Average*) violated the assumption of equal variances, so Welch's two sample t-test with unequal variances was used to compare misalignments between the two groups on those scales. The scales with significantly different values between Phases (using either the independent two-sample *t*-test or the Welch's test, as appropriate) are shown in Table 6.16; complete results for all scales are in Tables D13 and D14 in Appendix D.

Table 6. 16 Overall misalignments in CEFR levels between PYP tutors and MHCC teaching staff

CEFR Scales	Phase I (Levels Achieved)		Phase II (Levels required)		Differences (Misalignment)		t	df	P-value	Cohen's d
	M	SD	M	SD	M	SD				
Overall										
Written	6.11	2.12	4.95	1.68	-1.16	0.50	-2.24	262	0.021	-0.53
Interaction										
Overall										
Written	6.64	2.17	5.89	1.41	-0.74	0.35	-2.11	25	0.045	-0.35
Production										

M=Mean,SD=Standard deviation,df=degrees of freedom
Coding scheme for CEFR Scales: 1 (A1); 2 (A2); 3 (A2+); 4 (B1); 5 (B1+); 6 (B2); 7 (B2+); 8 (C1); 9 (C2)
Misalignment score calculated as phase II level MINUS phase I level.
Cohen's-d calculated as $2 \times t / \sqrt{df}$. Cohen's-d calculated as $2 \times t / \sqrt{df}$, 0.2=small effect; 0.5=medium; 0.8=large

Generally, there were no significant misalignments between the CEFR levels the students have achieved, based on PYP tutors' perceptions, compared to the ones required in the medical colleges as indicated by the teaching staff. This gives an impression that the CEFR levels attained are quite similar to the levels required. This might suggest that the students are joining these medical colleges with proficiency levels similar to the required levels. This is true to a certain extent; however, the qualitative analysis of the data from the focus groups and interviews (summarised in the second half of this chapter) offers deeper insights

into possible reasons behind the close similarity between the levels achieved and the levels required (the vicious circle).

The only significant difference between tutors' and staffs' assessment from Phase I and II is in the *Overall Written production* and *Interaction*, where staff's assessment in Phase II is significantly lower than in Phase I. However, the results need to be treated with caution given that the number of participants in each phase varied greatly. It is also difficult to investigate the analysis across colleges and PYP levels due to the limited number of participants (teaching staff) in phase II. The qualitative data below explore these results further.

6.2.6. Misalignment in CEFR levels between raters' ratings and MHCC teaching staff

An independent sample t-test analysis was conducted to compare the average CEFR levels assigned by the raters to the written texts ($n=105$) and the perceptions of the CEFR levels required by the staff, to identify any misalignment between the two phases (Table 6.17). Equality of variances was checked and met using the Levene's F test.

Table 6. 17 Overall misalignments in CEFR levels between raters's ratings and MHCC teaching staff

Phase I (level achieved) $n=105$		Phase II (level required) $n=19$		Difference					
<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>t</i>	<i>df</i>	<i>P-value</i>	<i>Cohen's d</i>
4.18	0.96	5.62	1.00	1.45	0.24	6.01	119	<.0001	1.47

M=Mean, SD=Standard deviation, df=degrees of freedom,
Coding scheme for CEFR Scales: 1 (A1); 2 (A2); 3 (A2+); 4 (B1); 5 (B1+); 6 (B2); 7 (B2+); 8 (C1); 9 (C2)
Misalignment score calculated as phase II level MINUS phase I level.
Cohen's-d calculated as $2 \times t/\sqrt{df}$. Cohen's-d calculated as $2 \times t/\sqrt{df}$, 0.2=small effect; 0.5=medium; 0.8=large

There was a significant difference indicating that a higher level is required in Phase II compared to that achieved in Phase I; i.e. students' written texts are not meeting the CEFR levels required in these medical colleges.

However, the sample size in Phase II is small, so that smaller differences between colleges may not be detected, and the low target level of the task that the students used to produce the texts might not have provided a true reflection of their proficiency.

6.3. Identifying the gaps between the writing skills perceived achieved (Phase I) and the ones required (Phase II)

This part aims to answer RQ6. To answer this question, first I analysed the students' responses to the table in the Phase II questionnaire (see Appendix A2, Part Two (1)). The students were asked to fill in the table with the writing skills they have been working on and dealing with during their first year in college (whether directly related to their academic subjects or not). Using a Likert scale that accompanied the table, they then evaluated the extent to which their PYP tuition in these skills prepared them for the college phase. In addition, I compared the qualitative findings from phases I and II to identify any potential gaps between the two phases.

6.3.1. Writing at the PYP curriculum: where is the gap?

Figure 6.2 summarises the writing skills as perceived by the students in the first-year of MHCCs when they were asked to fill in the table in Part Two of the questionnaire (Appendix A2), presenting the identified categories according to the number of participants who listed them as required for the first-year of college. The red bars represent the categories that are mostly identified by the students (each identified by >40 students). Then, figure 6.3 visualizes the mean scores of the students' evaluation of how well-prepared they felt (using the

Likert scale) by the PYP curriculum regarding those skills (0=*Not prepared at all* to 3=*Very well prepared*).

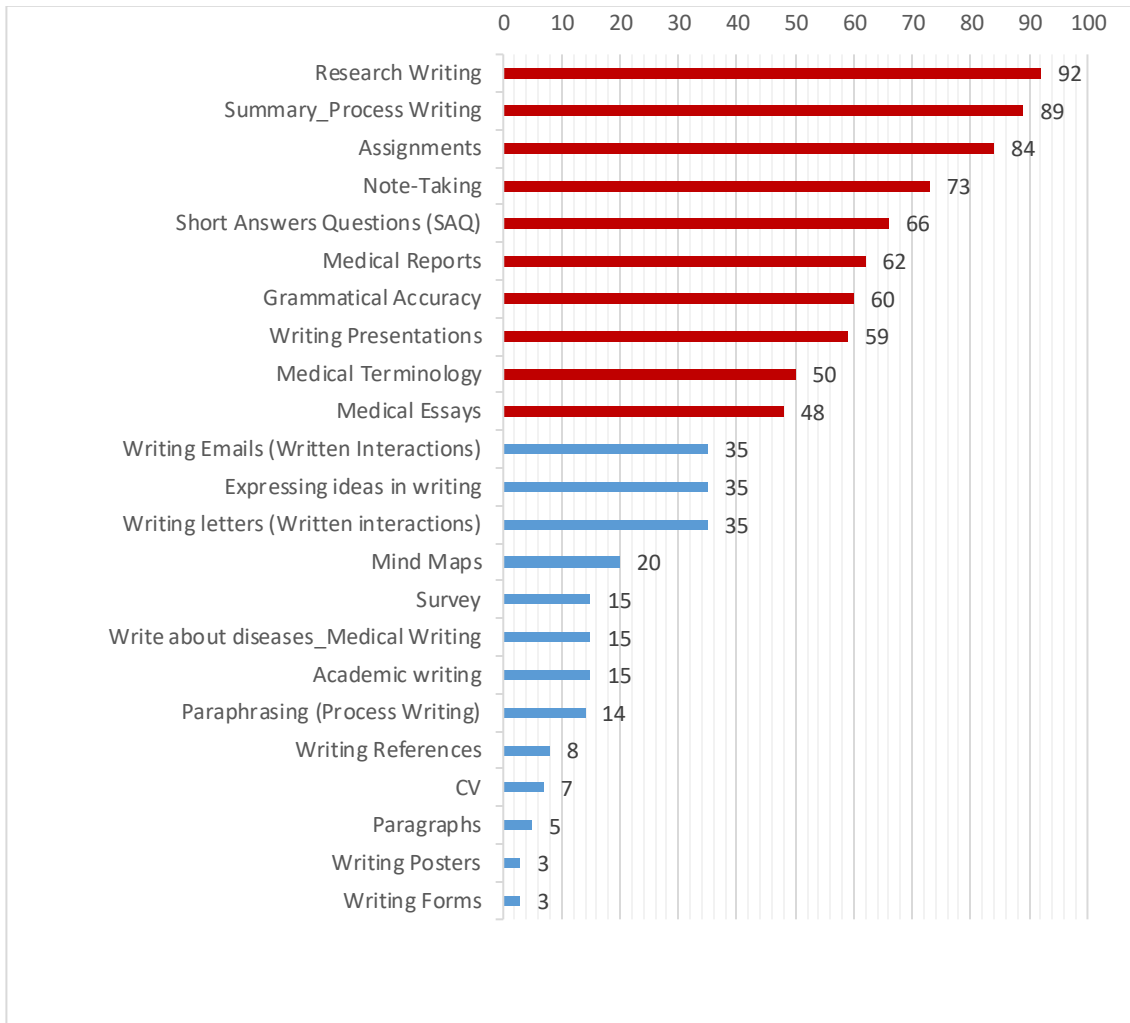


Figure 6. 2 Distribution of writing skills as identified by the number of students for each skill

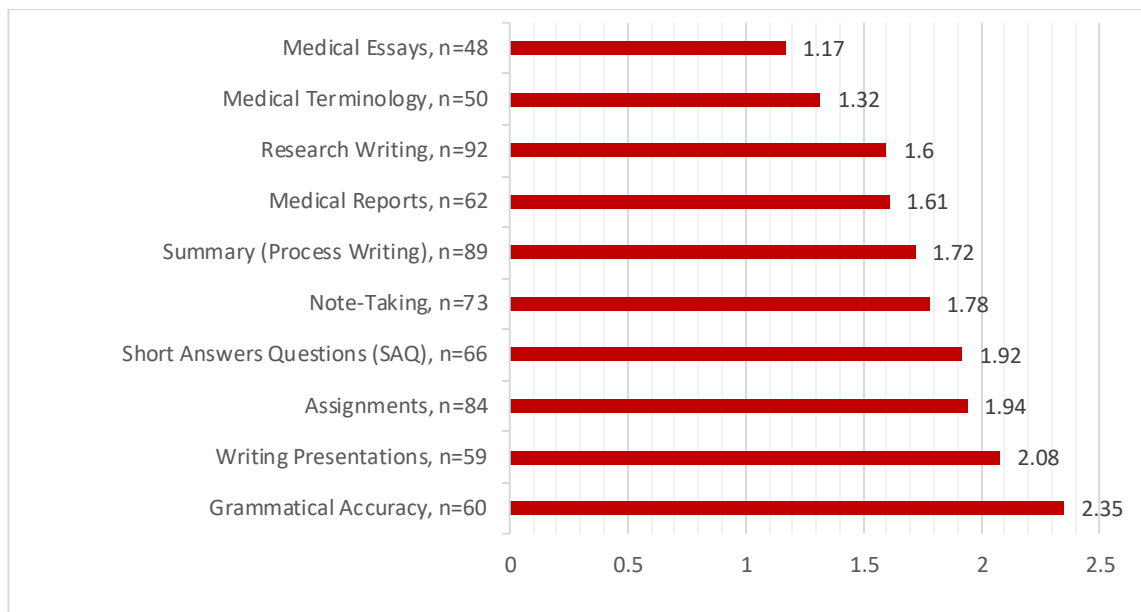


Figure 6.3 The ranking of how well students felt prepared by the PYP curriculum. 0=Not prepared at all to 3=Very well prepared.

As can be seen, the use of grammar (*grammatical accuracy*) received the highest mean rating. Indeed, most students felt very well-prepared to deal with the grammatical requirements of the first year of their programmes. *Writing presentations* was a close second. It seems that the students felt more prepared regarding general writing skills than medical writing (e.g. medical essays) and academic skills (e.g. research writing).

FG discussions supported the findings of the questionnaire. Students in the CAMS, for example, were satisfied with the PYP grammar curriculum because “*basic grammatical skills and not complicated structures*” [ST2_PH2_FG1_CAMS_208] were needed during their first year at the college. Staff members, however, disagreed that the PYP offered satisfactory instruction in grammar and syntax. One staff member at the CN pointed out that “*when you check the students’ reports, you cannot get away without a concern for their grammar*” [SF7_PH2_CN_175].

The FGs analysis also found that the PYP focuses on tutoring students in English for general purposes, especially such basic writing skills as paragraph structure,

the use of punctuation, some grammatical rules, and general vocabulary. There were few opportunities for student to practice these skills beyond such most basic general topics as writing about ‘*a member in your family*’, ‘*your mum*’ or ‘*someone you admire the most*’. Students in the medical colleges need more advanced writing skills to express their opinions and analyse medical topics.

Medical and academic writing are absent from both the PYP and MHCC curricula. Students rarely, if ever, submit written assignments of any kind, not just research papers. In both contexts, the students’ poor writing proficiency, their fear of losing marks, and instructors’ frustration with plagiarism were reasons for avoiding the practice of writing in graded assignments. Likewise, such forms of professional medical writing as clinical reports, documentation, and patient histories are also absent from the PYP curriculum. Although these medical writing skills are not a key component of success for first-year students in the medical colleges, they will be later on in their academic studies.

Students in Phase II argued that the writing skills in the PYP curriculum:

need to be related to their future fields of study. Students will not be interested unless it is related. I studied general English my whole life, so what does PYP add by teaching general English? [ST4_PH2_FG3_CM_488-489].

For most of the students in Phase II, writing as taught by the PYP curriculum was not appropriate to their needs as future medical professionals: *PYP improved [my] general English, but not medical English*” [ST7_PH2_FG2_CAMS_134-135].

[The PYP]...actually did not care about medical terminology. It was only one class a week and I think we have covered only 10 pages of the book [EMP textbook] and when we came here they told us you should have known these words already from the PYP [ST3_PH2_FG1_CAMS_159-161].

Students in CAMS were especially critical of the PYP’s general approach. When required to submit essays and reports related to their speciality, these students struggled because “*no one has taught us how to write a report...and* [university

staff] *d[o] not expect [us] not to know*" [ST6_PH2_FG2_CAMS_312]. Another student described how "*at the PYP, we were only requested to write general essays. It is different here, where we need to write in detail about specific topics*" [ST8_PH2_FG2_CAMS_185-186].

Staff at the medical colleges believed that additional support to improve students' writing was necessary not only because the PYP curriculum falls short, but also because students must learn how to continuously improve their academic literacies and professional communication skills to succeed:

I think that students after the preparatory year need an additional preparatory year, like a general foundation year [SF1_PH2_CM_108].

I suggest a preparatory year independently just for the college of pharmacy, another one just for the college of medicine. So we have our students from the beginning...and we prepare our students as we feel like...but not like what is happening now in PYP...most of the universities are like that around the world [SF3_PH2_CPH_129-130].

As Phase I demonstrated, PYP tutors and coordinators not only expected their advanced-level students to enrol in a medical college, but also expected their elementary-level students not to be offered admission to these same institutions. An analysis of Phase II data confirms their expectations. Most student participants enrolled in the medical colleges had, indeed, been in advanced and intermediate levels from the PYP (see Table 5.9 & Figure 5.2). Very few PYP elementary-level students were among the participants in Phase II samples, with the notable exception of the CN and a few in CAMS. Phase II student participants flagged the absence of lower-level students enrolled in the Medical College as well: "*I haven't met any student[s] from the elementary level who have enrolled in the college of medicine, NONE ... and I feel it's unfair*" [ST6_PH2_FG3_CM_189-193].

Most of the students were so critical of the PYP as to describe it as a '*waste of time*', '*easy*', '*boring*', or '*irrelevant*'. The course was seen as particularly

unsuitable for advanced-level students: “*Writing was boring. We had to make the same thing over and over again*” [ST1_PH2_FG1_CAMS_192].

That such repetition failed to either introduce or reinforce the skills actually needed to meet the writing requirements of their future colleges made the PYP curriculum all the more frustrating. One student lamented that both students and tutors at the PYP alike “*did not take the medical book seriously... it should have been taken seriously*” [ST275_PH2_CAMS_OQ]. Another commented that, “*there was not much focus on the medical terminology and the medical book [EMP textbook] at the PYP. We were not even assessed on it*” [ST3_PH2_FG3_CM_318-319]. Students also criticised the topics they used to practice writing skills when they were at the PYP and suggested that the PYP limit “*unnecessary topics and increase topics related to our field*” [ST109_PH2_CPH_OQ].

From the comparison of the two phases, I found that in the PYP curriculum, EMP, i.e., medical terminology and writing on medical-related topics, was not a main component of students’ course of study. This dissonance between the policy and practice of teaching English for MT students emerged, according to PYP tutors and coordinators, from its difficulty and students’ demand that it be excluded from the assessment process for fear of losing marks (and future hopes of college admission). The analysis of Phase II, however, found that students in the medical colleges were mostly engaged in writing about topics related to their subjects, especially during exams.

Staff members at the medical colleges advocated for the setting of common objectives between the PYP and the medical colleges:

I suggest we have a scaffolding system in teaching writing. You start with the basics at the PYP, and we should have the same objectives so we can build on what you have started...in the coming years...similar to taking a ladder, level-by-level and step-by-step...but we need to know what you [the PYP] are doing. We don’t know what you are doing at the PYP [SF2_PH2_CM_248-253].

Staff were asked in an open-ended question for suggestions as to what the PYP might focus on to improve students' English writing. They recommended a specific course on writing for medical students with an emphasis on medical terminology. They emphasised the importance of teaching students how to write simple, error-free sentences to express their thoughts and opinions:

I suggest they focus on teaching vocabulary and medical terminology and to be able to form simple correct sentences...a focus on writing words with their correct spelling [SF19_PH2_CM_OQ].

More practice, especially in "academic writing, reflective writing and writing about their own opinions" [SF6_PH2_CN_OQ], rather than rote memorisation of grammatical rules, was also a common suggestion.

6.4. Causes of the gaps: a vicious circle

This part aims to answer RQ7. To answer this research question, I looked at the key factors emerging from the data (FGs and interviews) which seem to be causing the misalignments between what was achieved and what is required.

Identifying the gaps between the two phases is important, and along with an analysis of their main causes can help improve the programme (Brown, 2016). This section addresses the last research question regarding the causes of the identified gaps. Inter-related factors were found to be contributing to the gaps between the two phases. The key ones included: a lack of proper NA, a lack of appropriate language admission criteria, and, more specifically, an ineffective approach to assessing writing. A better understanding of why these gaps affect students will be beneficial to curriculum designers and instructors alike for an awareness of the different, context-specific factors that influence students' writing which will help improve students' writing skills. In other words, the analysis of how the institutional and personal obstacles that impede students' progress in English writing came to be is the first step toward developing

solutions. Figure 6.4 below offers a visual representation to accompany the discussion of the interconnected causes discussed in this section.

6.4.1. Lack of proper needs analysis

The PYP English curriculum was designed by policy makers and administrators who drew on both their observations and small-scale studies targeting students in the PYP (Alfehaid, 2014; Alharby, 2005, Shukri, 2014). My study indicates that PYP participants – including tutors and coordinators – have little or no information about actual MHCC writing requirements. One of the PYP tutors discussed their willingness to help students practice the different writing skills required of them in college, but complained that:

[W]e do not have...enough information from the medical colleges to say they need X, Y and Z and then we [can] use that to adapt our curriculum to their needs [TUT5_PHI_FG4_493-494].

Teaching staff at the MHCC are equally unfamiliar with what the students are being taught in the PYP curriculum. For example, when one staff member was asked in an interview whether the PYP had prepared the students for the language skills they require in their college, she responded with another question: *“What are you teaching at the PYP? I don't know the content. I never knew the content at the PYP...or what are you offering there?”* [SF2_PH2_MC_120-121].

English-language programmes for ESP students in Saudi Arabia are too often designed in the absence of a proper NA (Al-Jurf, 1994; Al-Tamimi & Shuib, 2010; Alhuqbani, 2013) and the PYP of the current study is no exception. From the Phase I and Phase II data, when students find pedagogical content irrelevant to their more immediate goals, they fail to engage with and master it. In other words, a curriculum that fails to demonstrate its relevance to the students undermines the learning process (Richards, 2001). The current study offers a detailed NA which aims to bridge this gap between the PYP and the MHCCs. Table 6.18 below

summarizes the results of the gap analysis approach that was followed in this NA study and which will contribute to the development of the writing course at the PYP.

6.4.2. Lack of proper admission policies

Neither the admissions policies of the PYP MT nor those of the MHCCs promote the acquisition of English writing proficiency. This was found to contribute to the identified gap.

Students to be enrolled into PYP MT, they must meet a general standard of achievement, including high scores on the national Achievement Test, Aptitude Test and on their High School diploma (including English) to gain admittance to the PYP medical track (Al Alwan, *et al.*, 2013). There are, however, no specific language requirements for admission, and high school English scores are an unreliable indicator of actual proficiency. Most students graduate from Saudi high school with poor English proficiency (Almulhim, 2001; Al-Gorashi, 1988; Alhawsawi, 2013; Al-Hazmi, 2003; Al-Seghayer, 2005; Sheshsha, 1982; Zaid, 1993), and although the quality of English instruction in high school is questionable (Al-Hazmi, 2003; Al-Jarf, 2008a; Al-Sadan, 2000; Rabab'ah, 2005; Alharby, 2005; Abdulghani *et al.* (2014), high school English scores are still considered for enrolment into the PYP.

After the enrolment into the PYP, student proficiency levels are evaluated by means of a placement test of matriculated students who are divided into three groups (*Elementary, Intermediate* and *Advanced*). One result of this policy is the continued lowering of benchmarks in the PYP curriculum and assessment to address student complaints of perceived unfairness. Consequently, the assessment had to be benchmarked against criteria more suited to the lower level students (A2 on the CEFR as identified from the data).

Yeah. It's all based on what level A [elementary-level students] can accomplish as well. That's one of the main things is that level A must be able to complete the tasks [CO08_PH1_FG6_179-180].

Based on the students' GPA in PYP, students are admitted into the different MHCCs. It is important to mention here, again, that very few students in elementary-level English in the PYP joined the medical colleges, a reality of which both students and tutors alike were aware:

Right, but there's a lot of competitions going on. This is very high stakes for a lot of them because...many of them [elementary-level students] are iffy and many [are] not going to get in no matter what...but they think they're medical students. And we all know this. Come on, a category A [elementary-level]? They're not medical students. They're not going to make it [CO011_PH1_PG6_1089-1094].

Similarly, there are no specific language standards for admission to the different MHCCs. Students' overall GPAs are the main factor considered, yet although students join the MHCCs with very similar GPAs (i.e. the identified ceiling effect), their actual proficiency levels vary greatly. Having no clear language standards for admission to the different MHCCs is problematic, for without clear admissions standards, it would be difficult for the PYP programme to design a curriculum that helps students to meet their college's expectations.

6.4.3. Insufficiently rigorous assessment

Assessment is an integral component not only of any curriculum, but also of teaching, learning, and student achievement as a whole (Kellaghan & Greaney, 2001). This is especially the case for exam-centered students (Javid *et al.*, 2012), where marks play a main role in determining the students' attitudes towards learning. In the current study, my analysis of the ways in which writing is assessed showed that the students, PYP tutors, and MHCC teaching staff follow

similar patterns of behaviour with regards to the evaluation of student achievement. When the assessment targets, higher benchmarks and writing tasks are considered challenging, the pressure to maintain a high GPA leads students to complain and demand both lower benchmarks and easier assessments to secure higher grades (see figure 6.4 below). Staff members accommodate the students' demands. "*But the students will complain*" [SF8_PH2_CN_66] is the response of most participants when talking about why students are not requested to complete more challenging written tasks. Indeed, these students are accustomed to rote memorization of texts for evaluative purposes (Swan & Smith, 2001) and share a culture where marks and scoring high grades are what matter the most (Al-Sadan, 2000). Nevertheless, the complexity of advanced English writing skills is challenging for students, and students currently fail to dedicate the time and effort necessary for improvement. The problems related to assessment are not only relevant to the PYP contexts as can be seen in the following sub-sections:

6.4.3.1. The assessment of writing in secondary education: "*a regurgitation of texts*"

Secondary school education in English writing is, it seems, is under-valued. As one student put it bluntly, "*we have a problem in our school education system. They don't care about English writing that much*" [ST16_PH1_FG3_256-257]. One tutor described the high schoolers' approach to completing the English requirements as the problem when she said they "*learn to pass the exam and not to improve their writing*" [TUT2_PH1_FG5_140].

Many students, however, were critical of an approach to English writing in secondary schools that fails to demand even proper grammar and spelling.

Writing was not that important in school...I haven't written a paragraph on my own before...The teachers make writing very easy, and I know that wasn't for our benefits [ST5_PH1_FG1_85-186].

A common method of instruction and assessment of English "writing" in secondary schools is actually a feat of memorization of pre-given texts. The process of writing, of choosing words, constructing a sentence, and editing it for clarity is replaced with plagiarism in its purest form. Students are given two or three texts to memorise before a test, each on a different topic. The students, then, are evaluated based on their ability to write on one of the topics from memory in the exam. These students do not produce original texts; "*They just regurgitate*" [ST5_PH1_FG4_545]. One tutor explained how this examination model affects post-secondary education from its beginning at the PYP:

Before the exam [at the PYP] they ask us about the topics...they want to memorize [the text] and they [want to] come and write. If you told them to write something they are not ready for, they [would be] terrified [TUT2_PH1_FG5_267-268].

A coordinator "*was shocked [by] students telling [her] that in high school, they never wrote. They just memorized texts...[W]riting is extremely weak.*" [COO1_PH1_FG8_142-143] Students complete their secondary schooling with '*common fossilized errors*' that have never been identified, let alone corrected [TUT4_PH1_FG4_87]. The students' awareness of their poor spelling and punctuation reinforces the habit of using someone else's wording instead of going about the difficult work of writing: "*A lot of these [errors] came from high school, their previous experience*" [TUT4_PH1_FG4_162].

6.4.3.2. Writing assessment in the PYP programme: “*One big blanket for all*”

In the PYP, the students get accepted onto a track (*MT*) based on different admissions criteria (see Chapter 2). The students then sit a placement test at the start of the PYP and are placed into one of the three PYP levels. However, this test does not include a writing component. At the end of the PYP, all students (regardless of their levels) must sit and perform sufficiently well on a high-stakes standardised exam, which includes writing, in order to enrol in the different MHCCs based on their total (GPA) scores. The students’ score in this exam bears a heavy weight affecting the students’ GPA and consequently affecting their enrolment into the different MHCCs.

These policies for placement and admission undermine efforts to improve the students’ writing because, as one of the PYP coordinators stated, “*they [the students] should be selected early in the year so that they don’t even worry it. If they know from the beginning that I am going into this college, so I...should be at a level that I’m ready for*” [CO08_PH1_FG6_500-505]. This is because the students were worrying about their marks more than their actual progress. The students, because of their worries, start to demand easier assessment to more easily obtain high scores and thus successfully enrol in their chosen medical colleges. However, as suggested by participants, if students would have already enrolled into a specific college from the start based on specific criteria, they would give more attention to improving their skills rather than focus on their fear of losing marks.

PYP coordinators have suggested that the selection could be based on high scores in the placement test. Another coordinator suggested that the PYP should not only stop accepting low-scoring individuals, but also consider writing as a component of the placement test:

in this way you have some records to see their performance at the beginning and this is their performance when they exit...Now we cannot measure [CO011_PH1_FG6_509-513].

Requiring high scores on the placement test not only demands high proficiency levels of students but also avoids students placing themselves deliberately in the lower levels to begin with to guarantee high marks later from an easier course: *[W]e'd also avoid the ones who like to score low on the placement test because that's an easy category*" [CO011_PH1_FG6_506-508].

Participants believed that it would be difficult for students from the elementary level of the PYP to join the medical colleges, especially the more competitive ones. They therefore argued that students at the elementary level should not be placed in the MT from the start (as suggested by some coordinators and tutors). One of the coordinators mentioned that:

Some of them [elementary level students] will never get in, but they think they are medical students and we all know...They are not going to make it [CO08_PH1_FG6_1092-1094].

Joining the PYP with low writing proficiency, especially at the elementary level, and fearing to lose marks, students started to complain and demanded easier assessment. Therefore, PYP assessment benchmarks were lowered to respond to students' complaints and satisfy their demands.

The PYP writing exam is standardised across all levels and tracks. In my data analysis, it was found that this standardisation is one of the main causes for low levels of student writing proficiency in the PYP: *"Ever since this whole thing with the standardised against all three levels has come in...there is no progress; it's just gone downhill"* [CO09_PH1_FG6_1121-1122]. This is because the standardised exam is very easy as described by one of the coordinators: *"...the final [writing] exam is very very simple in terms of the word counts, in terms of the prompts, I mean in everything [it is very easy]"*. Most students receive full marks on the final writing exam (see Chapter 1). The assessment in this way turned out to be *'disappointing for most of the students'* [TUT4_PH1_FG4_127], particularly those in the advanced level. One tutor remarked:

[the students] say, 'ok. I have met the benchmark they need me to be at. Then why am I sitting in this English class? Since students do not feel challenged, teachers have hard time keeping their interest [TUT5_PH1_FG4_519-520].

This leads "*the students [to] not take writing seriously*" [TUT2_PH1_FG5_378]. The problem of standardisation was summarised by one of the coordinators as follow:

This is the inherent problem in having a standardised test across the courses, across the tracks. You're going to have to teach to the lowest level and designing to the lowest level. At least for a good portion of the exam. There are a few questions that are a bit more difficult, that are ok but the majority of it is aimed to satisfy elementary and intermediate levels. Advanced is sort of left out there and that's where most of the medical students are going to be [CO08_PH1_FG6_190-196].

In addition, the assessment rubric "*hasn't been changed for years and...also has not been changed from the first to the second semester*" [CO012_PH1_FG6_692]. Assessments include a '*fixed-form prompt*' that is structured in a certain way on general and descriptive topics with tiny word counts.

The assessment team, therefore, considers the current standardised proficiency exam as "*not appropriate*". Students at all levels of the PYP receive very high scores, and there is therefore no way of differentiating between them in terms of their writing proficiency levels; indeed, the exam was very easy for even elementary level students. The students '*get full marks*' on the exam not only because the task has a '*very simple prompt*', but also because the grading rubric is very simple and uses low benchmarks; one of the coordinators commented that "*the rubric has a very low benchmark of A2 according to the CEFR, which is very basic for the whole year-group for both semester one and two...in this way we are not aiming any progress*" [CO09_PH1_FG6_658-660]. Indeed, many of the PYP

coordinators and tutors considered the standardization of assessment as the root cause of lower writing proficiency (the gap). Without specific, tailored, track-based and level-based exams, the students' proficiency in writing is far from being developed during the PYP.

A few other coordinators had a different take on standardisation. They argued that, if properly applied, exams are more challenging as the lower level students would have to compete with high proficiency level students. In this way standardisation would be more effective. Therefore, they recommend better integration of writing throughout the curriculum. It was said that it "*should be institutionalized*" and receive more weight in the assessment to offer an "*external reason*" for the students' interest and teachers' attention. They also specifically highlighted how writing is not located in the syllabus and that limiting students' writing to continuous assessment is not adequate. They too believe that students "*come into the PYP and leave the PYP with very similar skill levels*" [COO1_PH1_FG8_147].

All students in Phase I, regardless of their proficiency levels, agreed on the ineffectiveness of the standardised exam, although their reasons for this belief varied. Elementary level students perceived the test as too difficult for their proficiency. One of the students in the elementary level wrote in the *letter to a friend* (in English), "*You are not going to be tested on what you have [studied] in textbooks because the exam is a standardised exam across the three levels*" [ST27_PH1_LETTER_27]. Another student wrote; "*you need to work very hard in order to be able to pass the exam*" [ST35_PH1_LETTER_35]. Other students mentioned that the exam does not draw on the textbook, so they need to work harder to improve their level in English. Another said "*exams are making extra pressures for us as students in the elementary level*" [ST56_PH1_LETTER_56].

Intermediate level students viewed the exam as easy, but required work to excel or obtain a '*full mark*'. Intermediate students emphasised the importance of preparations before the test in order to do well in the exam: "*in the first semester with some focus you will get the full mark*" [ST200_PH1_LETTER_200].

Students at the advanced levels found the test too easy. One described it as *“below our level,”* [ST489_PH1_LETTER_489] another as *“easier than IELTS tests”* [ST503_PH1_LETTER_503]. These students did not find preparations as necessary as their intermediate peers: *“you only need one day to prepare for the test”* [ST512_PH1_LETTER_512] and *“You can only revise the list of the vocabulary prepared by the university [to be able to pass the exam]”* [ST399_PH1_LETTER_399]. Many advanced students were disappointed by the assessment’s focus on general English: *“I am sorry to tell you [you] will be basically wasting your time per week in order to take a very general exam”* [ST433_PH1_LETTER_433].

The impact of the PYP assessment was clear in Phase II. Medical staff criticized the PYP admissions policy:

what is the use of the PYP when we want to select between students to be enrolled into our college, in English, they all come having full mark...how can a student get such a full mark and when you ask them to write a paragraph, they struggle! [SF4_PH2_CPH_157-158].

Assessment at the end of the PYP is not a true reflection of the students’ proficiency. It was found to be one of the main constraints to students meeting their expected levels. As mentioned above, staff have noticed the different proficiency levels of the students though almost all of them join with similar GPAs: *“we have around 350 students in this college and they all came [to college] with 4.5 GPA or more out of 5.0!”* [SF3_PH2_CPH_173].

After joining the MHCCs, students appreciate the importance of assessment being emphasized in the PYP if they are to develop their writing proficiency. First, the students attributed the problem of the constraints in the PYP English curriculum to the lack of a proper writing placement test: *“writing is different from one student to another...they need to divide us based on our levels”* [ST4_PH2_FG1_CAMS_184-187]. Consequently, they advocate the implementation of a writing skills placement test:

I think we should have been assessed from the beginning and been evaluated on our writing. They should screen those who are good in writing from those who are terrible in their writing. We shouldn't all have the same grammar and activities [ST17_PH2_FG9_CM_686].

They also pointed to the importance of assessing students in English for medical purposes: “[T]here was not much focus on the medical terminology and the medical book at the PYP. We were not even assessed on it” [ST3_PH2_FG3_CAMS_457]. They believed that standardisation in assessment negatively affected the students: “The problem is that they assess you as if you are in the elementary level...therefore, there was no noticed progress with the advanced level students” [ST5_PH2_FG2_CAMS_342-343]. The students also referred to the exam at the PYP that it was easy and the students recommend the need for the PYP assessment “to be more difficult” [ST2_PH2_FG1_CAMS_134]. “One big blanket for all” was an expression given to describe the assessment at the PYP. The way I see it, however, is that the standardised assessment in PYP is a One-Size-Fits-None approach.

6.4.3.3. The assessment of writing in college: *content vs. linguistic elements*

Medical colleges most frequently use both multiple choice questions (MCQs) and short answer questions (SAQs) to assess students. In the latter, students are required to answer questions related to their field of study. Although only requiring the writing of a short text, SAQs are considered demanding prompts because of the different skills required to answer this style of question (as explained in Chapter 5). The students might be asked to “list, analyse, recall, interpret and to be creative”, for example [SF2_PH2_CM167-169]. The staff reported that the first question students ask before any exam is “whether there is a writing [component] in the exam or not? And if yes, [they ask] whether language will be assessed” [SF4_PH2_CPH_41-42]. Often, an understanding of the course

content is not a problem, but rather the students' abilities to express themselves cogently: "[They] can tell you the answer and what they mean, but it is difficult when it comes to writing" [SF1_PH2_CM_186]. Also "it is easier for the students to give you the answer in points but not to write the answer in one comprehensive, complete paragraph" [SF3_PH2_CPH_47-49]. In this respect, one of the staff members mentioned that "during the exam most of those who call me to ask questions [have] questions that are language-related and not scientific-related" [SF4_PH2_CPH_53-54].

Generally, staff do not deduct marks for spelling or grammar, although it was reported that grading exams riddled with grammar and spelling errors was a very annoying process. Some staff members do include language as part of their assessment and "consider 2% of the total mark on their language, though students do not like that" [SF3_PH2_CPH_44-45]. Many others do not consider the language and the linguistic elements in their assessment at all. When I asked why this was such a common practice, one instructor replied, "because some students are really...concerned about marks, right?" [SF8_PH2_CN_173-174]. Students' poor writing proficiency and a desire to accommodate the students' actual competencies were also justifications for this practice. One of the staff explained how the choice to include written questions depends on different factors: "depends on the subject, the part we are assessing, my time and the number of the students...if I have like about 120 students, it will be impossible to mark them all and MCQs or matching will be the easiest way to mark" [SF4_PH2_CPH_28-29]. This lack of common criteria regarding assessment was noticed by the students themselves: "it depends on the doctor [staff members]. They have no clear criteria. One is requesting us to write a lot [while] another wants us only to mention points...[when answering exam questions]" [SF3_PH2_FG1_CAMS_134-135].

Instructors also limit the writing of reports and essays in their courses because students "have poor English". Staff are aware that the students are being short-changed by this decision: "[Students] are supposed to, but...we know their capacity"

[SF6_PH2_CN_421]. Some staff members justified eschewing essay questions because:

They are still making mistakes. Writing for them is still difficult because they do not have the linguistic medical terminology. The students can express their ideas in Arabic, but they cannot in English. They need more time to gain more medical and nursing terminology [SF10_PH2_CN_114-115].

Most others, however, cited frustration with the received assignments and despair at improving the situation as their main reason for avoiding written evaluations of any kind. Indeed, when asked to submit written assignments, staff expect most students “*will plagiarize and copy-paste from the internet*” [SF5_PH2_CPH_158]; “*we do not ask for written assignments because students will plagiarize*” [SF4_PH2_CPH_36]. The CPH staff accommodate students’ weaknesses and sometimes “*have to give up requesting and give up penalizing for the weak writing skills*” [SF3_PH2_CPH_55]. Another followed with:

[H]onestly, this is a problem we are facing. I cancelled many assignments because what the students will do is to go to the website and copy and paste it and I will find the same exact answer with other students...so why should I bother? I decided not to give assignments...It is just an extra effort, and they still don’t know how to paraphrase and summarise [SF7_PH2_CN_266-274].

Students find “*writing using their own words*” [SF2_PH2_CM_74] very difficult. Another staff member echoed that “*students in general are not strong in writing. They have phobia of writing... [W]hen they are asked to write even a short note or to complete a gap they beg me not to include such questions in the exams*” [SF5_PH2_CPH_171-174]. Instructors give in to the pressure: “[B]ecause the students will complain...because they say it is difficult for us to memorise and

remember the spelling...the students will complain...[about] being tested on their language and not on their knowledge.” [SF7_PH2_CN_66-69].

To summarise, the assessment practices in the PYP and in the MHCCs were identified as one of the main causes for misalignment in the current study. Because of the students’ low proficiency, writing was considered too difficult to include in a “fair” assessment of students’ performance in a course. Being omitted from the assessment, however, the practice of writing was not taken seriously by staff, who preferred to avoid conflict with their students, whose future performance suffered as they failed to acquire the skills to communicate complex thoughts in written English.

6.4.4. Students’ proficiency in writing

The students’ descriptions of their problems when writing in English reflect those of lower proficiency English writers in general. They identified difficulties when ‘*writing different types of texts*’, ‘*express[ing] their opinions in writing*’ and ‘*writing using the correct grammatical structure*’. This generally low level of proficiency in writing was also reflected in other studies exploring Saudi students’ writing (Al-Ghamdi, 2006; Almoallim *et al.*, 2010; Ghobain, 2014, Nazim & Hazarika, 2017). To address this low level of proficiency in writing, the PYP curriculum must overcome certain customs if it is to help students succeed in their communicative tasks and excel in the MHCCs.

Students plagiarising in written assignments and tasks is often the norm in both the PYP and the MHCCs, as it is elsewhere. El Tantawi *et al.* (2016), for example, found that in completing a written task, 87% of the first-year dentistry students had plagiarised. The unreliability of written assignments as an indicator of a student’s actual proficiency and the lack of clear action taken against plagiarism

by staff members and university policy-makers has encouraged PYP tutors and MHCC teaching staff to avoid written assignments.

The great disparity in English proficiency between the PYP and the MHCCs has encouraged staff to change how they approach the task of teaching. Indeed, the PYP tutors and MHCC staff repeatedly mentioned that they have to lower standards to accommodate lower proficiency levels and maintain a positive learning environment. One PYP tutor described how students considered writing a threatening skill because they risk losing marks due to poor writing; consequently, tutors avoid assessing students' writing wherever possible:

PY[P] is so high stakes, there is this kind of fear of giving something a huge chunk of marks when could of end up very subjectively marked like writing...just that at the end of the day we want to be as fair, you know in quotation marks as fair as possible and as objective as possible [COO2_PHI_FG8_148-155].

Both the PYP and the MHCCs have accommodated to the students' focus on competitive grading rather than real learning by lowering standards and avoiding the assessment of student writing. Students, unfortunately, perceive the exclusion of writing from their assessments as a sign that it is not essential for their overall success.

The inter-related factors working in a vicious circle (discussed above in section 6.4) contributing the identified gaps are visualised in Figure 6.4 below. To give an example of how this vicious circle might work, we can start by saying that the students are under tremendous pressure to obtain a high GPA and the teachers, therefore, are forced to act to accommodate the students' fears and pressures (e.g. by lowering standards, leading to writing being considered less important) which contributes to the vicious circle.

6.5. Phase III Discussion

This third phase of the study was a gap analysis. I compared the students' actual proficiency and writing skills achieved in PYP, as discussed in Phase I, with the proficiencies required in the MHCCs, as discussed in Phase II. The discussion suggests an answer to the main overarching research question, namely, whether the English course at the PYP has adequately prepared students to succeed in the first-year MHCCs. The last part of this chapter discusses the implications of this study and offers recommendations to minimize potential problems for future PYP cohorts. Wherever possible, I connect my suggestions to the relevant scholarly literature.

6.5.1. Does the PYP curriculum adequately prepare students to meet the writing requirements of the MHCCs?

6.5.1.1. General overview

The PYP, in general, has adequately prepared its students to meet the writing requirements of first-year students in the MHCCs, since the quantitative findings indicate, for six of the CEFR scales based on the student and staff data (*Overall written interactions, What students can write, Types of texts students can write, Grammatical accuracy, Processing texts, Note-taking*), there was no statistically significant difference between the two phases. Additionally, in two of the CEFR scales (based on students' data) significant variances were identified, namely *Overall written production* and *Orthographic control* scales, where the students felt that their proficiencies were higher than required. These two observations of the analysis at the macro-level suggest that the PYP has adequately prepared its students to meet their colleges' expectations.

There were, however, important factors that might expose problems behind this similarity when we explored the data in-depth. For example, an analysis of the

qualitative data demonstrated that although neither the PYP nor the first-year MHCC curriculum required much writing, the MHCC staff largely regretted this fact and felt it was not to students' benefit. In other words, they complained that in the MHCCs they had to 'dumb down' the writing requirements to accommodate a gap that is too large and complex and, thus, too difficult to address in any other way (discussed further below).

An analysis of the sample demographics (in the cross-tabulation analysis, see Table 5.9 and Figure 5.2) unveils important points which may explain the similarities identified in the quantitative analysis. For example, most of the students who participated in Phase II came from the higher PYP levels (intermediate and advanced). In other words, those students from the elementary level rarely enrolled in Phase II, yet if their data had been included in the analysis, we would likely have had a different picture.

In addition, the students who usually join the MHCCs are the highest achievers at high school in comparison to the other non-medical disciplines (Alshehri, 2001; Al Alwan *et al.*, 2013). This is particularly true of those joining the medical colleges in Saudi Arabia (Alharby, 2005; Al-Shehri *et al.*, 2013). For example, Alharby (2005) found that the students who join the medical colleges are the ones "who were likely among the top of their classes" especially in English (p.145). Thus, the similarities identified in the quantitative data are not clearly sufficient to provide an accurate picture of how well the PYP English curriculum prepares all students for their first-year of medical college. The few misalignments identified by the quantitative analysis, along with the in-depth exploration of the data, however, offer an excellent starting point for further analysis and discussion.

6.5.1.2. The identified gaps

The quantitative analysis of Phases I and II revealed significant variances in the *Vocabulary range and control* and *Reports and essays* scales. In addition to those two scales, *Grammatical Accuracy* and *Note-taking* scales also required higher levels than achieved (when the data were compared between PYP tutors and MHCC students). Moreover, the sample group's CEFR levels, as identified by the raters, further confirm the MHCC students' and staff's perceptions that the PYP does not adequately prepare students to meet MHCCs' requirements.

Considering the fact that the CEFR scales refer to general language skills (CoE, 2001; Saville, 2005) and not the more specific language requirements of medical students' chosen disciplines, the findings should be especially worrying for the curriculum planners of the PYPs in Saudi Arabia, given that their main goal is "to provide a strong foundation of English required for specific professional disciplines and...to achieve the desired proficiency level required for various professional disciplines" (Nazim & Hazarika, 2017, p.145).

Students in the PYP were required to use only very basic general writing skills, with very limited practice. Not much attention was given to writing compared to the other language skills, and the curriculum focused on descriptive (as opposed to analytical or synthetic) topics such as '*family members*' or the '*daily routine at the university*'. In the MHCCs, however, students were required to write exclusively on medical and subject-related topics (see summarised list in Table C6 in appendix C). The complex writing skills that teaching staff and students alike identified as required by the MHCC curriculum include '*summarising*', '*paraphrasing*', and '*synthesising information from different sources*'. Students were also expected to '*analyse*', '*argue*', '*describe*' and '*write a reflection on a case*' as part of their course assessments in the MHCCs. Students felt they needed a larger lexicon, a better methodological approach to research, argument construction and the ability to write in different genres to succeed in many of their courses. These were not skills they were able to practise at the PYP.

The significant differences between the two phases in respect of the *Reports and essays* and *Vocabulary range and control* scales were largely consistent across the different colleges and proficiency levels. This means that even advanced students have difficulty writing essays and reports and writing with the specialised terminology of their fields in their first-years in MHCCs. These difficulties with writing and vocabulary were also reported in other similar contexts in Saudi Arabia (Al-Eissa, 2008; Alharby, 2005; Doushaq, 1986; Fageeh, 2003; Shukri, 2008; Shukri 2014). Writing academically in discipline-related topics is, of course, a complex task and requires a high level of language proficiency (Kroll, 1990) involving the control of vocabulary, grammar, spelling, structure and cohesive devices in the context of specialised scientific and medical topics (Hedge, 2001; Shokrpour & Fallahzadeh, 2007; Al-Qhatani, 1999; Lucas *et al.*, 1997; Shukri, 2008).

Two MHCCs, the CN and CAMS, revealed larger significant gaps between the CEFR levels achieved versus required, compared to the other colleges in relation to the *Reports and essays* scale. My analysis of the FGs, interviews and students' mind maps revealed that the students in these two colleges were asked to write essays and reports more often than their peers enrolled elsewhere. In addition, these two colleges had a higher proportion of elementary students (see Table 5.9). Perhaps the additional writing exposure at the MHCC led these students to select a higher CEFR level to describe what was required of them, leading to a larger gap in those two colleges. Alternatively, it may be that these students lack the proficiency to assess their skills (see Chapter 4 above; Kruger & Dunning, 1999). Or the level required was the same as the one for students at other colleges, but these students at these colleges on average start off with a lower attainment level, so the gap is bigger to reach the required level than those in higher PYP levels to start with. I prefer the last two explanations here because a detailed analysis of the results shows that although the students (in CN and CAMS) have not chosen higher requirement levels in these two colleges, the gap is nevertheless bigger because of their lower proficiency.

Furthermore, the students in the first year of the MHCCs are expected to have at least some familiarity with basic medical and academic terminology. Although the textbooks for *English for Medical/Academic Purposes* (EMAP) are integral to the PYP curriculum, little attention was paid to them by either the PYP tutors or students in the teaching-learning process. The PYP curriculum largely focused on general vocabulary, whereas the medical terminology that forms an integral component of medical textbooks and classroom comprehension was neglected. The very few parts of the textbook that were covered in class were ignored by PYP students because they were not a component of their assessments. Further analysis of this gap in the students' CEFR level in the *Vocabulary range and control* scale revealed that some participants were not even familiar with such basic medical terms as *injection* and *abdomen*. Students repeatedly spoke of their difficulty with the basic medical terminology in both the questionnaire and in the FGs. This problem is also not unique to my study, but similar findings are reported (for example, in Al-Eissa, 2008; Alharby, 2005; Doushaq, 1986; Fageeh, 2003; Shukri, 2008; Shukri, 2014).

The complexity of L2 writing is widely acknowledged (Khuwaileh & Al Shoumali, 2000; Lucas *et al.*, 1997; Shukri, 2008). It is even more demanding when it occurs within an ESP context where students are expected to not only master content but also the vocabulary used to express their understanding of that content in their writing (Shukri, 2008). In his 2017 study of international students, Franklin-Landi found that the students identified medical terminology as one of their primary needs (about 51%). Slang and idiom use by staff has also been identified as another vocabulary-related obstacle to English as a Second Language (ESL) student success (Önder Özdemir, 2014). The PYP students' lack of vocabulary, and more specifically medical terminology, was identified as the main source of difficulty, which affected not only their writing skills, but their general communication skills in the language of instruction (Al Makoshi, 2014; Nazim & Hazarika, 2017; Alfehaid, 2014; Shukri, 2014). This difficulty with writing among Arab-speaking learners has been investigated at length (Ahmed,

2010; Doushaq, 1986), also for Saudi-learners (Al-Ghamdi, 2006; Almoallim *et al.*, 2010; Ghobain, 2014, Nazim & Hazarika, 2017).

Orthography (and mainly spelling) is especially important for MHCC students studying in high-stakes environments where a misplaced letter or two has the potential to cause sickness or even death. Yet, here too, the analysis identifies how its neglect in the PYP curriculum and its exclusion from direct assessment at the MHCCs created obstacles to student success. The staff were explicitly frustrated by students' frequent spelling mistakes and the fact that they had to accept such poor spelling. Teaching orthography and emphasising the importance of writing with correct spelling for clear and professional communication should start in the PYP or maybe even at school from an early stage. MHCCs, should likewise include students' spelling as a marking criterion in writing assessments to foster the improvement of this important skill.

Table 6.18 below summarises this current gap analysis to NA study. The first column pinpoints the main findings of Phase I (present-situation). The second column focuses on Phase II key findings (target-situation). The third refers to gaps identified from comparing the two Phases in Phase III. The fourth column summarises the potential causes based on the gap analysis. The last column presents some potential solutions and recommendations to bridge these gaps.

Table 6. 18 Summary of Gap Analysis

Present situation Phase I	Target situation Phase II	Gap Analysis Phase III	Potential causes	Recommendation for improvement
Average CEFR level achieved ranges between A2+ and B2	Average CEFR level required is B2	<ul style="list-style-type: none"> - Some overlaps between the CEFR levels achieved and required - Generally small gaps between the two phases in some scales - Significant gaps in the following CEFR scales: <ul style="list-style-type: none"> • Vocabulary range and control • Reports and essays • Grammatical accuracy • Note-taking 	<ul style="list-style-type: none"> - Lack of proper preparations due to lack of proper needs analysis - Lack of proper admissions procedures - Limited practice - Dumbing-down requirements at the PYP and MHCCs 	<ul style="list-style-type: none"> - PYP needs to set clear learning objectives and design/utilize materials that meet those objectives - Give more attention to writing in terms of teaching, practice and assessment. - Clear standards students are required to meet to be set by the PYP and the MHCCs

<ul style="list-style-type: none"> - Very limited writing practices - Focus on very basic writing skills 	<ul style="list-style-type: none"> - The students need writing for exam purposes and occasional written assignments (<i>highly</i> required) - The students need to write reports, essays and a 'mini research' project (<i>partially</i> required) - Students need to write in specific genres for their specific profession (<i>marginally</i> required) 	<ul style="list-style-type: none"> - Different focus between the two phases - Limited time and practice of the required writing skills 	<ul style="list-style-type: none"> - Writing is not sufficiently covered in PYP curriculum - Writing is perceived less important - Lack of focus on writing may be due to limited proper needs analysis - Insufficient assessment procedures - Students' complaints 	<ul style="list-style-type: none"> - PYP curriculum needs to be modified to give more attention to teaching writing. - Writing also needs to be given more weight in assessment to be taken seriously - Practice writing that is relevant to students in the first year
<ul style="list-style-type: none"> - General topics 	<ul style="list-style-type: none"> - Medical/subject-related topics 	<ul style="list-style-type: none"> - Gap in the choice of topics (general topics do not ensure students acquire the language related to medical discourse) 	<ul style="list-style-type: none"> - Plagiarism - Dumbing down of requirements - Accommodation of students' proficiency 	<ul style="list-style-type: none"> - Choose topics that are related and general enough to meet the writing requirements

				of the students in the different colleges
- General vocabulary	- Medical and subject-related terminology	- Gap in the vocabulary focus		- Beside the general vocabulary, basic medical and academic terminology needs to be practiced in PYP
- Many grammatical rules, but decontextualized	- The students need to have correct grammar when writing (correct and clear sentence structure)	- Gap in the way grammar is taught at the PYP		- Grammar should be taught and practiced while writing
- Strong focus on punctuation - No consideration on teaching spelling rules	- Limited focus on punctuation - Spelling is considered important	- Gap in the focus and the use of punctuation		- The use of punctuation marks will be mastered through the practice of writing

				<ul style="list-style-type: none"> - PYP needs to prepare students to improve their spelling and to practice common spelling rules, focusing in particular on students' most common errors
Descriptive skills	<p>Argumentative, narrative, descriptive skills</p> <p>Summarising, paraphrasing, analysing skills</p>	<ul style="list-style-type: none"> - Gap in the focus of the rhetorical modes and writing skills introduced in PYP 		<ul style="list-style-type: none"> - Different writing skills and modes need to be accounted for in the PYP curriculum

<p>Assessment of writing in the PYP:</p> <ul style="list-style-type: none"> - Very easy - Requires descriptive skills - On general topic - Carries the least weight in assessment 	<p>Assessment of writing in MHCCs</p> <ul style="list-style-type: none"> - Short-answer questions (SAQs) - Discipline-related topics - Requires more complex writing skills - Quality of language not taken into account 	<ul style="list-style-type: none"> - Writing is not perceived important - Writing is not taken seriously - Ceiling effect affects admission to MHCCs - No perceived progress/improvement - Dissatisfaction with the PYP programme 	<ul style="list-style-type: none"> - No clear objectives - No constructive alignments between the students' learning outcomes and the assessment tasks - Insufficiently rigorous assessment 	<ul style="list-style-type: none"> - Level-based assessment - Stringent rules for assessment - Context-related standards - Constructive alignments between student learning outcomes (SLOs) and assessment
---	--	--	--	--

6.6. Implications

From the summary of the gap analysis to needs identified in table 6.18, the following implications can be considered for improvement. These implications can be generalized not only to similar contexts where English writing is taught in EMI, but also to language programmes nationally and internationally involving Saudi and Arab students.

6.6.1. The CEFR-based-context-related standards

The CEFR, as stated by its proponents, focuses on general skills and should be used as a reference tool to be adapted to the context where it is used (CoE, 2001). McNamara *et al.* (2018) found that whereas CEFR descriptors have some relevance to academic writing skills, they do not cover the full complexity of academic writing. Indeed, Deygers, Zeidler, Vilcu, and Carlsen (2018) argue that the CEFR needs to be recognized “as a general theoretical framework that needs to be supplemented by language-specific and context-specific descriptors” (p.13). Table 6.19 below summarises the standards proposed based on the data collected for this study and are intended as proposed learning outcomes for the PYP curriculum. “In general terms, students learning outcomes (SLOs) are statements of what the NA shows the students should be able to do by the end of training (that is, by the end of the course or program)” (Brown, 2016, p.155). These standards can also be used as admission criteria for the medical colleges.

As suggested by West (1994), the results of NA should be translated into learning objectives or standards, as referred to in this study. In the table below, the standards written in black are those based on the means of the CEFR levels required. Those in blue are common to all the colleges and especially relevant to

the PYP. The other colours in the last column represent the required skills specific to each college: green for the CM, purple for the CAMS, red for the CN and orange for the CPH. No standards could be identified for the CD as no one from this college - neither students nor staff – participated in the qualitative part, as mentioned in Chapter 3.

When formulating these standards, I stated the descriptors identified from the analysis of the quantitative data when participants selected the CEFR scales that are minimally required (i.e. the descriptors of the B2 level on CEFR scales). I have also added *supplementary* descriptors for the identified skills that are not covered within the CEFR descriptors and, as suggested by North (2014), these descriptors are colour coded and follow the same style as the CEFR descriptors; as such they are “positive, brief, clear, concrete, standalone criterion statements” (p.143).

The students’ MHCC requirements as identified in the current study are intended to be appropriate for use both as student learning outcomes (SLO) for the PYP and as admissions criteria for the MHCCs. With common and transparent standards, the gaps can more easily be bridged and potentially all the participants would be on the same page. The proposed standards would apply to all MHCCs but would also include some college-specific standards as identified from the data.

Table 6. 19 CEFR-based-context-related standards

	General Standards	College-specific standards
Overall written production	<p>Can write clear, detailed texts on a variety of subjects related to his/her field of interest, synthesising and evaluating information and arguments from a number of sources (B2)</p> <p>Can write up research using academic vocabulary without plagiarising</p> <p>Can write a well-structured paragraph</p>	
Overall written interactions	<p>Can convey information and ideas on abstract as well as concrete topics, check information and ask about or explain problems with reasonable precision. Can write personal letters and notes asking for or conveying simple information of immediate relevance, getting across the point he/she feels to be important (B1)</p> <p>Can write and respond to staff emails</p> <p>Can write a formal letter to university administrators</p> <p>Can write their curriculum vitae (CV)</p>	
Types of texts	<p>Can write a variety of different texts (B2)</p> <p>Can answer clearly short questions with correct structure and spelling</p> <p>Can write assignments using different resources without plagiarising</p> <p>Can write and respond to teaching staff emails</p> <p>Can write up research on general-medical related topics</p> <p>Can write (short) essays on general, medical-related topics</p> <p>Can write a poster/brochure/survey on a medical-related topic (e.g. depression)</p> <p>Can write scientific research without plagiarising</p> <p>Can write references and cite correctly</p>	<p>Can write clinical reports</p> <p>Can write documentation</p> <p>Can write a diagnosis</p> <p>Can write a case study</p> <p>Can write lab reports and diagnostic report</p> <p>Can write medical/patients reports</p> <p>Can write a case report</p> <p>Can write a reflective essay</p> <p>Can write medical prescriptions</p>

Vocabulary range and control	<p>Has a good range of vocabulary for matters connected to his/her field and most general topics. Can vary formulation to avoid frequent repetition, but lexical gaps can still cause hesitation and circumlocution (B2).</p> <p>Lexical accuracy is generally high, though some confusion and incorrect word choice does occur without hindering communication (B2).</p> <p>Can write using basic medical terminology</p> <p>Can write using a range of vocabulary (using synonym)</p> <p>Can choose vocabulary that is suitable for the context</p> <p>Has a large range of vocabulary and is able to deploy it appropriately</p> <p>Can write using academic vocabulary</p>	<p>Can write and use the vocabulary related to each discipline (see summary table C6 appendix C)</p>
Grammar	<p>Shows a relatively high degree of grammatical control. Does not make mistakes which lead to misunderstanding (B2)</p> <p>Can write comprehensible sentences with no grammatical mistakes</p> <p>Can express ideas and thoughts in writing using correct grammar</p> <p>Can connect sentences accurately using <i>and, but, or</i></p> <p>Can write using simple and understandable grammar</p>	
Orthography	<p>Can produce continuous writing which is generally intelligible throughout. Spelling, punctuation and layout are accurate enough to be followed most of the time (B1)</p> <p>Can write the medical terminology with correct spelling</p> <p>Can write with correct spelling especially when answering exam questions</p>	
Processing texts	<p>Can collate short pieces of information from several sources and summarise them for somebody else. Can paraphrase short written passages in a simple fashion, using the original text wording and ordering (B1)</p> <p>Can write an analysis of a situation</p> <p>Can write a description of a situation</p>	

	<p>Can summarise and paraphrase short texts and short articles</p> <p>Can write a reflective piece</p> <p>Can understand and write a summary of a text</p>	
Reports and essays	<p>Can write an essay or report which develops an argument, giving reasons in support of or against a particular point of view and explaining the advantages and disadvantages of various options. Can synthesise information and arguments from a number of sources (B2)</p>	<p>Can write discipline-specific essays and reports</p>
Note-taking	<p>Can take notes during a lecture which are precise enough for his/her own use at a later date, provided the topic is within his/her field of interest and the talk is clear and well-structured (B1+)</p> <p>Can understand a clearly structured lecture on a familiar subject, and can take notes on points which strike him/her as important, even though he/she tends to concentrate on the words themselves and therefore to miss some information (B2)</p> <p>Can take notes about the important points in a lecture</p>	

The standards written in black are those based on the means of the CEFR levels required. Colour-coded standards are the supplementary ones that I have created from the data in the FGs and interviews: standards in blue are common to all MHCCs, in green related to the CM, in purple related to CAMS, in red related to CN And in orange related to CPH.

6.6.2. Implications for the PYP curriculum

6.6.2.1. The PYP curriculum: bridging the gaps between present and target situations

The first identified gap concerned the minimum CEFR levels required to join the MHCCs. The CEFR levels in the current curriculum framework do not reflect reality. Students leave the PYP with lower CEFR levels than those specified in the curriculum framework as the exit levels, and also lower than those required by the MHCCs.

If we are to consider the CEFR scales as those focusing on basic general skills (CoE, 2001), and B2 was found to be the minimum requirement, we would expect to find students facing difficulties in their general writing proficiency post-PYP. This was confirmed by the analysis of the qualitative data, where the students and staff members highlighted the students' writing difficulties and showed concern for their limited writing abilities in areas such as summarizing, paraphrasing, writing using their own words, and expressing their opinions (basic general writing skills for the CEFR B2 level).

PYP policy-makers and curriculum designers would do well to modify the current curriculum framework so that it reflects the CEFR levels required of the MHCC students. Furthermore, in order to use the CEFR in a more "meaningful" and "useful" way (Lowie *et al.* 2010), the levels identified for the curriculum framework should be constructively aligned to the teaching activities and the assessment of writing at the end of the PYP (Biggs & Tang, 2011; Fulcher, 2013). The CEFR-based standards in Table 6.19b are offered as a starting point for a modification of the current approach.

The data also unveil the pressing problems posed by students' limited vocabulary and their lack of medical terminology. These findings suggest that the PYP needs to take serious steps to work on improving students' vocabulary use in writing

and their knowledge and control of basic medical terminology. One simple, and cost-effective solution would be to utilise more fully the EAP and EMP textbooks in teaching and in the final assessment. It would also be helpful to identify the basic medical terminology that is common across the MHCCs and to promote its use among students' in their written assignments.

PYP tutors can also address students' poor orthography by understanding its roots. El-Hibir and Al-Taha (1992) found that students' L1 (Arabic) is the main reason for poor orthography and recommended teaching spelling with reference to the common errors (El-Hibir & Al-Taha, 1992). Furthermore, it is important to highlight the differences between the pronunciation and the orthography of a given word to help students make connections to using correct spelling. Moreover, some weight in assessment should be dedicated to spelling so as to emphasise its importance to students.

Overall, writing needs to be taken more seriously in the PYP curriculum. Curriculum designers need to assign more time in the PYP curriculum to expose students to different genres and on different topics, including both general and medical topics. Students' writing should be "*intensified*" in the curriculum and practised consistently: "*Consistency is the key*" [TUT5_PH1_FG4_490].

The PYP's focus on grammar, and particularly punctuation, is typical for many similar institutions (Javid & Umer, 2013), yet research suggests that when students are exposed to different grammatical rules without enough practice in applying them in their writing, they show little improvement (Nunan, 1998; Weaver, 1996). The PYP curriculum, then, would likely be improved by integrating the teaching of punctuation into a writing program that is focused on the interplay of syntax, grammar and orthography, and how these aspects in conjunction contribute to text quality, rather than focusing on one, minor grammatical element over all the others.

6.6.2.2. The PYP: who is it for?

Most PYP programmes in Saudi Arabia admit any student who meets the admission criteria regardless of their actual English proficiency. In other words, even students with very advanced proficiency in English join the programme alongside those with low skills (Rauf, 2015). The current PYP curriculum is especially unsuitable for advanced-level students, who are most likely to join the medical colleges for which the curriculum is intended to prepare them. PYP policy-makers and curriculum designers should reconsider their objectives and materials to target the more advanced and challenging learning outcomes required for the MHCCs. By raising these standards, lower-level students would also benefit from increased standards as well as more practice. Currently, advanced-level students consider the PYP to be 'boring' and a 'waste of time.'

Another alternative would be the exemption of students who can demonstrate a good mastery of English from the PYP, as is the case for many international students from English-speaking countries (Rauf, 2015). However, it is important to mention that those students are expected to join specialised colleges where English is the main medium of instruction. Even high-level students find the specialized subjects difficult at the MHCCs and a curriculum at the PYP redesigned to teach medical terminology would be of benefit to all. Another important point to be raised is the future of the students with low English proficiency. It was found that students in the elementary level (though having close results to their peers in more advanced levels as seen in Chapter 1) could not make it to some specific colleges (e.g. the CM) because of the small discrepancies in their GPA compared to students in higher levels. The questions to be raised here then, are whether it is fair for students with poor proficiency to be deprived from joining specific colleges just because of their low English proficiency. Do students with better English make better doctors? Or would those less proficient in English make equally good doctors, given more language support? Or are we are doing justice (*being fair*) to higher level students by

lowering the assessment benchmarks at the PYP to accommodate the students in the elementary level who probably would not end up joining the MHCCs? These are important issues that warrant further investigation.

6.6.3. Implications for the MHCCs

6.6.3.1. The need for additional writing support

The PYP is similar to many university English language programmes that prepare students for EMI disciplines through the provision of an intensive English programme (Murray & Nallaya, 2016). The PYP, however, is the *only* language support the students have before the MHCCs.

MHCC staff and students have blamed students' low proficiency in writing on the PYP. High expectations of proficiency upon arrival are based on the assumption that students who have already secured a place at the university should join these colleges (MHCCs) equipped with the academic literacies they need for their study degree (Gunn, Hearne & Sibthorpe, 2011). Yet the PYP and many similar English programmes usually "fail to reflect the pluralistic nature of academic literacy and the fact that different academic disciplines are characterised by specialised vocabularies, concepts, and knowledge" (Murray & Nallaya, 2016, p.1299). To prepare the students in all the specialised literacies (Rex & McEachen, 1999) which include specific genres, arguments, rhetorical structures and terminology for each discipline (Murray & Nallaya, 2016), is an enormous task beyond the scope of such a short curriculum that also needs to devote time to grammar, syntax and structure.

Indeed, even among the MHCCs, each college has its own specific literacies that need to be acquired along with the general skills that are common to these colleges (Moore & Harrington, 2016). Therefore, as suggested by the present study and previous research, continuous language support is needed (Nazim & Hazarika, 2017, Shukri, 2014). Students join the medical colleges with varied

levels of proficiency. Targeted language support should be offered based on the students' needs: some need support at the sentence level and others at the level of structuring arguments relevant to a particular genre. Students need to have language support and guidance to help them construct their writing in a meaningful way that is suitable for the specialised genres related to their field of study. Having English language teachers based in each college to collaborate with the subject teachers to help students improve their writing could be one possibility. A writing centre or an advisory service at the university is another possible option. It is worth quoting a CM student at length here:

Usually, they give us short courses about research basics...It includes how to use SPSS and how to do analysis...These things are easy to understand but what we need is to focus on is the linguistic parts [when doing research]...we need them to focus on writing scientific research, which unfortunately is unavailable...there are no courses on how to use vocabulary when writing...there is a difference between HOW TO DO which is not as problematic as HOW TO WRITE. [ST_PH2_FG3_133-139].

The student here complains that the mechanics of academic and medical writing are taught without any focus and support around the language of their writing at the MHCCs. Without such additional language support, they will “struggle to produce written work of the required standard” (Murray & Nallaya, 2016, p.1298).

What becomes evident from the data is that both students and staff often confuse basic writing skills and the skills that should be developed as part of becoming conversant in the academic literacies of their discipline and developing the relevant professional communication skills. As suggested by Murray and Nallaya (2016), it is important to understand the distinction between academic literacies required for each discipline and general language skills – or general proficiency. For example, the data suggested that the students needed to practice writing a summary on a general medical topic (*general proficiency*). The data also suggested the need, for example, to prepare students to write specific clinical

reports that are required of pharmacists or to write nurses' patient documentation (*academic literacies*), which are difficult to deal with while they are at the PYP. As discussed in Chapter 5, the PYP should prepare students to join the MHCCs with a general proficiency that enables them to cope with the complex writing demands of their first year in college. The PYP should also equip students with basic academic and medical writing skills that can serve as the basis for mastering the academic literacies and professional communication skills following entry into their respective colleges (Murray, 2013). Students at the PYP should be trained in reading about more elementary medical topics and then given opportunities to practise writing a summary or an essay based on what they have read.

Some participants have called for a preparatory year tailored specifically to reflect the particular language demands of each college; that is, students would begin to develop conversancy in the future academic literacies they will require whilst in the PYP. Another possibility is for the PYP to focus solely on improving the students' basic and general writing skills which, as Russell *et al.* (2009) note, address the students' writing difficulties at the surface level, focusing, for example, on grammar. Additional language support could then be made available, following their completion of the PYP, to help improve students' academic literacies, to be taught in parallel with their studies for their discipline. This additional language support post-PYP could be organized in collaboration between the PYP English language tutors and the content teachers, to help students in their content writing. The collaboration would allow the English tutor to look at the linguistic elements of the students' writing whereas the content teachers would assess the content. This collaborative approach has been shown to be beneficial in developing students' academic literacies (Jacobs, 2007; Murray & Nallaya, 2016; Shukri, 2008). Here I suggest that the teaching staff continues to assign written materials and give feedback on the content, while the English language teacher gives feedback on the linguistic elements. Another suggestion is to train content teachers (academic teaching staff in the MHCCs) to assess and help students improve their writing. This would be a cost-effective solution in

terms of resources, and also more focused, since students are going to get feedback on both language and content simultaneously from the same person. This technique, however, is not always practical as we are dealing mostly with non-native English teaching staff who themselves may require improvement in their skills in writing and providing feedback on the students' writing.

Given the current situation at the PYP and the MHCCs, I suggest that students be prepared during the PYP to write academically on a range of general topics (based on the students' interests). In their preparations for academic writing, the students need to practice basic research writing skills and to be introduced to the concept of plagiarism and how to avoid it in their writing. I also suggest offering the students a glimpse of some of the different genres required in the MHCCs, along with basic medical topics. This would allow for a broader exposure to medical terminology.

6.6.4. The need for rigorous policies

6.6.4.1. Admission policies

As mentioned previously, specific and clear language criteria do not exist either for admission to the MT at the PYP or to the MHCCs. This lack of criteria contributes to the gaps identified in the study and affects the quality of learning as students join with noticeable variations in their English proficiency, which make it hard to have fair assessment at the end of the PYP. With this lack of appropriate admission criteria, learning objectives became unclear to stakeholders (including the students) who are joining the PYP or when enrolled onto the MHCCs.

As the current study has introduced, specific and clear English requirements, which can be used as admission criteria, should be based on standards identified following an empirical NA study. These standards can be identified using an

international language proficiency framework (such as the CEFR in this study). These standards should also be contextualized to suit the context where they are used (Harsch, 2018). This contextualization of standards will help, as intended by the CEFR proponents (CoE, 2001), to unify communications between stakeholders and institutions (both nationally and internationally). The suggested standards at the end of the study are based on the findings from both the qualitative and quantitative analyses. The CEFR-context-related standards identified in this study (Table 6.19) can be used as a starting point to prepare the students for the writing demands/requirements of the first-year in college. They can be used to modify the curriculum outcomes, to consider assessment processes, to design standardised tests, and as criteria for admission.

6.6.4.2. Assessment policy

As pointed out by Kellaghan and Greaney (2001), appropriate assessment leads to relevant and effective instruction, which will in turn raise the quality of learning. Macdonald argues that “the design of assessment is critical in determining the direction of student efforts” (2004, p.218). A clear and detailed policy and procedure for the assessment of writing (both in the PYP and in the MHCCs) should be provided to all stakeholders before starting the course. The CEFR-based-context-related standards (see Table 6.19) can be recognized and applied by the different stakeholders both in the PYP and the MHCCs. Student assessment should also be benchmarked against these standards and levels (i.e. B2 as identified in this study) to ensure that students have minimally reached the required levels that allow them to cope with the writing demands in their colleges.

When assessment is benchmarked against required standards, assessment and teaching materials become relevant. Murphy (2003) reminds us that problem is exacerbated when challenging material is not presented as relevant to students’

needs. Students will put in the minimum effort to complete a task, and this often prevents them from achieving the main objectives of the course. Brown and Hirschfeld (2008) point out that students “are quite sensitive to assessment which they perceive to be unfair, bad or irrelevant to them, regardless of their place in the system (*i.e.*, elementary, secondary and tertiary)” (p.5). Assessment has a strong ‘washback’ effect on both teaching and learning. This washback can be positive, negative or both (Alderson, Brunfaut and Harding, 2017) and has the power to ‘change [learners and teachers’] behaviours’ (Baird et al., 2017, p.320 cited from Alderson, Brunfaut and Harding, 2017, p. 381).

Writing should be part of the PYP placement test. This will provide PYP policy-makers, tutors and coordinators with a clearer vision of the students’ actual proficiency in writing at the beginning of the programme and this can be used to measure students’ progress and achievement at the end of the course.

It is also important to have clear curriculum objectives and specific standards to be used as benchmarks and to be constructively aligned with the PYP assessment, as indicated in Biggs & Tang (2011). In this current study, as suggested by the current data, B2 is the minimum CEFR levels (see Table 6.19). But, given the current situation at the PYP, where students are accepted with variable levels of language proficiency, the standardised exam is problematic, even if it is designed with higher benchmarks. As suggested by Tsai and Tsou (2009), the assessment tasks should be modified “on the basis of [the] different levels of the students’ competence in English and learning situations” (p.327). Hence, I suggest the use of level-specific writing tasks for the end of year examination instead of the standardised test in the PYP, ranging from A2+ to B2. This proposal is not without its problems, however. For example, students with higher level of proficiency might deliberately place themselves into lower levels to raise their overall grade average (Alobaid, 2016). Therefore, it might be better to introduce the concept of criterion-referenced assessment (based on agreed criteria *i.e.* the CEFR) where students need to meet instead of the use of grades (norm-referenced test) at least for classroom assessment (Fulcher & Davidson, 2007).

6.6.4.3. Institutional codes of ethics and misconduct policies

Another important area that needs to be considered (both at the PYP and in the MHCCs) is the students' tendency to plagiarise. The students in the present study, like many other students in the Arab world, typically do not have to abide by research ethics procedures that are strictly adhered to by institutions (Muthanna, 2016). Resnik, Rasmussen and Kissling (2015) studied research misconduct in 40 countries and found that 18 countries, including Saudi Arabia, do not have clear codes of ethics and national misconduct policies in their education systems. In the university where this study was conducted, while detailed ethics-related documents exist, they tend not to be strictly enforced, especially for students. University policy-makers need to improve practices in this regard and train both students and staff in how to avoid plagiarism and the consequences of a failure to do so.

It is also important for policy-makers and teaching staff to understand the specific reasons why their students plagiarise if they are to deal with the problem effectively. Many Asian students, including Saudi students, consider memorization and imitation as a way to respect knowledge (Shukri, 2014). This "immature writing" (Bereiter & Scardamalia, 1987) is considered the norm for many Arab and Saudi learners, and only through proper instruction and training (Hyland, 2003), along with the implementation of plagiarism codes of ethics, will students eventually improve their ethical practices in writing and avoid plagiarism (Muthanna, 2016; Shukri, 2014).

6.6.5. 'Handle the students with kid gloves' a concluding remark

It is of vital importance to understand student behaviour, particularly why they care more about marks than about their progress in writing. It would be beyond the scope of this study, however, to talk in detail about the students' attitudes, but some key factors need to be considered.

In addition to the students' weak writing proficiency, the PYP is putting a great deal of pressure on students to maintain a higher GPA than their peers, who are also applying to the same MHCCs. Students will do almost anything to secure higher grades, making their progress in the subjects least important. These students have not been exposed to basic and different writing skills since high school and writing academically, and more specifically on medical-related topics, is a new experience for them.

Students' difficulties in writing should be addressed from the moment they commence the PYP and more time needs to be allocated to writing practice throughout, starting from basic skills and moving to more complex skills. The students' progress in writing should be observed and proper feedback given for them to improve. Self-assessment of their writing can also be introduced from the start of the PYP to promote greater awareness among themselves of their progress throughout the year. Criterion-referenced assessment could be one of the possibilities to overcome these problems.

Furthermore, as Cox (2014) suggested, there is a need to "put the faculty member in the L2 writer's shoes" (p.305) in order to understand why students are having these difficulties in their writing. Systemic changes are needed to solve writing problems, not only by teaching different writing skills and vocabulary, but also assuaging students' insecurities about their grades without affecting their future success. When writing is poorly incorporated into the curriculum, given little weight in assessment, and simplified to the point that neither student nor teacher is paying attention, improvement seems unlikely. In other words, "*the students need to be handled with kid gloves*" [COO14_PH1_FG7_1502].

Chapter 7

Conclusion

7.1. Summary of the thesis

The objectives of this mixed-methods longitudinal study were threefold. (1) It explored the achieved CEFR levels and writing skills of Preparatory Year Programme (PYP) students in the medical track. (2) It explored the CEFR levels and writing skills required of these same students during their first year at the university Medical and Health Care Colleges (MHCCs). (3) The analysis of data collected over a period of two years permitted the identification of misalignments between the writing proficiency required at MHCCs with that actually achieved at the PYP, as well as potential gaps in the PYP writing course curriculum.

This study was presented across seven chapters. The first chapter set the stage by articulating research aims and questions, how these related to observed problems, and the rationale behind the current study. The second chapter offered a review of the existing body of relevant scholarly literature that guided and shaped the study's design as a gap analysis mixed-methods approach with longitudinal aspects. Chapter 3 more explicitly explained and justified the study's methodology in terms of its design, the paradigm, ethical considerations and the methods used for data collection and analysis. It also included a description of the pilot stages for each phase of the study. Chapters 4 through 6 discussed the qualitative and quantitative results of the study. In Chapter 4, the findings from Phase I (*present-situation analysis*) were presented. This chapter focused on an exploration of the students' writing proficiency, including their CEFR levels at the end of the PYP. Chapter 5 focused on the findings from Phase II (*the target-situation analysis*), which explored the writing proficiency required of first-year students in the targeted colleges. Chapter 6 offered a comparative analysis of the two phases in order to identify the gaps. This seventh and final chapter provided

a short summary and an overview of the thesis as well as a discussion of the main contributions of the study to the current body of knowledge.

7.2. Summary of results

This study explored the writing proficiency of a group of students who completed both the PYP and attended the first year of several MHCCs in one of the universities in Saudi Arabia. It employed a gap NA, with some longitudinal aspects, in a mixed-methods study that helped to identify gaps between the *present* and *target* situations. A wide array of stakeholders participated in the two phases of the study, namely students, tutors, raters, coordinators and academic staff. To identify misalignments in writing skills between the PYP and MHCCs, the same student cohort participated in both phases of the study. The findings from this study can assist PYP policy-makers and curriculum designers to bridge the gaps between the two curricula and implement changes to better prepare students for their first academic year at a MHCC.

To answer the main overarching question, this study also explored seven sub-questions. RQ1 identified the students' actual CEFR levels by the end of the PYP. This was done through self- and tutor assessments using the CEFR scales, as well as ratings of a sample of the same students' written texts by seven independent raters. My analysis of the given ratings reveals that, in contrast to the ceiling effect found in the students' PYP final exam results, in actuality, the students' proficiency levels vary significantly across the three PYP levels (elementary, intermediate and advanced).

The CEFR levels of the MT students at the end of the PYP (**RQ1**) differed between PYP levels and were identified as:

- Elementary-level students: Students' self-assessment: B1; tutors: A2+ and independent raters: A2+.
- Intermediate-level students: Self-assessment: B1, tutors: B1+ and raters: A2+.

- Advanced-level students: Self-assessment: B2, tutors: B2 and raters: B1.

In phase II of the study, I identified the CEFR level required of students in their first-year of college (**RQ3**) i.e. B2. Comparing these results (CEFR levels achieved in Phase I (**RQ1**) with the ones required in Phase II (**RQ3**), misalignments were recognized (**RQ5**). According to my analysis, misalignments appeared between the actual levels achieved in Phase I and the levels required by students at the MHCCs in Phase II. These misalignments were especially critical in the areas of *Vocabulary range and control*, and *Reports and Essays* scales. This finding brings to light the fact that upon the completion of the PYP, many elementary and intermediate students fail to meet the minimum requirements of MHCCs (**RQ5**), and this is the most obvious source of dissatisfaction with students' writing proficiency in the MHCCs.

Yet, it is worth noting again that the CEFR scales only reflect general proficiency (CoE, 2001) rather than the specific proficiency required in medical colleges. As such, the issues faced in Phases I and II could, potentially, be deeper, considering the specific academic language achieved and required. Therefore, in this study, to answer RQs 2 and 4, a panoramic overview of the writing skills and proficiency achieved and required were obtained from the in-depth exploration of the students' writing proficiency in the two Phases. This in-depth analysis of qualitative data not only explained the findings from the quantitative results, but also helped to understand the gaps and the potential causes of these gaps.

In order to answer **RQ2**, qualitative data were gathered. Based on the data, it was found that the improvement of writing skills, in general, was not prioritized by either the students or staff at the PYP. Little effort was dedicated to teaching, practicing and assessing writing. It can thus be stated that the development of writing skills was not institutionally recognized in the PYP. For example, although the PYP curriculum has a dedicated textbook for teaching English for academic and medical purposes, little time was allotted to using the textbook for the teaching of specialised English writing skills (**RQ2**). This is the case despite the fact that proficiency with specific medical vocabulary, terminology and

genres is required of all medical students (RQ4). Regrettably, this lack of prioritization in developing students' writing skills was also true in the MHCCs; yet, among individual teaching staff members of the MHCCs, there was emphasis on its importance to student success (RQ4).

This lack of development of writing skills in the PYP is further compounded by the single summative assessment of students' writing skills at the end of the year. This exam is identical across all three PYP levels, deals with very general and simple topics, and does not foster the development of any of the specific skills required to enable the students to cope at their future colleges. Furthermore, the homogeneity of student results on this PYP final exam does serious harm by misrepresenting the skills of an incoming cohort to both MHCC staff and the students themselves. As such, unrealistic expectations tend to hinder improvement as much as they do success.

Indeed, data collected in Phases I and II (RQ6) demonstrated that the limited opportunities to practice writing, limited exposure to medical terminology, and the limitations of the standardised summative exam were the main gaps flagged by student and teaching staff participants.

As a result of these factors, students finished the PYP intensive programme with writing proficiency levels lower than the minimal requirements for the first-year MHCCs, especially among students coming from the elementary and intermediate levels (RQ5). The study's findings demonstrate that student success in the MHCCs requires not only the ability to write sentences using a clear structure, appropriate and specific terminology related to each field, and correct spelling, but also the ability to summarize, paraphrase, or describe a situation before expressing a critical opinion on it (RQ4). Students did not have the chance to practice these skills while at the PYP. Furthermore, students' poor writing skills in examinations have affected how the MHCCs approach the development of writing skills as well. Written assignments, projects, and research reports are increasingly disregarded by faculty members despite the importance of such practice to students' improvement (RQ6).

An analysis of the misalignments revealed how the many factors contributing to these identified gaps reinforce each other. Indeed, all stakeholders have a role to play in improving the success rates of MHCC students. Of all the factors identified, three identifiable causes stand out and can be used as a starting point for designing solutions and answering RQ7. These are:

- (1) The lack of a proper needs analysis of the actual requirements of the target-situation as the foundation for the PYP intensive programmes;
- (2) No clear admissions procedure to either the MT at the PYP or the MHCCs;
- (3) Students' demands for easier assessments to inflate test results.

7.3. The study's contribution

This exploratory longitudinal mixed-methods gap analysis study of students' writing proficiency has generated different contributions summarised in three key points. The first is the study contribution to the Saudi context, specifically university PYPs and MHCCs. This includes producing CEFR-based context-related standards which can guide and improve the PYP curriculum design, improve learning and teaching materials and the assessment of writing. The second contribution is to add to our knowledge and understanding of the use of the CEFR scales for needs analysis studies. The third contribution refers to practical and methodological aspects of the study.

7.3.1. Contribution to the Saudi context

The first contribution of this study is the addition it makes to the currently sparse literature regarding needs analysis studies within the context of Saudi universities. Adopting as it does a mixed-methods, gap analysis approach to needs analysis with some longitudinal aspects, in tandem with its application of CEFR scales is, to the best of my knowledge, the first of its kind to be conducted

within the context of Saudi universities. This approach allowed for the exploration and identification of gaps in the PYP curriculum writing course at tertiary level and the potential causes of these gaps.

This study, which utilizes the CEFR as the international framework to identify students' proficiency levels at phase I (*present-situation analysis*) and the required levels at phase II (*target-situation analysis*), addressed a gap in Saudi literature. To my knowledge, the students' proficiency levels were not previously empirically defined nor aligned to an international proficiency framework such as the CEFR. More specifically, no research, to my knowledge, has identified students' writing proficiency levels as part of a needs analysis study for tertiary level students, specifically in relation to medical students within the Saudi context. As such, an important contribution of this study is the identification of students' writing proficiency levels in relation to the CEFR scales (both achieved and required). This is especially important because, as mentioned in Chapter 1, the PYP has already introduced the CEFR into its curriculum. However this was done intuitively and without any empirical basis. These identified CEFR levels can then be used to re-define the PYP curriculum framework by providing an overview of the students' proficiency levels at the end of the programme in each of the three PYP levels and the levels the student need to acquire at the end of the PYP.

This step was important not only because it maps the students' levels on a proficiency scale, but it also flagged up the significant variations of abilities between PYP levels (elementary, intermediate and advanced) at the end of the course, which had previously been masked by the ceiling effect identified from these students' scores during their summative assessment, the final exam at the PYP. From this finding, it is recommended that the PYP modify their assessment practices in order to provide a more accurate picture of the students' proficiency levels at the end of the programme.

One important outcome from the current needs analysis is the CEFR-based-context-related standards where the minimum required proficiency level for

first-year MHCC students was identified along with more supplementary standards. These standards were identified based on the analysis of the qualitative and quantitative data derived from students' and staff perspectives.

A number of important insights can be yielded from these CEFR-based-context-related standards. Firstly, according to previous needs analysis studies, the only (overly vague) specification of language requirement for students to enter medical colleges was that they achieved an advanced level in English, without providing a clear definition of what constituted high proficiency. However, this study gives a clearer, more precise definition of the actual requirements for first year students of MHCCs and the proficiency they should attain with reference to the CEFR proficiency scales. This was done by locating the minimum level of proficiency needed, and what students could and should (minimally) be able to do to successfully undergo the academic rigours faced by first-year students in the MHCCs.

Secondly, these standards can be used as 'a point of departure', as expressed by North (2014), to modify the PYP curriculum to better meet the needs of the students and the recipient MHCCs. These standards can be used to set curriculum objectives and outcomes that students need to attain at the end of the programme. These standards can also be used to guide the development and selection of teaching materials to allow for specific skills to be developed to achieve these contextual standards. Moreover, these standards can be used to modify the types of assessment conducted at the PYP, as well as be utilized as a benchmark for the exams to ensure that students are meeting these standards prior to their acceptance into the MHCCs.

Thirdly, these CEFR-based context-related standards can also be used to negotiate the possibilities of setting language criteria for university admissions, particularly in colleges where English is the main medium of instruction in Saudi Arabia (e.g. MHCCs). This is important especially to those who are taught using international curricula in 'English as the medium of instruction' contexts (Alblowi, 2016).

Additionally, these standards can be used as a point of reference for discussion (Harsch, 2018) between policy makers, English teachers and subject academic staff. Thus, CEFR can be utilized as a *common language* between the stakeholders at the PYP and the MHCCs through which to discuss students' proficiency levels and agree on certain levels and skills that students need to acquire prior to their first year in the colleges. To this end, I call for CEFR training to be conducted for the various stakeholders within the PYP and the MHCCs to foster a clear understanding of the CEFR and what each proficiency level entails. This will go a long way towards influencing the development of the curriculum, the teaching and the assessment methods at the PYP, and help to bridge the current gap between learners' proficiency levels at the point of exit and the expectations at point of entry to the MHCCs.

This study's in-depth qualitative analysis, which was guided by the CEFR categories in phases I and II, also yielded valuable insights regarding how writing is perceived and treated by the various participants in both the PYP and the MHCCs. This enabled a comparison of the data collected between the two phases, resulting in the identification of gaps. The analyses also pinpointed the key elements (identified in the form of a vicious circle - see Chapter 6) that contribute to the gaps between what is achieved and what is required between the two phases. So far as I know, these insights have never previously been documented within the Saudi context.

Among the new insights gained through the analyses of the qualitative data in phase I, was the student participants' own perception of their writing proficiency at the end of the PYP. The qualitative data also provided information as to how the development of writing proficiency was approached within the PYP. More significantly, we were also provided with information regarding the status and importance of writing as perceived by the different participants.

A similar in-depth analysis of the data in phase II revealed the same learners' perception of their writing skills a year later, as well as the perception of writing among various participants involved in the first year in medical colleges. The

comparison of the qualitative data collected in these two phases consequently helped me to identify the gaps between what is achieved and what is required. The causes of these gaps were attributed to various factors including:

- (1) lack of proper needs analysis,*
- (2) lack of rigorous assessment policy,*
- (3) lack of clear admission criteria,*
- (4) students' varying proficiency in writing,*
- (5) writing is not well-considered in the curriculum,*
- (6) students' need for high GPA,*
- (7) students' fear of losing marks,*
- (8) students' complaints demanding easier assessment of writing,*
- (9) the tendency to lower the requirement,*
- (10) easier assessment of writing, and*
- (11) writing given less weight in teaching and assessment.*

These are significant because these elements are found to affect students' proficiency levels and further contribute to the gaps that have already been identified.

Further compounding the problem is the fact that the vicious circle does not simply affect the students' proficiency levels in the PYP. Rather, it is found to continuously hinder these students' progress towards higher proficiency levels at the respective MHCCs. As such, I strongly suggest that the identified factors, along with the aspects in the vicious circle, are addressed seriously in order to bridge the gaps that have been identified, while improving the students' writing proficiency (as indicated in the implications, see Chapter 6 section 6.6).

Another contribution that can be garnered from the in-depth analysis is our understanding of how various skills in writing are perceived (*highly, partially and marginally important*) by participants, based on what is required at each stage of the students' education in their MHCCs. This finding gives rise to an important implication in terms of the development of a writing skills curriculum at the PYP. The data from this study have yielded a list of *highly required* general proficiency writing skills for first year MHCC students (mentioned in the CEFR-based context-related standards). This list can serve as a guideline to develop a PYP writing curriculum that allows students to practice and develop these skills as a priority.

The data analysed also revealed areas of writing proficiency that were identified as *partially* and *marginally* required. This finding in itself is significant and carries with it a serious implication for what skills are partially or marginally required of first-year students in college. As discussed in Chapter 5, the reason that these proficiencies were identified as only partially and marginally required could be attributed to the effect of the vicious circle identified from the data. It also could be because the students only needed these proficiencies in later stages while they were studying in these medical colleges, as part of their academic literacies and professional development (Murray, 2016b). Therefore, another important implication for policy makers is to reconsider having additional writing support post-PYP, through the development of writing centres at the MHCCs. This is opposed to the current practice of solely relying on the PYP to fully prepare and meet the needs of the students in every aspect of developing their writing skills and academic literacies that are required by specific colleges, within the one year PY programme.

7.3.2. Contribution to the field (the CEFR)

CEFR-based needs analysis

This study comes as a response to a call by both North (2014) and Harsch (2018) regarding the utilization of the CEFR to carry out needs analysis studies in order to enrich the sparse literature in this area. Many educational (and other) institutions in Europe and around the world set a CEFR-based language requirement, usually with little prior effort expended on any needs analysis (Harsch, 2018). Harsch (2018) suggested carrying out studies where researchers analysed the local needs and then formulated context-specific and context-relevant CEFR-based requirements. North (2014) has called for conducting needs analyses in order to select specific objectives “related to real world tasks the learners are going to have to perform in the language” (p.111). The current study, following an extensive needs analysis, contributed to the limited literature where the CEFR is used as the basis for needs analysis, to identify the CEFR level achieved and required, and any misalignments between them. It also contributed to the literature of contextualizing the CEFR for specific contexts, and adding *supplementary* contextualised descriptors that have emerged from the data analysed.

Required CEFR levels

The CEFR has been widely used to set proficiency entrance requirements in higher education (Deygers, Zeidler, Vilcu and Carlsen, 2018). According to Deygers, Zeidler, Vilcu and Carlsen (2018), the current common practice in Europe is to set entrance requirements without conducting an extensive needs analysis, but rather to just copy the practice of competitors. B2 is the most commonly chosen minimum required level (Papageorgiou *et al.*, 2015) for university entry, usually without any empirical justification (Alderson, 2007).

Unlike this common practice, my study contributed to the body of literature by identifying the minimally-required level for entrance to MHCCs in Saudi Arabia, based on an empirical needs analysis. While B2 is commonly stipulated as the minimum required level for first-year students in many international universities around the world, now, through this study, this level can be empirically confirmed as the most appropriate proficiency level for Saudi students at Medical Colleges, based on empirical data garnered from different methods and perspectives. In doing so, it supports other studies which empirically underpin B2 as the CEFR level minimally required for academic success (see, for example, Harsch, Ushioda and Ladroue, 2017 (in a UK context) and Carlsen, 2018 (in Norway)).

The identification of supplementary descriptors

CEFR proponents have emphasized the point that the CEFR is only to be used as a guide and reference tool, and needs to be adapted to the context where it is to be used (CoE, 2001): “The CEFR offers guidance for building contextually relevant standards [of] what knowledge and skills...[learners] have to develop to be able to act effectively [in specific contexts]” (Green, 2018, p.2). In addition to the CEFR levels identified, my study also identified *supplementary* descriptors relevant to the context and based on the outcomes of a needs analysis study. These supplementary descriptors illustrate the versatility of the CEFR framework to be customised to make it more relevant to local contexts. Identifying these supplementary descriptors was an important contribution to the study, especially since the CEFR scales focus on general proficiency and given the need for scales that are more specific and address academic skills. Although North (2014) had mentioned that there are some elements in the CEFR scales that refer to academic skills, these skills can only be used as a point of departure. McNamara *et al.* (2018) have called for the incorporation of such specific descriptors (like the ones identified in this study) relating to academic writing,

as the current CEFR descriptors fail to reflect the complexity and high cognitive demands required for academic writing skills.

7.3.3. Practical and methodological contributions

The Usability and Reliability of the CEFR beyond the European Context

The purpose of this study was not to test the reliability and validity of the CEFR scales, however, the results gained from the current study supported and pinpointed the scales' reliability and usability.

First of all, reliable results were obtained following the innovative design using the CEFR scales both for assessing the proficiency achieved in Phase I ($\alpha=0.884$ and 0.951 for student and tutor data respectively), and for students selecting the proficiency levels required in Phase II ($\alpha=0.847$). In Phase I, I formulated an innovative design of the CEFR scales used to measure language proficiency (which I called *Controlled Utilization of CEFR Descriptors*), which helped guide participants when they were deciding the levels to choose for each descriptor. The design helped participants read the descriptors in ascending order, one level after another, until they reached the level that they found above their current level of proficiency. This guidance procedure is unique and contrasts with observations from other studies, where the CEFR utilization methods are either not clearly stated, or less guided methods are used, such as when participants freely select certain statements from the provided descriptors. Such effort to control the usage of the CEFR by subjects is of great importance in the effort to establish the reliability of the results (Sebba *et al.*, 2008). My study, therefore, confirmed the usability and suitability of the CEFR as a generally adaptable and reliable tool to measure students' proficiency in the Saudi university context.

Furthermore, the data in this study also showed that participants were able to use the CEFR scales, which gave results with face validity. For example, in Phase I, students in the advanced PYP level were assigned CEFR levels (in self- and

tutors' assessment) higher than those for intermediate and elementary levels. This indicated that the scales, in general, were interpreted in a way which reflected students' PYP proficiency, and that they could be used to identify students' proficiency levels in the Saudi PYP contexts.

In this phase, I also identified students' attained CEFR levels through the comparison of three measures; self-assessment, tutors' assessment and raters' ratings of exam texts. Though there were some discrepancies in their assessment (i.e. overestimations and underestimations), the findings were similar to many other studies, even those using measurement tools other than the CEFR. This was another indication that the CEFR scales were valid (working similarly to other proficiency measures) and supported my argument for the usability of the CEFR scales in measuring language proficiency in the Saudi context. For example, the results of the self-assessments resonated with the results from other studies such as those by Blanche and Merino (1989), Ross (1998), Sahragard & Mallahi (2014) and Üstünlüoğlu *et al.* (2012), where more proficient students tended to underestimate their abilities, whereas less proficient students tended to overestimate their proficiency. This finding also reflects the Dunning-Kruger effect (Kruger and Dunning, 1999) of over and under-estimation of abilities.

A comparison of the results across student proficiency levels yielded another interesting observation that could contribute to the field of needs analysis. This was that the higher the proficiency levels of the students, the higher these students perceived the required levels to be. It would be interesting to examine further how students' proficiency levels affect their perception of the required abilities, although this aspect was beyond the scope of the current study.

Moreover, another observation that could also be significant to the field of needs analysis was that being satisfied or dissatisfied might also have affected the students' self-assessment. For example, students in the elementary level felt satisfied with their progress and this could explain the overestimation observed in their self-assessment. On the contrary, students in the intermediate level were less satisfied and as a result they tended to underestimate their levels. This could

be related to learners' internal drive to mastery, or even their understanding of the details of various elements that contribute to language mastery. However, further empirical studies are required to explore the relationship between feeling satisfied with the course and self-assessment of proficiency.

Comparability

The results obtained from this current study yielded significant insights in terms of comparing different data using the same CEFR scales.

In Phase I, I compared student, tutor and rater data to triangulate the results and obtain an overview of the students' proficiency achieved as perceived by the different participants. Although, as expected, some significant differences were identified among the different participant groups, these differences were, as argued in the *Manual* (CoE, 2009) within the acceptable range (mostly within two adjacent CEFR levels). Someone might argue that having differences among and within the different assessors (self-, tutors and raters), albeit close to each other, could still present problems especially for assessment. However, one of the aims of this study was to identify the students' CEFR levels and locate their proficiency on the CEFR levels. Therefore, variability among different groups, at this stage, was expected and accepted.

These differences can be attributed to other factors besides the CEFR itself (Harsch, 2018). They could, for example, be related to lack of training (Oscarson, 1989; Leach, 2012; Jafarpur, 1991), or the students' proficiency level (Heilenman, 1990; Oskarsson, 1984; Shimura, 2006), or the wording of the CEFR descriptors (Harsch, 2018). Nevertheless, even with all these factors, the results obtained in this study showed that the CEFR can be used and compared between different participants. This means that the CEFR can be considered suitable to identify learners' proficiency levels and levels required, and to be used in similar contexts where participants have no or limited knowledge of the CEFR, and limited training to familiarize them with its use.

Another unique feature of this study was that the results from Phases I and II were compared not only to outline misalignments in students' proficiency between the two phases, but also between different participating groups (i.e. students' and their tutors/staff assessment). By checking the assessment results between groups, we gained some understanding as to how, and to what extent, the groups agreed in identifying the students' levels. Additionally, in Phase I, the comparison was made across the three PYP levels to compare students' achieved CEFR levels across the three PYP levels (elementary, intermediate and advanced). In Phase II, I also compared the differences in perceptions between students and teaching staff of the minimum CEFR levels required for first-year students. The comparison was made across the different MHCCs and the PYP levels. At this juncture, the CEFR scales were used to compare participants across PYP levels and contexts (i.e. MHCCs).

In Phase III, the comparison of the two earlier phases, and the identification of the misalignment between the levels achieved and required, was the most significant finding of this study. To the best of my knowledge, no other study has been conducted in a similar way to identify the misalignment between levels achieved and expected, to define the gaps between the PYP and MHCCs. The misalignment and gaps identified were further supported by the results from the qualitative analysis in these two phases. For example, significant misalignments were especially found in the *Vocabulary range and control* and *Reports and essays* scales, and these were the same two areas where most participants expressed greatest difficulty.

Given the reliable results presented in this study and which indicate the usability of the CEFR scales, there is a strong case for saying that the CEFR scales, if used with proper design and clear instructions, can be used in other contexts, such as Saudi Arabia (Abdulhaleem & Harsch, 2018). It can also be used reliably by participants with no/little experience of working with the CEFR and with limited training.

7.4. Limitations and recommendations of further studies

On the whole, the study findings can be generalized to other Saudi and Arab universities, especially those with similar contexts (i.e. those using international curricula and those where English is the main medium of instruction). Future studies can aim to replicate this study to explore students' proficiency in other disciplines and universities. Whereas this study only focuses on students' needs for writing proficiency, more research is recommended focusing on other skills and disciplines.

Another significant limitation faced in this study is the lack of an objective measure (e.g. students' scores from well-developed tests or international well-recognized tests that are properly aligned to the CEFR) to obtain a general view of the students' proficiency levels. It was not possible, given the restricted logistics followed in the context of this study, to compare participants' use of this study tool with such objective measures. Nevertheless, having results from different participant group perspectives helped this study in obtaining an overview of these students' language proficiency. However, future studies are required to compare participants' identified CEFR levels with more objective measures, to assess the validity of the measure.

At the time this study was conducted, B2 was the minimum identified required level for first-year students in the MHCCs. B2 ranges "from the upper portion of IELTS band 5.0 ... to the lower portion of the 6.5" (Green, 2018, p. 9). In the PYP, sitting of IELTS is an option only available to advanced-level students. Even if these students were to score beyond band 6.5, they would not be exempted from the PYP (although those scoring 5.0 and above are exempted from the final English language assessment). At this point, there has yet to be a study conducted to examine the need for B2 as the minimum level required for academic success at the MHCCs in Saudi Arabia. Therefore, predictive validity studies are needed to investigate the relationship between students' subject achievements and their academic success in MHCCs with e.g. their IELTS band scores or other measures aligned to the B2 level.

Due to the segregation policy in Saudi Arabia, the current study also focused on collecting data only from female candidates at the PYP and the MHCCs. Therefore the results of this study cannot be generalized to include the male section in the MHCCs as well. Future studies are required to compare other data collected from the male cohorts. Self-assessment results in particular should be viewed with caution, as it has been found that gender differences can affect the results of self-assessment (Denies & Janssen, 2016).

It is important for other researchers who want to carry out similar needs analysis studies, particularly in Saudi contexts, to be cautious of the different factors contributing to the gaps, including the elements in the vicious circle, as those might affect their findings.

The *Controlled Utilization of CEFR Descriptors* that I used in the questionnaire for data collection also had its limitations. This document was paper-based, which required the printing of many copies and took time for data entry. I suggest, for future studies, that an electronic version of the questionnaire be developed to minimise printing and to ease data collection, entry and the process of analysis, and potentially enabling the use of a larger cohort of students across various locations.

From the qualitative data analysed, it was revealed that while the awareness of the importance of writing is very high among all the stakeholders, little was done to embed the development of writing proficiency at both the PYP and MHCCs, and this lack of focus was further emphasised in both the continuous and final assessment at the PYP. The qualitative data in this study provided insights into steps that could be taken at both the PYP and MHCCs to potentially break the vicious circle and close the gap between actual and expected proficiency levels between the two institutions, including incorporating the writing skills that are deemed highly required for first year students at MHCCs into the writing curriculum and assessment at the PYP. Whereas, at the MHCCs, skills that were deemed partially and marginally required could be included as part of an on-going writing development course at specific writing centres to serve the

academic literacies needs of the students. Based on this, future research is required to address the specific academic literacies required at each college, be it in a specific discipline, or type of college. This would be useful to enable each of these colleges to establish writing programmes that would address the specific skills, terminologies, genres and other writing requirements needed by their students. This would be especially important given the fact that “the CEFR descriptors underrepresent the complexity of the challenges of academic writing...[and] a new and rather different approach will be required to inform assessments used to manage the admission of students in to academic writing contexts and monitoring their progress” (McNamara *et al.*, 2018).

7.5. Concluding remarks

Since its publication in 2001, the CEFR has been used in different areas in language teaching, learning and assessment. However, the detailed and intensive way the CEFR scales were used in this study to carry out a gap analysis approach to needs analysis, to my knowledge, is unique. This intensive usage of the CEFR, targeting two places (PYP and MHCCs), and following the same student cohort in the second phase, has added to the literature, indicating the usability, practicality and suitability of the CEFR scales for various purposes. The standards in the current study were identified based on empirical data and research rather than from intuitions or following what others do (Deygers, Zeidler, Vilcu, & Carlsen, 2018)

In concluding this thesis, I have shown that the CEFR scales were an appropriate tool to carry out gap analysis in needs analysis studies. Through the design of the *Controlled Utilization of CEFR Descriptors*, it was evident that various participants were able to use the CEFR descriptors reliably with minimal experience and training in a new context, outside of Europe. The utilization of CEFR scales and descriptors also enabled me to identify the perceived CEFR levels that the students achieved at the end of the PYP, and the writing skills that were

minimally required for first year students at the MHCCs within the university. Through the comparison of data from both the present and target situations, I was able to identify the gaps between the two. The qualitative data analysis also provided insights regarding the vicious circle that contributed to the potential causes of these gaps. These causes provide important issues for consideration in terms of the development of a writing syllabus, teaching methodologies and the focus of assessment practices at the PYP, as well as potentially establishing writing centres at the MHCCs to serve the specific academic literacies needs of the students.

This study contributed insights into the applicability of the CEFR for NA, and in particular in a non-European (Saudi) context. Most importantly, this study identified the gaps between actual proficiency and expectations between the PYP and MHCCs. While doing so, it also revealed the causes of these gaps. Only by understanding the nature and causes of these gaps can improvements be made to the writing curriculum, teaching methods and assessment at the PYP and the MHCCs.

References

- Abalhassan, K. (2002). English as a foreign language instruction with CALL multimedia in Saudi Arabian private schools: A multi-case and multi-site study of CALL instructors' pedagogies and beliefs. (Doctoral dissertation, Indiana University of Pennsylvania.)
- Abbad, A. (1988). An Analysis of Communicative Competence Features in English Language Texts in Yemen Arab Republic (Doctoral Dissertation, University of Illinois at Urbana-Champaign).
- Abbasabady, M. M. (2009). The reported activities and beliefs of the students preparing for the specialised English test (SPE) (Doctoral dissertation, Lancaster University).
- Abdulghani, H. M., Al-Drees, A. A., Khalil, M. S., Ahmad, F., Ponnampereuma, G. G., & Amin, Z. (2014). What factors determine academic achievement in high achieving undergraduate medical students? A qualitative study. *Medical teacher*, 36(Suppl. 1), pp. S43-S48. doi: 10.3109/0142159X.2014.886011.
- Abdulhaleem, E. & Harsch, C. (2018). Using the CEFR Scales to Assess Students' Proficiency Levels in a Saudi-Arabian Higher Education Context. In Brandt, Buschmann-Göbels & Harsch (Eds.). *Der Gemeinsame Europäische Referenzrahmen für Sprachen und seine Adaption im Hochschulkontext. Erträge des 6. Bremer Symposions. Fremdsprachen in Lehre und Forschung Bd. 51.* Bochum: AKS.
- Abomrifa, F. (2013, June 29). Preparatory Year, Suggestions and Solutions. Al-Riyadh Newsletter. Retrieved July 15, 2017 from <http://www.alriyadh.com/847807>
- Abu-Rizaizah, S. (2010). *Evaluation of an English Language Programme for Engineers in a Saudi Arabian University: a case study* (Doctoral dissertation, University of Newcastle Upon Tyne).
- Ahmed, A. H. (2010). Students' problems with cohesion and coherence in EFL essay writing in Egypt: Different perspectives. *Literacy Information and Computer Education Journal (LICEJ)*, 1(4), pp. 211-221.
- Al Alwan, I., Al Kushi, M., Tamim, H., Magzoub, M., & Elzubeir, M. (2013). Health sciences and medical college preadmission criteria and prediction of in-course academic performance: a longitudinal cohort study. *Advances in Health Sciences Education*, 18(3), pp. 427-438.
- Al-Kahtany, A. H., Faruk, S. M. G., & Al Zumor, A. W. Q. (2015). English as the medium of instruction in Saudi higher education: necessity or hegemony?. *Journal of Language Teaching and Research*, 7(1), 49-58.

- Al Kathiri, S. N. (2014). Preparatory Year (First Year Experience). *The Saudi Journal of Higher Education*, 11, pp. 65-70. Retrieved 22 May, 2016 from <http://chers.moe.gov.sa/en/Publications/Publications/HEJ-0011-En.pdf>
- Al Makoshi, M. A. (2014). Discourse markers and code-switching: Academic medical lectures in Saudi Arabia using English as the medium of instruction (Doctoral dissertation, University of Birmingham. Order No. U640792). Available from ProQuest Dissertations & Theses A&I. (1689622311). Retrieved July 17, 2016 from <http://0-search.proquest.com.pugwash.lib.warwick.ac.uk/docview/1689622311?accountid=14888>
- Al Rukban, M. O., Munshi, F. M., Abdulghani, H. M., & Al-Hoqail, I. (2010). The ability of pre-admission criteria to predict performance in a Saudi medical school. *Saudi Medical Journal*, 31(5), pp. 560–564. Retrieved October 10, 2016 from https://www.researchgate.net/publication/44593584_The_ability_of_the_pre-admission_criteria_to_predict_performance_in_a_Saudi_medical_school
- Alaqueeli, A. S. (2014). The preparatory year: Global perspectives and local practices. *The Saudi Journal of Higher Education*, 11, pp. 45-64.
- Albishri, J. A., Aly, S. M., & Alnema, Y. (2012). Admission criteria to Saudi medical schools. Which is the best predictor for successful achievement? *Saudi Medical Journal*, 33(11), pp. 1222-1226.
- Alblowi, A. (2016). An evaluation of the effectiveness and validity of the preparatory year programme in preparing students for studying in Taibah university in Saudi Arabia (Doctoral dissertation, Dublin City University). Retrieved January 4, 2018 from http://doras.dcu.ie/21363/1/Alblowi_Thesis_sep_2016.pdf
- Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. London: Continuum.
- Alderson, J. C., Brunfaut, T., & Harding, L. (2014). Towards a theory of diagnosis in second and foreign language assessment: Insights from professional practice across diverse fields. *Applied Linguistics*, 36(2), pp. 236-260.
- Alderson, J. C., Brunfaut, T., & Harding, L. (2017). Bridging assessment and learning: a view from second and foreign language assessment. *Assessment in Education: Principles, Policy & Practice*, 24(3), pp. 379-387.
- Alderson, J. C. (2007). The CEFR and the need for more research. *Modern Language Journal*, 91(4), pp. 659–663.
- Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Cambridge: Cambridge University Press.

- Alderson, J. C., & Huhta, A. (2005). The development of a suite of computer-based diagnostic tests based on the Common European Framework. *Language Testing*, 22(3), pp. 301-320.
- Al-Eisa, E., & Smith, L. (2013). Governance in Saudi higher education. In Smith, L. and Abouammoh, A. (Eds.). *Higher Education in Saudi Arabia: Achievements, Challenges, and Opportunities* (pp.27-37). New York: Springer.
- Al-Eissa, M. A. (2008). Determining the Language Proficiency Needs Relevant for Students in The Colleges of Health Sciences (Master's thesis, University of Essex). Retrieved from The Saudi Cultural Bureau Library in London.
- Alfadly, H.O. (2003). The English Language Teaching Situation in Yemen: A Case Study. *Bulletin of Higher Education Research*, 2, pp. 8-9.
- Alfehaid, A. F. (2014). English for future healthcare professionals in Saudi Arabia: A needs analysis proposal. *Journal of Teaching English for Specific and Academic Purposes*, 2(2), pp. 275-280.
- Al-Ghamdi, H. M. (2006). The effectiveness and appropriateness of a medical English course: an eclectic approach to language programme evaluation (Doctoral dissertation, University of Essex).
- Alghamdi, O. S. (2015). Satisfaction of preparatory year students with university services. *World Journal of Education*, 5(5), pp. 117-129.
- Al-Gorashi, A. K. (1988). The English communication needs of military cadets in Saudi Arabia as perceived by junior officers in the Saudi Army and air difference (Doctoral dissertation, Indiana University).
- Alhaisoni, E. (2012). Language learning strategy use of Saudi EFL students in an intensive English learning context. *Asian Social Science*, 8(13), pp. 115-127.
- Alhaisoni, E. M., Al-Zuoud, K. M., & Gaudel, D. R. (2015). Analysis of spelling errors of Saudi beginner learners of English enrolled in an intensive English language program. *English Language Teaching*, 8(3), pp. 185-192.
- Al-Haq, F. A. A., & Ahmed, A. S. (1994). Discourse problems in argumentative writing. *World Englishes*, 13(3), pp. 307-323.
- Alharbi, H. A. (2015). Improving students' English speaking proficiency in Saudi public schools. *International Journal of Instruction*, 8(1), pp. 105-116.
- Alharby, M. (2005). ESP target situation needs analysis: The English language communicative needs as perceived by health professionals in the Riyadh area (Doctoral dissertation, University of Georgia). Retrieved August 2, 2015 from http://faculty.ksu.edu.sa/majed/Documents/alharby_majid_200505_phd.pdf, (Accessed 12 January 2012)
- Alhawsawi, S. (2013). Investigating student experiences of learning English as a foreign language in a preparatory programme in a Saudi university (Doctoral

dissertation, University of Sussex). Retrieved December 23, 2016 from http://sro.sussex.ac.uk/48752/1/Alhawsawi%2C_Sajjadllah.pdf

- Al-Hazmi, S. (2003). EFL teacher preparation programs in Saudi Arabia: Trends and challenges. *TESOL Quarterly* 37(2), pp. 341–344.
- Al-hazimi, A., Al-hyiani, A., & Roff, S. (2004). Perceptions of the educational environment of the medical school in King Abdul Aziz University, Saudi Arabia. *Medical teacher*, 26(6), 570-573.
- Alhojailan, A. (2015). Perceptions of academic writing by some Saudi graduate students studying in American universities (Doctoral dissertation, Oklahoma State University). Retrieved July 7, 2018 from https://shareok.org/bitstream/handle/11244/45334/Alhojailan_okstate_0664_D_14134.pdf?sequence=1&isAllowed=y
- Alhossaynee, H. (2006). *Obstacles to the Efficiency and Performance of Nurses in Riyadh Region*. Riyadh: MOH (Research Series of Assistance Agency for the Preparation and Development of Manpower).
- Alhuqbani, M. N. (2013). Assessing the Academic English Needs of King Fahd Security College Officers: Implications for the Development of an EAP Program. *Journal of Teaching English for Specific and Academic Purposes*, 1(2), pp. 143-167.
- Aliakbari, M., & Boghayeri, M. (2014). A needs analysis approach to ESP design in Iranian context. *Procedia-Social and Behavioral Sciences*, 98, pp. 175-181.
- Aljamhoor, A. A. (1997). The English writing process of two Saudi graduate Students before and after ESL instruction. *DAI-A*, 57(9), 3854. Retrieved 16 May 2017 from http://www.tirfonline.org/wp-content/uploads/2011/05/TIRF_SheikhFellowship_2008_Abdellatif.pdf
- Al-Jarf, R. (2008a). A call for new benchmarks at Saudi language and translation schools. *Asian EFL Journal*, 10, pp. 60–72.
- Al-Jarf, R. (2008b). The impact of English as an international language (EIL) upon Arabic in Saudi Arabia. *Asian EFL Journal*, 10(4), pp. 193-210.
- Al-Jarf, R. (2011). Teaching spelling skills with a mind-mapping software. *Asian EFL Journal Professional Teaching Articles*, 53, pp. 4-16
- Al-Jarf, R. S. (2001, March 14-16). Barriers to internet integration in ESL instruction. Paper presented at TESOL Arabia 2001 conference. Dubai, United Arab Emirates.
- Aljumah, F. H. (2012). Saudi learner perceptions and attitudes towards the use of blogs in teaching English writing course for EFL majors at Qassim University. *English Language Teaching*, 5(1), pp. 100-116.
- Al-Jurf, R. (1994). An ESP Program Model for Graduate Students at King Saud University Based on Academic and Occupational Needs. *King Saud University Journal of Education and Islamic Science*, 6, pp. 67-95.

- Al-Khairiy, M.A. (2013). Saudi English-major undergraduates' academic writing problems: A Taif university perspective. *English Language Teaching* 6(6), pp. 1-12.
- Alkubaidi, M. (2017). Investigating perceived challenges in English language writing instruction: An action research study in a Saudi university preparatory programme (Doctoral dissertation, Trinity College Dublin). Retrieved March 25, 2018 from <http://www.tara.tcd.ie/handle/2262/82012>
- Alkubaidi, M. A. (2014). The relationship between Saudi English major university students' writing performance and their learning style and strategy use. *English Language Teaching*, 7(4), pp. 83-95.
- Allen, L. K., Crossley, S. A., & McNamara, D. S. (2015, June). Predicting misalignment between teachers' and students' essay scores using natural language processing tools. In Conati C., Heffernan N., Mitrovic A., Verdejo M. (Eds.), *Artificial Intelligence in Education. AIED 2015. Lecture Notes in Computer Science*, vol 9112. (pp. 529-532). Cham, Switzerland: Springer.
- Almoallim, H., Aldahlawi, S., Alqahtani, E., Alqurashi, S., & Munshi, A. (2010). Difficulties facing first-year medical students at Umm Alqura University in Saudi Arabia. *Eastern Mediterranean Health Journal*, 16(12), pp. 1272-1277.
- Almulhim, A. (2001). An English language needs assessment of Saudi college-of-technology students with respect to a number of business sectors in Saudi Arabia (Doctoral dissertation, University of Mississippi).
- Al-Murabit, I. A. (2012). A closer look at an English language curriculum of a community college in Saudi Arabia. *English Language Teaching*, 5(8), pp. 226-242.
- Alnassar, S. A., & Dow, K. L. (2013). Delivering high-quality teaching and learning for university students in Saudi Arabia. In Smith L., Abouammoh A. (Eds.), *Higher Education in Saudi Arabia* (pp. 49-60). Dordrecht: Springer.
- Alobaid, A. (2016). Testing, assessment, and evaluation in language programs (Doctoral dissertation, The University of Arizona). Retrieved May 20, 2018 from <https://repository.arizona.edu/handle/10150/613422>
- Al-Omar, B. (2014). A study for King Saud University: Preparatory year is a positive contribution to university education. *Al-Riyadh Newsletter*. Retrieved April 4, 2017 from <http://www.alriyadh.com/900665>
- Alqahtani, M. (2011). An investigation into the language needs of Saudi students studying in British postgraduate programmes and the cultural differences impacting on them (Doctoral dissertation, University of Southampton). Retrieved January 29, 2018 from <https://eprints.soton.ac.uk/198175/1/Binder1.pdf>

- Al-Qhatani, M. F. M. (1999). Approaches to study and learning environment in medical Schools with special reference to the Gulf countries (Doctoral dissertation, University of Dundee).
- Alqurashi, F. (2016). English for medical purposes for Saudi medical and health professionals. *Advances in Language and Literary Studies*, 7(6), pp. 243-252. Retrieved October 1, 2017 from <http://www.journals.aiac.org.au/index.php/all/article/view/2936/2471>
- Al-Sadan, I.A. (2000). Educational assessment in Saudi Arabian schools. *Assessment in Education: Principles, Policy & Practice*, 7, pp. 143-155.
- AlSaif, A. (2011). Investigating vocabulary input and explaining vocabulary uptake among EFL learners in Saudi Arabia (Doctoral dissertation, Swansea University, UK).
- Al-Seghayer, K. (2005). Teaching English in the Kingdom of Saudi Arabia: Slowly but steadily changing. In G. Braine (Ed.), *Teaching English to the world: History, curriculum, and practice* (pp. 125-134). Mahwah, NJ: L. Erlbaum Associates.
- Alshehri, M. Y. (2001). Medical curriculum in Saudi medical colleges: current and future perspectives. *Annals of Saudi Medicine*, 21(5/6), pp. 320-323.
- Al-Shehri, M. Y., Campbell, S., Daud, M. Z., Mattar, E. H., Sayed, M. G., & Abu-Eshy, S. A. (2013). Development of Medical Education in Saudi Arabia. In *Higher Education in Saudi Arabia* (pp. 137-149). Dordrecht: Springer.
- Al-Shumaimeri, Y. (2011). Perceptions of Saudi preparatory year students about the intensive English program. *Annals of the Faculty of Arts, Ain Shams University*, 39, pp. 597-625. Retrieved March 15, 2015 from <http://www.m5zn.com/newuploads/2013/10/21/pdf/3c1e5a71edb7221.pdf>
- Al-Tamimi, A. S., & Shuib, M. (2010). Investigating the English language needs of petroleum engineering students at Hadhramout University of Science and Technology. *The Asian ESP Journal*, 6(1), pp. 1-30.
- Al-Wassia, R., Hamed, O., Al-Wassia, H., Alafari, R., & Jamjoom, R. (2015). Cultural challenges to implementation of formative assessment in Saudi Arabia: An exploratory study. *Medical teacher*, 37(Suppl. 1), pp. S9-S19.
- Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory into practice*, 48(1), pp. 12-19. Retrieved February 8, 2015 from <https://www.tandfonline.com/doi/pdf/10.1080/00405840802577544?needAccess=true>
- Ariail, J., Thomas, S., Smith, T., Kerr, L., Richards-Slaughter, S., & Shaw, D. (2013). The value of a writing center at a Medical University. *Teaching and Learning in Medicine*, 25(2), pp. 129-133. doi: 10.1080/10401334.2013.770739

- Ashton, K. (2014). Using self-assessment to compare learners' reading proficiency in a multilingual assessment framework. *System*, 42, pp. 105-119.
- Atai, M., & Shoja, L. (2011). A triangulated study of academic language needs of Iranian students of computer engineering: Are the courses on track?. *RELC Journal*, 42(3), pp. 305-323. doi:10.1177/0033688211419392.
- Athanasiou, A., Constantinou, E. K., Neophytou, M., Nicolaou, A., Sophocleous, S. P., & Yerou, C. (2016). Aligning ESP courses with the Common European Framework of Reference for Languages. *Language Learning in Higher Education*, 6(2), pp. 297-316.
- Babaii, E., Taghaddomi, S., & Pashmforoosh, R. (2016). Speaking self-assessment: Mismatches between learners' and teachers' criteria. *Language Testing*, 33(3), pp. 411-437.
- Bachman, L. F., & Palmer, A. S. (1989). The construct validation of self-ratings of communicative language ability. *Language Testing*, 6(1), pp. 14-29.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bailey, K. M. (1998). *Learning about language assessment*. Pacific Grove, CA: Heinle & Heinle.
- Ballard, B., & Clanchy, J. (1991). *Teaching students from overseas: A brief guide for lecturers and supervisors*. Melbourne: Longman Cheshire.
- Bazerman, C. (1988). *Shaping written knowledge: the genre and activity of the experimental article in science*. Madison, WI: University of Wisconsin Press.
- Bawazeer, K. (2015). A system for teaching English in Saudi Arabia. In A. K. Hamdan (Ed.), *Teaching and Learning in Saudi Arabia* (pp. 31-48). Rotterdam: SensePublishers.
- Bechger, T., Kuijper, H., & Maris, G. (2009). Standard setting in relation to the common European framework of reference for languages: The case of the state examination of Dutch as a second language. *Language Assessment Quarterly*, 6(2), pp. 126-150. doi: 10.1080/15434300802457521
- Bee, F., & Bee, R. (2003). *Learning needs analysis and evaluation*. London: Chartered Institute of Personnel and Development.
- Belcher, D. (2009). What ESP is and can be: An introduction. In D. Belcher (Ed.), *English for Specific Purposes in Theory and Practice* (pp. 1-20). Michigan: University of Michigan Press.
- Benesch, S. (2001). *Critical English for academic purposes: Theory, politics, and practice*. Mahwah, NJ: Erlbaum
- Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Lawrence Erlbaum. Carter.

- Bérešová, J. (2011). The impact of the Common European Framework of Reference on teaching and testing in Central and Eastern European context. *Synergies Europe*, 6, pp. 177-190.
- Berkenkotter, C., & Huckin, T. N. (1995). *Genre knowledge in disciplinary communication: cognition / culture / power*. Hillsdale, N.J.: L. Erlbaum Associates.
- Berwick, R. (1989). Needs assessment in language programming: from theory to practice. In R. K. Johnson (Ed.), *The Second Language Curriculum* (pp. 48-62). Cambridge: Cambridge University Press.
- Biggs, J., & Tang, C. (2011). *Teaching for Quality Learning at University* (4th ed.). Maidenhead, UK: Society for Research into Higher Education & Open University Press.
- Blanche, P., & Merino, B. J. (1989). Self-assessment of foreign-language skills: Implications for teachers and researchers. *Language Learning*, 39(3), pp. 313-338.
- Bland, J. M., & Altman, D. G. (1997). Statistics notes: Cronbach's alpha. *BMJ*, 314, 572. doi: 10.1136/bmj.314.7080.572
- Blatchford, P. (1997). Students' self assessment of academic attainment: Accuracy and stability from 7 to 16 years and influence of domain and social comparison group. *Educational Psychology*, 17(3), pp. 345-359.
- Blue, G. M. (1988). Self-assessment: The limits of learner independence. Individualization and autonomy in language learning. *ELT Documents*, 131, pp. 100-118.
- Blue, G. M. (1994). Self-assessment of foreign language skills: Does it work?. *Centre for Language in Education (CLE) Working Papers*, 3, pp. 18-35. Retrieved August 3, 2017 from <https://files.eric.ed.gov/fulltext/ED396567.pdf>
- Bosher, S., & Smalkoski, K. (2002). From needs analysis to curriculum development: Designing a course in health-care communication for immigrant students in the USA. *English for Specific Purposes*, 21(1), pp. 59-79.
- Boud, D. (1991). *Implementing student self-assessment*. Campbelltown, N. S. W.: Higher Education Research and Development Society of Australasia (HERDSA).
- Boud, D., & Falchikov, N. (1989). Quantitative Studies of Student Self-assessment in Higher Education: A Critical Analysis of Findings. *Higher Education* 18(5), pp. 529-549.
- Branden, K. (2006). *Task based language education: From theory to practice*. Cambridge: Cambridge University Press.
- Brantmeier, C., Vanderplank, R., & Strube, M. (2012). What about me?: Individual self-assessment by skill and level of language instruction. *System*, 40(1), pp. 144-160.

- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2), pp. 77-101. doi: 10.1191/1478088706qp063oa.
- Brindley, G. (1989). The Role of Needs Analysis in Adult ESL Program Design. In R. K. Johnson (Ed.), *The Second Language Curriculum* (pp. 63-78). Cambridge: Cambridge University Press.
- Broek, S. & van den Ende, I. (2013). The implementation of the Common European Framework of References for languages in Europe education systems [Policy document]. Retrieved from [http://www.europarl.europa.eu/RegData/etudes/etudes/join/2013/495871/I_POL-CULT_ET\(2013\)495871\(SUM01\)_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/etudes/join/2013/495871/I_POL-CULT_ET(2013)495871(SUM01)_EN.pdf)
- Brown, G. T., Andrade, H. L., & Chen, F. (2015). Accuracy in student self-assessment: directions and cautions for research. *Assessment in Education: Principles, Policy & Practice*, 22(4), pp. 444-457.
- .
- Brown, G. & Harris, L. (2013). Student self-assessment. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 367-393). Thousand Oaks, CA: SAGE Publications, Inc. doi: 10.4135/9781452218649.n21
- Brown, G. T., & Hirschfeld, G. H. (2008). Students' conceptions of assessment: Links to outcomes. *Assessment in Education: Principles, Policy & Practice*, 15(1), pp. 3-17.
- Brown, J. D. (1995). *The Elements of Language Curriculum: A Systematic Approach to Program Development*. New York: Heinle & Heinle Publishers.
- Brown, J. D. (2016). *Introducing needs analysis and English for specific purposes*. London: Taylor & Francis Ltd.
- Brunfaut, T. (2014). Language for Specific Purposes: Current and future issues. *Language Assessment Quarterly*, 11(2), 216-225.
- Bryman, A. (2006). Paradigm peace and the implications for quality. *International Journal of Social Research Methodology*, 9(2), pp. 111-126.
- Byram, M., & Parmenter, L. (Eds.). (2012). *The Common European Framework of Reference: The globalisation of language education policy*. Bristol: Multilingual matters.
- Byrnes, H. (Ed.). (2007). Perspectives. *Modern Language Journal*, 91(4), 641-685.
- Callies, M., Zaytseva, E., & Present-Thomas, R. L. (2013). Writing assessment in higher education: Making the framework work. *Dutch Journal of Applied Linguistics*, 2(1), pp. 1-15.
- Cambridge ESOL. (2011). *Using the CEFR: Principles of Good Practice*. Cambridge: University of Cambridge ESOL Examinations. Retrieved June 13 2018 from

<http://www.cambridgeenglish.org/images/126011-using-cefr-principles-of-good-practice.pdf>

- Campillo, M. (2006). Acquisition and transfer of a writing revision strategy: A self-regulatory analysis. *Dissertation Abstracts International. A, The Humanities and Social Sciences*, 67 (3).
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), pp. 1-47.
- Carlsen, C. (2012). Proficiency Level--A Fuzzy Variable in Computer Learner Corpora. *Applied Linguistics*, 33(2), pp. 161-183.
- Carlsen, C. H. (2018). The Adequacy of the B2 Level as University Entrance Requirement. *Language Assessment Quarterly*, 15(1), pp. 75-89.
- Chalhoub-Deville, M., & Deville, C. (1999). Computer adaptive testing in second language contexts. *Annual Review of Applied Linguistics* 19, pp. 273-299. doi:10.1017/S0267190599190147 Retrieved April 8, 2017 from <https://www.scribd.com/document/97882146/Computer-Adaptive-Testing-in-Second-Language-Contexts>
- Chang, Y. Y. (2010) English-medium instruction for subject courses in tertiary education: Reactions from Taiwanese undergraduate students. *Taiwan International ESP Journal*, 2(1), pp. 55-84.
- Charpy, J. P., & Carnet, D. (2014). The European sTANDEM project for certification in medical English: standards, acceptability and transgression(s). ILCEA. *Revue de l'Institut des langues et cultures d'Europe, Amérique, Afrique, Asie et Australie*, 19. Retrieved October 16, 2016 from <https://journals.openedition.org/ilcea/2475?lang=en>
- Chen, Y.-M. (2008). Learning to self-assess oral performance in English: A longitudinal case study. *Language Teaching Research*, 12(2), pp. 235-262.
- Chostelidou, D. (2011). Needs-based course design: the impact of general English knowledge on the effectiveness of an ESP teaching intervention. *Procedia-Social and Behavioral Sciences*, 15, pp. 403-409. Retrieved May 4, 2017 from https://ac.els-cdn.com/S1877042811002916/1-s2.0-S1877042811002916-main.pdf?_tid=d90eb307-2a2f-4aec-8544-cac5528c4a2&acdnat=1542490272_5358ee1689efdf76e06c5ff7103a26f0
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological assessment*, 6(4), 284.
- Coffey, B. (1984). State of the art article: ESP - English for specific purposes. *Language Teaching*. 17(1), pp. 2-16.

- Coffin, C., Curry, M. J., Goodman, S., Hewings, A., Lillis, T. M., & Swann, J. (2003). *Teaching academic writing: A toolkit for higher education*. London, UK: Routledge.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), pp. 37–46.
- Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin* 70(4), pp. 213–220. doi: 10.1037/h0026256. PMID 19673146.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Academic Press.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Coste, D. (2007, February 6-8). Contextualising uses of the Common European Framework of Reference for Languages. Paper presented at Council of Europe Policy Forum on use of the CEFR, Strasbourg. Retrieved July 18, 2018, from <https://rm.coe.int/09000016805ab765>.
- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press. Retrieved September 12, 2016 from <https://rm.coe.int/1680459f97>
- Council of Europe. (2008). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press (Arabic version).
- Council of Europe. (2009). *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment* (CEFR). Strasbourg: Council of Europe. Retrieved January 11, 2016 from <https://rm.coe.int/1680667a2d>
- Cox, M. (2014). In response to today's "felt need": WAC, faculty development, and second language writers. In T. M. Zawacki & M. Cox (Eds.), *WAC and second language writers: Research towards linguistically and culturally inclusive programs and practices* (pp. 299-326). Fort Collins, Colorado: The WAC Clearinghouse and Parlor Press. Retrieved August 1, 2018 from <https://wac.colostate.edu/docs/books/l2/chapter12.pdf>
- Cramer, D., & Howitt, D. (2004). Tamhane's T2 multiple comparison test in *The SAGE dictionary of statistics*. London: SAGE Publications Ltd. doi: 10.4135/9780857020123
- Creswell, J. W. (2007). *Qualitative inquiry and research design* (3rd ed.). Thousand Oaks, CA, USA: Sage Publications.
- Creswell, J., & Plano-Clark, V. (2011). *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage.

- Creswell, J.W. (2005). *Educational research: Planning, conducting and evaluating quantitative and qualitative research* (2nd ed.). Upper Saddle River, N.J.: Pearson Merrill Prentice Hall.
- Crocker, T. (1981). Scenes of endless science: ESP and education. In *The ESP teacher: Role, development and prospects*. *ELT Documents* 112, pp. 7-15. London: The British Council.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), pp. 297-334.
- Cummins, J. (1980). The construct of language proficiency in bilingual education. In J. E. Alatis (Ed.) *Current issues in bilingual education*, pp. 81-103.
- Cummins, J. (1984). Language proficiency and academic achievement revisited: A response. In C. Rivera (Ed.), *Language proficiency and academic achievement*. *Multilingual matters* 10, pp. 71-76. Retrieved March 24, 2016 from <https://files.eric.ed.gov/fulltext/ED240882.pdf>
- Cummins, J. (2009). Fundamental psycholinguistic and sociological principles underlying educational success for linguistic minority students. In T. Skutnabb-Kangas, R. Phillipson, A.K. Mohanty and M. Panda (Eds.), *Social justice through multilingual education* (pp. 19–35). Bristol: Multilingual Matters.
- Davidson, F., & Fulcher, G. (2007). The Common European Framework of Reference (CEFR) and the design of language tests: A matter of effect. *Language Teaching*, 40(03), pp. 231-241.
- Davis, L. (2015). The influence of training and experience on rater performance in scoring spoken language. *Language Testing* 33(1), pp. 117-135.
- Denies, K., & Janssen, R. (2016). Country and gender differences in the functioning of CEFR-based can-do statements as a tool for self-assessing English proficiency. *Language Assessment Quarterly*, 13(3), pp. 251-276.
- Deygers, B., Van Gorp, K., & Demeester, T. (2018). The B2 level and the dream of a common standard. *Language Assessment Quarterly*, 15(1), pp. 44-58.
- Deygers, B., Zeidler, B., Vilcu, D., & Carlsen, C. H. (2018). One framework to unite them all? Use of the CEFR in European university entrance policies. *Language Assessment Quarterly*, 15(1), pp. 3-15.
- Dickinson, L. (1982). *Self-assessment as an Aspect of Autonomy*. Edinburgh: Scottish Centre for Education Overseas.
- Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford: Oxford University Press.
- Doushaq, H. H. (1986). An investigation into stylistic errors of Arab students learning English for academic purposes. *English for Specific Purposes*, 5(1), pp. 27-39.

- Dragemark Oscarson, A. (2009). Self-assessment of writing in learning English as a foreign language. A study at the upper secondary school level. (Doctoral thesis, University of Gothenburg). Retrieved May 8, 2016 from <http://hdl.handle.net/2077/19783>
- Dudley-Evans, T. & St. John, M. J. (1998). *Developments in English for Specific Purposes, A multi-disciplinary approach*. Cambridge: Cambridge University Press.
- Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest*, 5(3), pp. 69-106.
- Ekbatani, G. (2000). Moving toward learner-directed assessment. In Ekbatani, G. and Pierson, H., (Eds.), *Learner-directed assessment in ESL* (pp. 1-11). Mahwah, NJ: Erlbaum.
- El Tantawi, M., Al Ansarir, A., Sadafiand, S., & Al Humaid, J. (2016). Evaluating the English language scientific writing skills of Saudi dental students at entry level. *Eastern Mediterranean Health Journal*, 22(2), pp. 148-153.
- Elder, C. (1993). How do subject specialists construe classroom language proficiency? *Language Testing* 10(3), pp. 235–254.
- Elder, C., & Harding, L. (2008). Language testing and English as an international language. *Australian Review of Applied Linguistics*, 31(3), pp. 34.1 -34.11 doi: 10.1075/aral.31.3.07eld. Retrieved April 23, 2017 from <https://benjamins.com/catalog/aral.31.3.07eld/fulltext>
- El-Hibir, B. I., & Al-Taha, F. M. (1992). Orthographic errors of Saudi students learning English. *Language Learning Journal*, 5(1), pp. 85-87.
- Engelhardt, M., & Pflingsthor, J. (2013). Self-assessment and placement tests—a worthwhile combination?. *Language Learning in Higher Education*, 2(1), pp. 75-89.
- Faez, F., Majhanovich, S., Taylor, S., Smith, M., & Crowley, K. (2011a). The power of "Can Do" statements: Teachers' perceptions of CEFR-informed instruction in French as a second language classrooms in Ontario. *The Canadian Journal of Applied Linguistics*, 14(2), pp. 1-19. Retrieved November 20, 2016 from <https://files.eric.ed.gov/fulltext/EJ967319.pdf>
- Faez, F., Taylor, S., Majhanovich, S., Brown, P., & Smith, M. (2011b). Teachers' reactions to CEFR's task-based approach for FSL classrooms. *Synergies Europe*, 6, pp. 109-120. Retrieved January 26, 2017 from <https://gerflint.fr/Base/Europe6/faez.pdf>
- Fageeh, A. I. (2003). Saudi college students' beliefs regarding their English writing difficulties. (Doctoral dissertation; Indiana University of Pennsylvania).

- Fahim, M., & Bijani, H. (2011). The effects of rater training on raters' severity and bias in second language writing assessment. *Iranian Journal of Language Testing*, 1(1), pp. 1-16.
- Falchikov, N. & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research*, 59, pp. 395-430.
- Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). Los Angeles: Sage Publications.
- Figueras, N. (2007). The CEFR, a lever for the improvement of language professionals in Europe. *Modern Language Journal*, 91(4), pp. 673-675.
- Fleckenstein, J., Leucht, M., & Köller, O. (2018). Teachers' judgement accuracy concerning CEFR levels of prospective university students. *Language Assessment Quarterly*, 15(1), pp. 90-101.
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5) pp. 378-382.
- Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, 33(3), pp. 613-619.
- Fox, S., & Dinur, Y. (1988). Validity of self-assessment: A field evaluation. *Personnel Psychology*, 41(3), pp. 581-592. doi: 10.1111/j.1744-6570.1988.tb00645.x
- Franklin-Landi, R. (2017). Identifying and responding to learner needs at the medical faculty: the use of audio-visual specialised fiction (FASP). In C. Sarré & S. Whyte (Eds.), *New developments in ESP teaching and learning research* (pp. 153-170). Research-publishing.net. doi: 10.14705/rpnet.2017.cssw2017.750 Retrieved March 12, 2018 from <https://files.eric.ed.gov/fulltext/ED578667.pdf>
- Fulcher, G. (2004). Deluded by artifices? The Common European Framework and harmonization. *Language Assessment Quarterly*, 1(4), pp. 253-266.
- Fulcher, G. and Davidson, F. (2007). *Language Testing and Assessment*. London & New York: Routledge.
- Fulcher, G., Davidson, F., and Kemp, J. (2011). Effective rating scale development for speaking tests: Performance Decision Trees. *Language Testing* 28(1), pp. 5-29.
- Fulcher, G. (2013). *Practical language testing*. London; New York: Routledge.
- Geertz, C. (1973). Thick description; Toward an interpretive theory of culture. In *The interpretation of cultures selected essays* (pp. 3-30). New York: Basic Books.
- Ghobain, E. A. (2014). A case study of ESP for medical workplaces in Saudi Arabia from a needs analysis perspective (Doctoral dissertation, University of Warwick). Retrieved from University of Warwick institutional repository (Accession number: <http://go.warwick.ac.uk/wrap/66885>) Retrieved February 21, 2015 from http://wrap.warwick.ac.uk/66885/1/WRAP_THESIS_Ghobain_2014.pdf

- Green, A. (2012). *Language functions revisited: Theoretical and empirical bases for language construct definition across the ability range (Vol. 2)*. Cambridge: Cambridge University Press.
- Green, R. (2013). *Statistical analyses for language testers*. Basingstoke, UK: Palgrave Macmillan.
- Green, A. (2018). Linking tests of English for academic purposes to the CEFR: The score user's perspective. *Language Assessment Quarterly*, 15(1), pp. 59-74.
- Green, A., Wang, F., Cochrane, P., Dyson, J., & Paun, C. (2012). English spreads as teaching language in universities worldwide. *University World News*, 229.
- Gubrium, J. F., & Holstein, J. A. (Eds.). (2003). *Postmodern interviewing*. Thousand Oaks, CA.: Sage.
- Gunn, C., Hearne, S., & Sibthorpe, J. (2011). Right from the start: A rationale for embedding academic literacy skills in university courses. *Journal of University Teaching and Learning Practice* 8(1), pp. 1-10.
- Hamdy, H., Telmesani, A. W., Al Wardy, N., Abdel-Khalek, N., Carruthers, G., Hassan, F., ... & El Din Ahmed, M. G. (2010). Undergraduate medical education in the Gulf Cooperation Council: a multi-countries study (Part 1). *Medical teacher*, 32(3), 219-224.
- Hamouda, A. (2011). A study of students and teachers' preferences and attitudes towards correction of classroom written errors in Saudi EFL context. *English Language Teaching*, 4(3), pp. 128-141. Retrieved April 14, 2018 from <https://files.eric.ed.gov/fulltext/EJ1080753.pdf>
- Handjani, F., & Habibzadeh, F. (2013). Medical writing in the Middle East. *Medical Writing*, 22(2), pp. 96-98.
- Harding, J. (2013). *Qualitative data analysis from start to finish*. London: Sage.
- Harris, M. (1997). Self-assessment of language learning in formal settings. *ELT Journal* 51(1): pp. 12-20. doi: 10.1093/elt/51.1.12.
- Harsch, C. (2014). General language proficiency revisited: Current and future issues. *Language Assessment Quarterly*, 11(2), pp. 152-169.
- Harsch, C. (2017). Proficiency. *ELT Journal*, 71(2), pp. 250-253.
- Harsch, C. (2018). How suitable is the CEFR for setting university entrance standards?. *Language Assessment Quarterly*, 15(1), pp. 102-108.
- Harsch, C., & Martin, G. (2012). Adapting CEF-descriptors for rating purposes: Validation by a combined rater training and scale revision approach. *Assessing Writing*, 17(4), pp. 228-250.
- Harsch, C., & Rupp, A. A. (2011). Designing and scaling level-specific writing tasks in alignment with the CEFR: A test-centered approach. *Language Assessment Quarterly*, 8(1), pp. 1-33.

- Harsch, C., Ushioda, E., & Ladroue, C. (2017). Investigating the predictive validity of TOEFL iBT® test scores and their use in informing policy in a United Kingdom university setting. *ETS Research Report Series*, 2017(1), pp. 1-80. Retrieved May 13, 2018 from <https://onlinelibrary.wiley.com/doi/epdf/10.1002/ets2.12167>
- Hasselgreen, A. (2005). Assessing the language of young learners. *Language Testing*, 22(3), pp. 337-354.
- Hawkins, J. A., & Buttery, P. (2010). Criterial features in learner corpora: Theory and illustrations. *English Profile Journal*, 1(1), pp. 1-23. Retrieved December 27, 2017 from <https://www.cambridge.org/core/services/aop-cambridge-core/content/view/S2041536210000103>
- Hedge, T. (2001). *Teaching and learning in the language classroom*. Oxford University Press.
- Heilenman, L. K. (1990). Self-assessment of second language ability: The role of response effects. *Language Testing*, 7(2), pp. 174-201.
- Hellmann, K. (2013). *What do I need to succeed: The case of Arab engineering graduate students' self-perceptions of writing* (Doctoral dissertation, University of Idaho).
- Hennink, M. M. (2014). *Focus group discussions*. New York: Oxford University Press.
- Hismanoglu, M. (2013). Does English language teacher education curriculum promote CEFR awareness of prospective EFL Teachers?. *Procedia - Social and Behavioral Sciences*, 93(3rd World Conference on Learning, Teaching and Educational Leadership), pp. 938-945. doi: 10.1016/j.sbspro.2013.09.307
- Hodges, B., Regehr, G., & Martin, D. (2001). Difficulties in recognizing one's own incompetence: novice physicians who are unskilled and unaware of it. *Academic Medicine*, 76(10), S87-S89. Retrieved July 14, 2017 from https://journals.lww.com/academicmedicine/Fulltext/2001/10001/Difficulties_in_Recognizing_One_s_Own.29.aspx
- Hoge, R., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research*, 59(3), pp. 297-313.
- Holliday, A. (1995). Assessing language needs within an institutional context: An ethnographic approach. *English for Specific Purposes*, 14(2), pp. 115-126.
- Hughes, A. (2002). *Testing for language teachers* (2nd ed.). Cambridge: Cambridge University Press.
- Huhta, A., Alanen, R., Tarnanen, M., Martin, M., & Hirvelä, T. (2014). Assessing learners' writing skills in a SLA study: Validating the rating process across tasks, scales and languages. *Language Testing*, 31(3), pp. 307-328.
- Huhta, M., Vogt, K., Johnson, E., & Tulkki, H. (2013). *Needs analysis for language course design: A holistic approach to ESP*. Cambridge: Cambridge University Press.

- Hulstijn, J. H. (2003). Incidental and intentional learning. In: C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 349-381). Oxford: Blackwell Publishing.
- Hulstijn, J. H. (2007). The shaky ground beneath the CEFR: Quantitative and qualitative dimensions of language Proficiency1. *The Modern Language Journal*, 91(4), pp. 663-667.
- Hulstijn, J. H. (2011). Language proficiency in native and nonnative speakers: An agenda for research and suggestions for second-language assessment. *Language Assessment Quarterly*, 8(3), pp. 229-249.
- Hulstijn, J. H. (2015). *Language proficiency in native and non-native speakers: Theory and research* (Language Learning and Language Teaching, Vol. 41). Amsterdam: John Benjamins Publishing Company.
- Hung, S. (2009). Promoting self-assessment strategies: An electronic portfolio approach. *Asian EFL Journal*, 11(2), pp. 129-146.
- Hutchinson, T. & Waters, A. (1987). *English for Specific Purposes: A Learning-centered Approach*. Cambridge: Cambridge University Press.
- Hyland, K. (2003). *Second language writing*. Cambridge: Cambridge University Press.
- Hyland, K. (2006). *English for academic purposes: An advanced resource book*. London: Routledge.
- Hymes, D. (1972). On communicative competence. In: J.B. Pride and J. Holmes (Eds.) *Sociolinguistics. Selected Readings* (pp. 269-293). Harmondsworth: Penguin.
- Ivankova, N. V., Creswell, J. W., & Stick, S. L. (2006). Using mixed-methods sequential explanatory design: From theory to practice. *Field methods*, 18(1), pp. 3-20.
- Labaree, R. V. (2002). The risk of 'going observationalist': negotiating the hidden dilemmas of being an insider participant observer. *Qualitative Research*, 2(1), pp. 97-122.
- Jacobs, C. (2007). Towards a critical understanding of the teaching of discipline-specific academic literacies: Making the tacit explicit. *Journal of Education*, 41, pp. 59-81. Retrieved from http://joe.ukzn.ac.za/Libraries/No_41_2007/Towards_a_critical_understanding_of_the_teaching_of_discipline-specific_academicliteracies_making_the_tacit_explicit.sflb.ashx
- Jacobs, G. M., & Farrell, T. S. (2003). Understanding and implementing the CLT (Communicative Language Teaching) paradigm. *RELC journal*, 34(1), pp. 5-30. Retrieved August 5, 2017 from <https://files.eric.ed.gov/fulltext/ED574120.pdf>
- Jafarpur, A. (1991). Can Naive EFL Learners Estimate Their Own Proficiency? *Evaluation & Research in Education* 5(3), pp. 145-157.

- Javid, C. Z., & Umer, M. (2013). Investigating English language needs: Medical undergraduates perspective in a Saudi context. *Pakistan Journal of Social Sciences*, 33(2), pp. 363-377. Retrieved June 20, 2016 from <https://pdfs.semanticscholar.org/c1b9/43231834592e1a8d0d6b03b63f81df32825f.pdf>
- Javid, C. Z., Farooq, U., & Gulzar, M. A. (2012). Saudi English-major undergraduates and English teachers' perceptions regarding effective ELT in the KSA: A comparative study. *European Journal of Scientific Research*, 85(1), pp. 55-70.
- Jefferies, D., McNally, S., Roberts, K., Wallace, A., Stunden, A., D'Souza, S., & Glew, P. (2017). The importance of academic literacy for undergraduate nursing students and its relationship to future professional clinical practice: A systematic review. *Nurse Education Today*, 60, pp. 84-91.. doi:10.1016/j.nedt.2017.09.020. Retrieved April 8, 2018 from <https://www.sciencedirect.com/science/article/pii/S0260691717302319?via%3Dihub>
- Jiménez-Muñoz, A. (2014). Measuring the impact of CLIL on language skills: a CEFR-based approach for Higher Education. *Language Value*, 6(1), pp. 28-50. Retrieved May 25, 2017 from <http://www.languagevalue.uji.es/index.php/languagevalue/article/view/96>
- Johns, A. M., & Dudley-Evans, T. (1991). English for Specific Purposes: International in Scope, Specific in Purpose. *TESOL Quarterly*, 25(2), pp. 297-314.
- Johnson, P. (1988). English language proficiency and academic performance of undergraduate international students. *TESOL Quarterly* 22(1), pp. 164-168.
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), pp. 14-27.
- Jones, N. (2002). Relating the ALTE framework to the Common European Framework of Reference. In J. C. Alderson (Ed.), *Common European Framework of Reference for Languages: Learning, teaching, assessment. Case studies* (pp. 167-183). Strasbourg: Council of Europe.
- Jones, N., & Saville, N. (2009). European language policy: Assessment, learning, and the CEFR. *Annual Review of Applied Linguistics*, 29, pp. 51-63. doi: 10.1017/S0267190509090059
- Jones, S. R., Carley, S., & Harrison, M. (2003). An introduction to power and sample size estimation. *Emergency Medicine Journal*, 20, pp. 453-458. Retrieved April 22, 2015 from <https://emj.bmj.com/content/emj/20/5/453.full.pdf>
- Kaur, S. (2003). Evaluating Teaching Effectiveness in Higher Education: A Case Study. *Bulletin of Higher Education Research* 2, pp. 2-4. Retrieved March 28, 2016 from <https://core.ac.uk/download/pdf/83543065.pdf>
- Kellaghan, T., & Greaney, V. (2001). Using assessment to improve the quality of education. Paris: UNESCO, International Institute for Educational Planning.

Retrieved September 23, 2017 from
<http://unesdoc.unesco.org/images/0012/001262/126231e.pdf>

- Khafaji, A. I. A. (2004). An Evaluation of the Materials used for Teaching English to the Second Secondary Level in Male Public High Schools in Saudi Arabia. (Master's thesis. University of Exeter, Exeter).
- Khan, I. A. (2011). Learning difficulties in English: Diagnosis and pedagogy in Saudi Arabia. *Educational Research*, 2(7), pp. 1248-1257. Retrieved February 21, 2016 from
https://www.researchgate.net/publication/234168115_Learning_difficulties_in_English_Diagnosis_and_pedagogy_in_Saudi_Arabia
- Khuwaileh, A. A., & Shoumali, A. A. (2000). Writing errors: A study of the writing ability of Arab learners of academic English and Arabic at university. *Language Culture and Curriculum*, 13(2), pp. 174-183.
- Knodel, J. (1993). The design and analysis of focus group studies: A practical approach. In D. L. Morgan (Ed.), *Successful focus groups: Advancing the state of the art* (pp. 35-50). Thousand Oaks, CA: SAGE Publications Ltd. doi: 10.4135/9781483349008
- Kroll, B. (1990). *Second Language Writing: Research Insights for the Classroom*. Cambridge: Cambridge University Press.
- Krueger, R. A. (1998). Moderating focus groups. *Focus group kit*, 4. Thousand Oaks, CA: Sage Publications Ltd.
- Krueger, R. A., & Casey, M. A. (2000). *Focus groups. A practical guide for applied research* (3rd ed.) Thousand Oaks, CA: Sage.
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of personality and social psychology*, 77(6), pp. 1121-1134. Retrieved November 30, 2016 from
<https://www.avaresearch.com/files/UnskilledAndUnawareOfIt.pdf>
- Kuiken, F., & Vedder, I. (2008). Cognitive task complexity and written output in Italian and French as a foreign language. *Journal of Second Language Writing*, 17(1), pp. 48-60.
- Kun, A. I. (2016). A comparison of self versus tutor assessment among Hungarian undergraduate business students. *Assessment & Evaluation in Higher Education*, 41(3), pp. 350-367.
- Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), pp. 159-174. Retrieved June 12, 2016 from
https://www.dentalage.co.uk/wp-content/uploads/2014/09/landis_jr_koch_gg_1977_kappa_and_observer_agreement.pdf

- Lea, M. R., & Stierer, B. (Eds.) (2000). *Student writing in higher education: New contexts*. Buckingham: Society for Research into Higher Education/Open University Press.
- Lea, M., & Street, B. (1998). Student writing in higher education: An academic literacies approach. *Studies in Higher Education*, 23(2), pp. 157-172.
- Lea, M. R., & Street, B. V. (2006). The “academic literacies” model: Theory and applications. *Theory Into Practice*, 45(4), pp. 368-377. doi: 10.1207/s15430421tip4504_11
- Leach, L. (2012). Optional self-assessment: some tensions and dilemmas. *Assessment & Evaluation in Higher Education*, 37(2), pp. 137-147. Retrieved July 18, 2016 from <https://srhe.tandfonline.com/doi/pdf/10.1080/02602938.2010.515013?needAccess=true>
- LeBlanc, R., & Painchaud, G. (1985). Self-assessment as a second language placement instrument. *Tesol Quarterly*, 19(4), pp. 673-687.
- Lederman, L. C. (1990). Assessing educational effectiveness: The focus group interview as a technique for data collection. *Communication Education*, 39(2), pp. 117-127.
- Lee, I. (2013). Research into practice: Written corrective feedback. *Language Teaching*, 46(1), 108-119.
- Lejk, M., & Wyvill, M. (2001). The effect of the inclusion of self-assessment with peer assessment of contributions to a group project: A quantitative study of secret and agreed assessments. *Assessment & Evaluation in Higher Education*, 26(6), pp. 551-561.
- Leki, I. (2007). *Undergraduates in a second language: Challenges and complexities of academic literacy development*. Mahwah, NJ: Lawrence Erlbaum.
- Levene, H. (1960). Robust tests for equality of variance. In I. Olkin (Ed.) *Contributions to Probability and Statistics* (pp. 278-292). Palo Alto, CA: Stanford University Press.
- Lewin, K., & Dunne, M. (2000). Policy and Practice in Assessment in Anglophone Africa: does globalisation explain convergence?. *Assessment in Education: Principles, Policy & Practice*, 7(3), pp. 379-399.
- Lincoln, Y. S. & Guba, E. G. (2000). Paradigmatic controversies, contradictions, and emerging confluences. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 191-216). London: Sage.
- Litosseliti, L. (2003). *Using focus groups in research*. London: Bloomsbury Publishing PLC.

- Little, D., & Perclová, R. (2001). *The European Language Portfolio: Guide for teachers and teacher trainers*. Strasbourg: Council of Europe. Retrieved from <https://rm.coe.int/1680459fa6>
- Little D. (2011). The Common European Framework of Reference for Languages: A research agenda. *Language Teaching*, 44(3), pp. 381-393. doi: 10.1017/S0261444811000097
- Little, D. (2002). The European Language Portfolio: Structure, origins, implementation and challenges. *Language Teaching*, 35(3), pp. 182-189. doi: 10.1017/S0261444802001805. Retrieved March 17, 2016 from https://www.researchgate.net/publication/231870127_The_European_Language_Portfolio_Structure_origins_implementation_and_challenges
- Little, D. (2005). The Common European Framework and the European Language Portfolio: Involving learners and their judgements in the assessment process. *Language Testing* 22(3), pp. 321–336. Retrieved October 28, 2017 from <https://www.dcu.ie/sites/default/files/carpe/Little%20%282005%29.pdf>
- Little, D. (2006). The Common European Framework of Reference for Languages: Content, purpose, origin, reception and impact. *Language Teaching*, 39(3), pp. 167-190. doi: 10.1017/S0261444806003557
- Little, D. (2007). The Common European Framework of Reference for Languages: Perspectives on the making of supranational language education policy. *Modern Language Journal*, 91(4), pp. 645-655.
- Little, D., & Erickson, G. (2015). Learner identity, learner agency, and the assessment of language proficiency: Some reflections prompted by the Common European Framework of Reference for languages. *Annual Review of Applied Linguistics*, 35, pp. 120-139.
- Liu, J. Y., Chang, Y. J., Yang, F. Y., & Sun, Y. C. (2011). Is what I need what I want? Reconceptualising college students' needs in English courses for general and specific/academic purposes. *Journal of English for Academic Purposes*, 10(4), pp. 271-280.
- Lok, B., McNaught, C., & Young, K. (2016). Criterion-referenced and norm-referenced assessments: compatibility and complementarity. *Assessment & Evaluation in Higher Education*, 41(3), pp. 450-465.
- Long, M. H. (Ed.) (2005). *Second language needs analysis*. Cambridge: Cambridge University Press.
- Lowie, W. M., Haines, K. B., & Jansma, P. N. (2010). Embedding the CEFR in the academic domain: Assessment of language tasks. *Procedia-Social and Behavioral Sciences*, 3, pp. 152-161.
- Lucas, P. W., Lenstrup, M., Prinz, I., Williamson, D., Yip, H., & Tipoe, G. (1997). Language as a barrier to the acquisition of anatomical knowledge. *Medical Education*, 31(2), pp. 81-86.

- Lumley, T., Diehr, P., Emerson, S., & Chen, L. (2002). The importance of the normality assumption in large public health data sets. *Annual Review of Public Health*, 23(1), pp. 151-169.
- Luoma, S., & Tarnanen, M. (2003). Creating a self-rating instrument for second language writing: from idea to implementation. *Language Testing*, 20(4), pp. 440-465. doi: 10.1191/0265532203lt267oa
- Macdonald, J. (2004). Developing competent e-learners: the role of assessment. *Assessment & Evaluation in Higher Education*, 29(2), pp. 215-226.
- Maher, M. (1986). The Development of English as an international language of medicine. *Applied Linguistics*, 7(2), pp. 206-218.
- McMullen, M. G. (2009). Using language learning strategies to improve the writing skills of Saudi EFL students: Will it really work?. *System: An International Journal of Educational Technology and Applied Linguistics*, 37(3), pp. 418-433. doi: 10.1016/j.system.2009.05.001.
- McMullen, M. G. (2014). The value and attributes of an effective preparatory English program: Perceptions of Saudi university students. *English Language Teaching*, 7(7), pp. 131-140. Retrieved June 4, 2015 from <http://www.ccsenet.org/journal/index.php/elt/article/view/37822>
- McNamara, T. (2006). Validity in language testing: The challenge of Sam Messick's legacy. *Language Assessment Quarterly*, 3(1), pp. 31-51.
- McNamara, T., Morton, J., Storch, N., & Thompson, C. (2018). Students' Accounts of Their First-Year Undergraduate Academic Writing Experience: Implications for the Use of the CEFR. *Language Assessment Quarterly*, 15(1), pp. 16-28. doi: 10.1080/15434303.2017.1405420
- Menard, S. (2002). *Longitudinal research (2nd ed.)*. Sage university papers series. Quantitative applications in the social sciences; no. 07-76. Thousand Oaks, CA: Sage Publications Inc.
- Mercer, J. (2007). The challenges of insider research in educational institutions: Wielding a double-edged sword and resolving delicate dilemmas. *Oxford Review of Education*, 33(1), pp. 1-17.
- Messuri, K. (2015). Clarity in medical writing. *The Southwest Respiratory and Critical Care Chronicles*, 3(12), pp. 56-58. Retrieved December 23, 2017 from <http://pulmonarychronicles.com/index.php/pulmonarychronicles/article/view/227/575>
- Miles, J., & Shevlin, M. (2001). *Applying regression and correlation: A guide for students and researchers*. London: Sage Publications Ltd.
- Miller, T. M., & Geraci, L. (2011). Unskilled but aware: Reinterpreting overconfidence in low-performing students. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(2), pp. 502-506.

- Ministry of Education (MoE). 2002. General directorate of curricula (2002): English for Saudi Arabia. Jeddah: Almadina Almunawara Press.
- Ministry of Education, English Department. (2012). Teaching English language course. (No. 912/9-2).
- Ministry of Higher Education (2010). Saudi Arabia: Kingdom of humanity [WWW Document]. Ministry of Higher Education. Retrieved January 15, 2015 from <http://www.mohe.gov.sa/en/studyinside/aboutKSA/Pages/default.aspx>
- Mohammad, T., & Hazarika, Z. (2016). Difficulties of learning EFL in KSA: Writing skills in context. *International Journal of English Linguistics*, 6(3), pp. 105-117. Retrieved March 22, 2017 from <http://www.ccsenet.org/journal/index.php/ijel/article/view/60143/32267>
- Moonen, M., Stoutjesdijk, E., de Graaff, R., & Corda, A. (2013). Implementing the CEFR in secondary education: Impact on FL teachers' educational and assessment practice. *International Journal of Applied Linguistics*, 23(2), pp. 226–246. doi: 10.1111/ijal.12000
- Moore, P. J., & Harrington, M. (2016). Fractionating English language proficiency: policy and practice in Australian higher education. *Current Issues in Language Planning*, 17(3-4), pp. 385-404.
- Morrow, K. (Ed.) (2004). *Insights from the Common European Framework*. Oxford: Oxford University Press.
- Morse, J. M. (1991). Approaches to qualitative-quantitative methodological triangulation. *Nursing Research*, 40(2), pp. 120-123.
- Munby, J. (1978). *Communicative Syllabus Design*. Cambridge: Cambridge University Press.
- Muñoz, A. J. J. (2014). Measuring the impact of CLIL on language skills: a CEFR-based approach for Higher Education. *Language Value*, 6(1), pp. 28-50.
- Murphy, J. (2003). Task-based learning: the interaction between tasks and learners. *ELT Journal*, 57(4), pp. 352-360.
- Murray, N. (2010). Considerations in the post-enrolment assessment of English language proficiency: Reflections from the Australian context. *Language Assessment Quarterly*, 7(4), pp. 343-358.
- Murray, N. (2011). University post-enrolment English language assessment: A consideration of the issues. *Journal of the Education Research Group of Adelaide*, 2(1), pp. 27-33.
- Murray, N. (2013). Widening Participation and English Language Proficiency: A Convergence with Implications for Assessment Practices in Higher Education. *Studies in Higher Education*, 38(2), pp. 299–311. doi: 10.1080/03075079.2011.580838

- Murray, N. (2016a). *Standards of English in higher education: Issues, challenges and strategies*. Cambridge: Cambridge University Press.
- Murray, N. (2016b). An academic literacies argument for decentralizing EAP provision. *ELT Journal*, 70(4), pp. 435-443.
- Murray, N., & Hicks, M. (2016). An institutional approach to English language proficiency. *Journal of Further and Higher Education*, 40(2), 170-187.
- Murray, N., & Nallaya, S. (2014). Embedding academic literacies in university programme curricula: a case study. *Studies in Higher Education*, doi: 10.1080/03075079.2014.981150. 1-17.
- Murray, N., & Nallaya, S. (2016). Embedding academic literacies in university programme curricula: a case study. *Studies in Higher Education*, 41(7), pp. 1296-1312. doi: 10.1080/03075079.2014.981150
- Muthanna, A. (2016). Plagiarism: A shared responsibility of all, current situation, and future actions in Yemen. *Accountability in Research*, 23(5), pp. 280-287. doi: 10.1080/08989621.2016.1154463
- Nagai, N., & O'Dwyer, F. (2011). The actual and potential impacts of the CEFR on language education in Japan. *Synergies Europe*, 6, pp. 141-152.
- Nation, I. S. P., & Macalister, J. (2009). *Language curriculum design*. New York: Routledge.
- Nazim, M., & Hazarika, Z. (2017). Efficacy of ESP in EFL Context: A Case Study of Saudi Arabia. *Arab World English Journal (AWEJ)*, 8(1), pp. 145-164.
- Neff, J., Bunce, C., Dafouz, E., Gallego, J., Rica, J. P., Genis, M., & McCabe, A. (2008). The question of descriptors for academic writing in European Language Framework: A critical view. *Kalbotyra*, 59(3), pp. 213-221. Retrieved September 7, 2016 from https://www.researchgate.net/publication/305202884_The_Question_of_Descriptors_for_Academic_Writing_in_European_Language_Framework_a_Critical_View
- Neff-van Aertselaer, J. (2013). Contextualizing EFL argumentation writing practices within the Common European Framework descriptors. *Journal of Second Language Writing*, 22(2), pp. 198-209.
- North, B. (2000). *The development of a common framework scale of language proficiency*. New York: Peter Lang.
- North, B. (2006, March). *The Common European Framework of Reference: Development, theoretical and practical issues*. Paper presented at the symposium *A New Direction in Foreign Language Education: The Potential of the Common European Framework of Reference for Languages*, Osaka University of Foreign Studies, Japan. Retrieved August 3, 2016 from

https://www.researchgate.net/publication/251995323_The_Common_European_Framework_of_Reference_Development_Theoretical_and_Practical_issues

- North, B. (2007) Descriptors for C2, C1 (and B2+): Calibrated/Non-calibrated/Qualitative Analysis. Eurocentres, 14.04.2007. Retrieved from http://www.ealta.eu.org/documents/resources/C2%20_C1%20descriptors.pdf [Accessed on May 14, 2016].
- North, B. (2014). *The CEFR in practice*. Cambridge: Cambridge University Press.
- North, S. (2005). Disciplinary variation in the use of theme in undergraduate essays. *Applied Linguistics*, 26(3), pp. 431-452.
- Nunan, D. (1988). *The learner-centred curriculum: A study in second language teaching*. Cambridge: Cambridge University Press.
- Nunan, D. (1998). Teaching grammar in context. *ELT Journal*, 52(2), pp. 101-09.
- Obeid, R. (2017). Second language writing and assessment: Voices from within the Saudi EFL context. *English Language Teaching*, 10(6), pp. 174-181.
- O'Cathain, A., & Thomas, K. J. (2004). "Any other comments?" Open questions on questionnaires – A bane or a bonus to research? *BMC Medical Research Methodology*, 4(1), 25. Retrieved January 12, 2017 from <https://bmcmmedresmethodol.biomedcentral.com/track/pdf/10.1186/1471-2288-4-25>
- Önder Özdemir, N. (2014). Diagnosing the EAP needs of Turkish medical students: A longitudinal critical needs analysis. *Ibérica*, 28, pp. 35-58. Retrieved January 14, 2016 from http://www.aelfe.org/documents/02_28_Onder.pdf
- Oraif, I. M. (2016). The right approach in practice: A discussion of the applicability of EFL writing practices in a Saudi context. *English Language Teaching*, 9(7), pp. 97-102. Retrieved April 15, 2017 from <https://files.eric.ed.gov/fulltext/EJ1101748.pdf>
- Oscarson, M. (2013a). Self-Assessment in the Classroom. In A. J. Kunnan (Ed.), *The companion to language assessment*, (Vol 2, pp. 712-729). New York: Wiley-Blackwell.
- Oscarson, M. (2013b). The challenge of student self-assessment in language education. *Voices in Asia Journal*, 1(1), 1-14. Retrieved May 10, 2017 from http://voiceinasiajournal.com/images/pdf/vol1/File_5_VIAJ_1.1_Mats_Oscarson_pp_1-14.pdf.pdf
- Oscarson, M. (1989). Self-assessment of language proficiency: rationale and applications. *Language Testing* 6(1), 1–13.
- Oskarsson, M. (1984). *Self-Assessment of Foreign Language Skills: A Survey of Research and Development Work*. Strasbourg: Council for Cultural Cooperation.

- O'Sullivan, B., & Weir, C. J. (2011). Test development and validation. In B. O'Sullivan (Ed.), *Language Testing: Theories and Practices* (pp. 13-32). Basingstoke, UK: Palgrave Macmillan.
- Pallant, J. (2013). *SPSS survival manual* (5th ed.). Maidenhead, UK: McGraw-Hill Education.
- Papageorgiou, S., Tannenbaum, R. J., Bridgeman, B., & Cho, Y. (2015). The association between TOEFL iBT® test scores and the Common European Framework of Reference (CEFR) levels. Princeton, NJ: Educational Testing Service. Retrieved March 24, 2018 from <https://www.ets.org/Media/Research/pdf/RM-15-06.pdf>
- Paris, S. G., & Winograd, P. (1990). How metacognition can promote academic learning and instruction. In B. F. Jones & L. Idol (Eds.), *Dimensions of thinking and cognitive instruction* (pp. 15-51). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Pecorari, D. (2006). Visible and occluded citation features in postgraduate second language writing. *English for Specific Purposes*, 25, pp. 4-29. Retrieved June 7, 2017 from <http://www.idt.mdh.se/kurser/computing/DVA403/DVA403-2012/Lectures/Pecorari%202006.pdf>
- Pill, J., & Harding, L. (2013). Defining the language assessment literacy gap: Evidence from a parliamentary inquiry. *Language Testing*, 30(3), pp. 381-402.
- Prescott, P. (2014). Student's t-Tests. In N. Balakrishnan, T. Colton, B. Everitt, W. Piegorisch, F. Ruggeri & J. L. Teugels (Eds.), *Wiley StatsRef: Statistics Reference Online*. doi: 10.1002/9781118445112.stat01861
- Present-Thomas, R. L., Weltens, B., & de Jong, J. H. (2013). Defining proficiency: A comparative analysis of CEF level classification methods in a written learner corpus. *Dutch Journal of Applied Linguistics*, 2(1), 57-76.
- Punch, K. (1998). *Introduction to social research: quantitative and qualitative approaches*. London: Sage Publications Ltd.
- Punch, K. F. (2005). *Introduction to social research: Quantitative and qualitative approaches* (2nd ed.). London: Sage Publications Ltd.
- Rabab'ah, G. (2005). Communication problems facing Arab learners of English. *Journal of Language and Learning* 3(1), pp. 180-197. Retrieved February 23, 2017 from http://webpace.buckingham.ac.uk/kbernhardt/journal/jllearn/3_1/rababah.pdf
- Rababah, G. (2002). Communication Problems Facing Arab Learners of English. *TEFL Web Journal*, 2(1). Retrieved October 17, 2017, from <https://files.eric.ed.gov/fulltext/ED473079.pdf>
- Radnor, H. A. (2001). *Researching Your Professional Practice: Doing Interpretive Research*. Philadelphia: Open University Press.

- Rauf, M. (2015). Best practices in English language testing at the university Preparatory Year Programs. In A. K. Hamdan (Ed.), *Teaching and Learning in Saudi Arabia* (pp. 185-205). Rotterdam: SensePublishers.
- Reed, J. F., III, & Stark, D. B. (1988). Robust alternatives to traditional analysis of variance: Welch W*, James II*, James III*, Brown-Forsythe BF*. *Computer Methods and Programs in Biomedicine*, 26(3), pp. 233-237.
- Reichelt, M. (1999). Toward a more comprehensive view of L2 writing: Foreign language writing in the US. *Journal of Second Language Writing*, 8(2), pp. 181-204.
- Resnik, D. B., Rasmussen, L. M., & Kissling, G. E. (2015). An international study of research misconduct policies. *Accountability in Research*, 22(5), pp. 249-266.
- Rex, L.A., & McEachen, D. (1999). If anything is odd, inappropriate, confusing, or boring, it's probably important: The emergence of inclusive academic literacy through English classroom discussion practices. *Research in the Teaching of English*, 34(1), pp. 65-127.
- Richards, J. C., Platt, J., & Platt, H. (1992). *Longman dictionary of language teaching & applied linguistics*. Essex: Longman.
- Richards, J. C. (2001). *Curriculum development in language teaching*. New York: Cambridge University Press.
- Richards, J. C., & Schmidt, R. W. (2010). *Longman dictionary of language teaching and applied linguistics* (4th ed.). Harlow: Longman.
- Richards, K. (2003). *Qualitative inquiry in TESOL*. Basingstoke, UK: Palgrave Macmillan.
- Robinson, P. C. (1991). *ESP today: A practitioner's guide*. Hemel Hempstead UK: Prentice Hall International (UK) Ltd.
- Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experiential factors. *Language testing*, 15(1), pp. 1-20.
- Ross, J. A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment, Research, and Evaluation*, 11(10), 1-13. Retrieved November 6, 2016 from <https://pareonline.net/pdf/v11n10.pdf>
- Royse, D., Staton-Tindall, M., Badger, K., & Webster, J. M. (2009). *Needs assessment*. New York: Oxford University Press.
- Russell, D. R., Lea, M., Parker, J., Street, B., & Donahue T. (2009). Exploring notions of genre in "academic literacies" and "writing across the curriculum": Approaches across countries and contexts." In C. Bazerman, A. Bonini, & D. Figueiredo (Eds.), *Genre in a Changing World: Perspectives on Writing* (pp. 459-91). Fort Collins, CO: WAS Clearinghouse/Parlor Press.

- Saavedra, R., & Kwun, S. K. (1993). Peer evaluation in self-managing work groups. *Journal of Applied Psychology*, 78(3), pp. 450-462.
- Sabbour, S. M., Dewedar, S. A., & Kandil, S. K. (2010). Language barriers in medical education and attitudes towards Arabization of medicine: Student and staff perspectives. *Eastern Mediterranean Health Journal* 16(12), pp. 1263-1271. Retrieved August 23, 2017 from http://applications.emro.who.int/emhj/V16/12/16_12_2010_1263_1271.pdf?ua=1
- Sahinkarakas, S., & Arifi, Q. (2007). The CEFR and the needs of the ESP students. Retrieved May 4, 2016 from https://www.researchgate.net/publication/237406920_The_CEFR_and_the_needs_of_the_ESP_students
- Sahragard, R., & Mallahi, O. (2014). Relationship between Iranian EFL learners' language learning styles, writing proficiency and self-assessment. *Procedia-Social and Behavioral Sciences*, 98, pp. 1611-1620.
- Saldaña, J. (2015). *The coding manual for qualitative researchers*. London: Sage Publications Ltd.
- Salkind, N. J. (Ed.). (2010). *Encyclopedia of research design*. Thousand Oaks, CA: SAGE Publications Ltd. doi: 10.4135/9781412961288
- Saville, N. (2005). INTERVIEW: An Interview With John Trim at 80. *Language Assessment Quarterly: An International Journal*, 2(4), pp. 263-288.
- Saville, N. (2010, October 1). The CEFR: Handle with care. *EL Gazette*, 201010.
- Sebba, J., Crick, R. D., Yu, G., Lawson, H., Harlen, W., & Durant, K. (2008). Systematic review of research evidence of the impact on students in secondary schools of self and peer assessment. Technical report. In *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.
- Sebolai, K. (2016). Distinguishing between English proficiency and academic literacy in English. *Language Matters*, 47(1), pp. 45-60.
- Sheshsha, J. A. (1982). *The qualifications of a competent teacher of English in Saudi Arabia as perceived by successful EFL teachers and selected TESOL specialists* (Doctoral dissertation, Indiana University).
- Shimura, M. (2006). Peer- and instructor assessment of oral presentations in Japanese university EFL classrooms: A pilot study. *Waseda Global Forum*, 3, pp. 99-107. Retrieved October 22, 2016 from <https://core.ac.uk/download/pdf/46872246.pdf>
- Shokrpour, N., & Fallahzadeh, M. H. (2007). A survey of the students and interns' EFL writing problems in Shiraz University of Medical Sciences. *Asian EFL Journal*, 9(1), pp. 147-163.

- Shukri, N. (2008). Exploring ESP/Medical biology teacher collaboration, and medical students' and their teachers' perceptions of writing needs in an Arab university: A case study (Doctoral dissertation) Retrieved June 13, 2017 from https://www.researchgate.net/publication/41500891_Exploring_ESPMedical_biology_teacher_collaboration_and_medical_students%27_and_their_teachers%27_perceptions_of_writing_needs_in_an_Arab_University_a_case_study
- Shukri, N. A. (2014). Second Language Writing and Culture: Issues and Challenges from the Saudi Learners' Perspective. *Arab World English Journal*, 5(3), pp. 190-207. Retrieved January 17, 2017 from https://www.researchgate.net/publication/305730164_Second_Language_Writing_and_Culture_Issues_and_Challenges_from_the_Saudi_Learners%27_Perspective
- Smeeton, N. C. (1985). Early history of the kappa statistic. *Biometrics*, 41(3), p. 795. Retrieved June 11, 2017 from <https://www.jstor.org/stable/2531300>
- Smith, L., & Abouammoh, A. (2013). Higher education in Saudi Arabia: Reforms, challenges and priorities. In Smith L., & Abouammoh A. (Eds.), *Higher Education in Saudi Arabia* (pp. 1-12). Dordrecht, the Netherlands: Springer.
- Snape, D., & Spencer, L. (2003) The foundations of qualitative research, in: J. Ritchie & J. Lewis (Eds.), *Qualitative Research Practice: a guide for social science students and researchers* (pp. 1-23). Thousand Oaks: Sage.
- Stefani, L. A. (1994). Peer, self and tutor assessment: Relative reliabilities. *Studies in Higher Education*, 19(1), pp. 69-75.
- Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104(3), pp. 743-762.
- Swales, J. M., & Feak, C. B. (2004). *Academic writing for graduate students: Essential tasks and skills*. Ann Arbor, MI: University of Michigan Press.
- Swan, M., & Smith, B. (Eds.). (2001). *Learner English: A teacher's guide to interference and other problems* (2nd ed.). Cambridge: Cambridge University Press.
- Tannenbaum, R. J., & Wylie, E. C. (2005). Mapping English language proficiency test scores onto the Common European Framework. ETS Research Report Series, RR-05-18, TOEFL-RR-80. Princeton, NJ: Educational Testing Service.
- Taras, M. (2001). The use of tutor feedback and student self-assessment in summative assessment tasks: towards transparency for students and for tutors. *Assessment & Evaluation in Higher Education*, 26(6), pp. 605-614.
- Tashakkori, A., & Teddlie, C. (2003). Issues and dilemmas in teaching research methods courses in social and behavioural sciences: US perspective. *International Journal of Social Research Methodology*, 6(1), pp. 61-77.

- Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics*, 29, pp. 21-36.
- Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. Thousand Oaks, CA: Sage Publications Inc.
- Tejeiro, R. A., Gómez-Vallecillo, J. L., Romero, A. F., Pelegrina, M., Wallace, A., & Emberley, E. (2012). Summative self-assessment in higher education: Implications of its counting towards the final mark. *Electronic Journal of Research in Educational Psychology*, 10(2), pp. 789-812.
- Thomas, M. (1994). Assessment of L2 proficiency in second language acquisition research. *Language Learning*, 44(2), pp. 307-336.
- Thomas, M. (2006). Research synthesis and historiography: The case of assessment of second language proficiency. In J.M. Norris, & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 279–298). Amsterdam: John Benjamins.
- Tonkyn, A. (1995). English proficiency standards for overseas students: Who needs what level? *Journal of International Education*, 6(3) pp. 37–61.
- Trim, J. L. M. (1978). *Some possible lines of development of an overall structure for a European unit/credit scheme for foreign language learning by adults*. Strasbourg: Council of Europe.
- Tsai, Y., & Tsou, C.-H. (2009). A standardised English language proficiency test as the graduation benchmark: Student perspectives on its application in higher education. *Assessment in Education: Principles, Policy & Practice*, 16(3), pp. 319-330.
- Ünaldı, İ. (2016). Self and teacher assessment as predictors of proficiency levels of Turkish EFL learners. *Assessment & Evaluation in Higher Education*, 41(1), 67-80. doi: 10.1080/02602938.2014.980223
- Üstünlüoğlu, E., Zazaoglu, K. F. A., Keskin, M. N., Sarayköylü, B., & Akdoğan, G. (2012). Developing a CEF based curriculum: A case study. *International Journal of Instruction*, 5(1), pp. 115-128.
- Wang, M. F., & Bakken, L. L. (2004). An academic writing needs assessment of English-as-a-second-language clinical investigators. *Journal of Continuing Education in the Health Professions*, 24(3), pp. 181-189.
- Wang, Y. (2012). Differences in L1 and L2 Academic Writing. *Theory and Practice in Language Studies*, 2(3), pp. 637-641.
- Weaver, C. (1996). *Teaching Grammar in Context*. Portsmouth: Boynton/Cook Publishers, Inc.
- Weigle, S. C. (1994). Effects of training on raters of ESL compositions. *Language Testing*, 11(2), pp. 197-223. doi: 10.1177/026553229401100206

- Weir, C. J. (2005). Limitations of the Common European Framework for developing comparable examinations and tests. *Language Testing*, 22(3), pp. 281-300. doi: 10.1191/0265532205lt309oa
- West, R. (1994). Needs Analysis in Language Teaching, In *Language Teaching*, 27(1), pp. 1-19. doi: 10.1017/S0261444800007527
- West, R. (1997). Needs Analysis: State of the Art. In R. Howard & G. Brown (Eds.), *Teacher Education for Languages for Specific Purposes* (pp. 68-79). Clevedon, UK: Multilingual Matters Ltd.
- Yurdugül, H. (2008). Minimum sample size for Cronbach's coefficient alpha: A Monte-Carlo study. *Hacettepe University Journal of Education*, 35, pp. 1-9.
- Yushau, B., & Omar, M. H. (2007). Preparatory year program courses as predictors of first calculus course grade. *Mathematics and Computer Education* 41(2), pp. 92 – 108.
- Zaid, M. A. (1993). Comprehensive analysis of the current system of teaching English as a foreign language in the Saudi Arabia intermediate schools (Doctoral dissertation, University of Colorado).
- Ziebland, S., & McPherson, A. (2006). Making sense of qualitative data analysis: an introduction with illustrations from DIPEX (personal experiences of health and illness). *Medical education*, 40(5), pp. 405-414. doi: 10.1111/j.1365-2929.2006.02467.x
- Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.) *Handbook of self-regulation* (pp. 13-39). New York: Academic Press.

Appendices

Appendix A

A1: PYP Students' Self-Assessment Questionnaire

Background Information:

You are invited to participate in a study that seeks to identify Preparatory Year Medical Track students' perceptions of their writing levels by self-assessing themselves using an open-ended questionnaire aligned to the *Common European Framework of Reference for Languages* CEFR. The questionnaire consists of three parts. In the first part, you are asked to write a *letter to a friend* joining the same track as yours in the Preparatory Year next year. In the second part, you are asked to assess your writing skills, using the CEFR grid (please familiarize yourself with the instructions provided on how to use the self-assessment grid before you start). The last part will request general information about you.

Part One:

Instructions:

Please read the following prompt carefully, and then write a letter to your friend. You have the option to write it in English or Arabic. (You are kindly asked to write about 150 to 200 words and spend about 20 to 30 minutes on the task.)

Imagine that you have a friend joining the Preparatory Year Programme (PYP) next year and she will be pursuing the same track as yours. Your friend asked you to write her a letter where you express your sincere opinions and thoughts on the following:

First:

1. Which English writing skills will your friend learn when she joins the PYP programme next year?

After that, tell your friend about:

2. The English writing skills you feel you have mastered (can do) by the end of the PYP (please specify those skills).
3. The English writing skills you still need to develop (please list them).

Then, tell your friend:

4. **How prepared you feel with regard to your writing skills after spending** one year in the PYP, and why?
And finally:
5. The advice you would offer in terms of improving their writing during the PYP?

Part Two:

Instructions:

A. The following is a self-assessment grid on which you are going to assess your writing skills. There are ten writing categories. Each category is presented in a separate row. Each row contains a number of descriptions about one of these categories, which have a group of *can do* writing skills. These skills are ordered, from the simplest (number 1) to the most complicated (number 9). Please read these descriptions and assess your writing skills using the instructions below. You should start reading the descriptions from number one, but please note that:

1. If you are able to do number three in the row, for example, this means you should be able to do number one and two as well.
2. If you chose that you are not sure that you can do/or you cannot do number four, for example, this means that you cannot do five and above.

Instructions on how to use the self-assessment grid:

1. Start reading the descriptions from left to right (from right to left in the Arabic version).
2. Read the first description (number one); if you feel that you can do all, or most, of what is mentioned in that description then put a tick (✓) beside the option **yes I can do**.
3. Move to the second description in the same row (number two in the same row) and do the same as above. If you feel you can do what is mentioned in that section, or most of it, then put a tick (✓) beside **yes I can do** and move to the third, and so on.
4. If you come to a description where you feel that you are not sure whether you can do this or not, put a tick (✓) beside the **Not Sure** phrase and move to the following row.
5. If you chose the **Not Sure** phrase, DO NOT continue reading the descriptions in the same row. Go down to the next row and start doing the same as in the previous row.

Please note that there are some empty boxes in some rows; when you reach such an empty box, skip it and move to the next one and so on.

	1	2	3	4	5	6	7	8	9
Overall written production	I can write simple isolated phrases and sentences.	I can write a series of simple phrases and sentences linked with simple connectors like 'and', 'but' and 'because'.		I can write straightforward connected texts on a range of familiar subjects within my interest, by linking a series of shorter discrete elements into a linear sequence.		I can write clear, detailed texts on a variety of subjects related to my field of interest, synthesising and evaluating information and arguments from a number of sources.		I can write clear, well-structured texts of complex subjects, underlining the relevant salient issues, expanding and supporting points of view at some length with subsidiary points, reasons and relevant examples, and rounding off with an appropriate conclusion.	I can write clear, smoothly flowing, complex texts in an appropriate and effective style and a logical structure which helps the reader to find significant points.
	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure
Overall written Interaction	I can ask for or pass on personal details in written form.	I can write short, simple formulaic notes relating to matters in areas of immediate need.		I can convey information and ideas on abstract as well as concrete topics, check information and ask about or explain problems with reasonable precision. I can write personal letters and notes asking for or conveying simple information of immediate relevance, getting across the point I feel to be important.		I can express news and views effectively in writing, and relate to those of others.		I can express myself in writing with clarity and precision, relating to the addressee flexibly and effectively.	As C1
	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure

Types of texts the student can write	I can write very short pieces of writing: isolated words and very short, basic sentences. For example, simple messages, notes forms and postcards.	I can usually write short, simple pieces of writing. For example, simple personal letters, postcards, messages, notes, forms.		I can write a continuous, intelligible text in which elements are connected.		I can write a variety of different texts.		I can write a variety of different texts. I can express myself with clarity and precision, using language flexibly and effectively.	I can write a variety of different texts. I can convey finer shade of meaning precisely. I can write persuasively.
	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure
What student can write	I can write numbers and dates, my name, nationality, address and other personal details required to fill in simple forms when travelling. I can write short, simple sentences linked with connectors such as 'and' or 'then'.	I can write texts typically describe immediate needs, personal events, familiar places, hobbies, work, etc. I can write texts typically consist of short, basic sentences. I can use the most frequent connectors (e.g. and, but because) to link sentences in order to write a story or describe something as a list of points.		I can convey simple information to friends, service people, etc. who feature in everyday life. I can get straightforward points across comprehensively. I can give, in written, news, expresses thoughts about abstract or cultural topics. I can describe experiences, feelings and events in some detail.		I can express news and views in writing effectively, and relate to those of others. I can use a variety of linking words to make clearly the relationships between ideas. My spelling and punctuation are reasonably accurate.		I can produce clear, smoothly flowing, well-structured writing, showing controlled use of organisational patterns, connectors and cohesive devices. I can qualify opinions and statements precisely in relation to degrees of, for example, certainty/uncertainty, beliefs/doubts, and likelihood. My layout, paragraphing and punctuation are consistent and helpful. My spelling is accurate apart from occasional slips.	I can create coherent and cohesive text making full and appropriate use of variety of organizational patterns and a wide range of cohesive devices. My writing is free of spelling errors.
	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure

Vocabulary range and control	I have a basic vocabulary repertoire of isolated words and phrases related to particular concrete situations.	I have a sufficient vocabulary for the expression of basic communicative needs. I have a sufficient vocabulary for coping with simple survival needs. I can control a narrow repertoire dealing with concrete everyday needs.	I have sufficient vocabulary to conduct routine, everyday transactions involving familiar situations and topics.	I have a sufficient vocabulary to express myself with some circumlocutions on most topics pertinent to my everyday life such as family, hobbies and interests, work, travel, and current events. I can show good control of elementary vocabulary but major errors still occur when expressing more complex thoughts or handling unfamiliar topics and situations.		I have a good range of vocabulary for matters connected to my field and most general topics. I can vary formulation to avoid frequent repetition, but lexical gaps can still cause hesitation and circumlocution. My lexical accuracy is generally high, though some confusion and incorrect word choice does occur without hindering communication.		I have a good command of a broad lexical repertoire allowing gaps to be readily overcome with circumlocutions; little obvious searching for expressions or avoidance strategies. Good command of idiomatic expressions and colloquialisms. I have occasional minor slips, but no significant vocabulary errors.	I have a good command of a very broad lexical repertoire including idiomatic expressions and colloquialisms; shows awareness of connotative levels of meaning. I have consistently correct and appropriate use of vocabulary.
	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure
Grammatical accuracy	I have only limited control of a few simple grammatical structures and sentence patterns in a learnt repertoire.	I can use some simple structures correctly, but still systematically makes basic mistakes – for example tends to mix up tenses and forget to mark agreement; nevertheless, it is usually clear what I am trying to write.		I can use reasonably accurately a repertoire of frequently used 'routines' and patterns associated with more predictable situations.	I can communicate with reasonable accuracy in familiar contexts; generally, good control though with noticeable mother tongue influence. Errors occur, but it is clear what I am trying to express.	I have a relatively high degree of grammatical control. I do not make mistakes which lead to misunderstanding.	I have good grammatical control; occasional 'slips' or non-systematic errors and minor flaws in sentence structure may still occur, but they are rare and can often be corrected in retrospect.	I consistently maintain a high degree of grammatical accuracy; errors are rare and difficult to spot.	I maintain consistent grammatical control of complex language, even while attention is otherwise engaged (e.g. in forward planning, in monitoring others' reactions).
	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure

Reports and essays	No descriptors available. Start from number 4	No descriptors available. Start from number 4	No descriptors available. Start from number 4	I can write very brief reports to a standard conventionalized format, which pass on routine factual information and state reasons for actions.	I can write short, simple essays on topics of interest. I can summarize report and give his/her opinion about accumulated factual information on familiar routine and non-routine matters within my field with some confidence.	I can write an essay or report which develops an argument, giving reasons in support of or against a particular point of view and explaining the advantages and disadvantages of various options. I can synthesise information and arguments from a number of sources.	I can write an essay or report which develops an argument systematically with appropriate highlighting of significant points and relevant supporting detail. I can evaluate different ideas or solutions to a problem.	I can write clear, well-structured expositions of complex subjects, underlining the relevant salient issues. I can expand and support points of view at some length with subsidiary points, reasons and relevant examples.	I can produce clear, smoothly flowing, complex reports, articles or essays which present a case, or give critical appreciation of proposals or literacy works. I can provide an appropriate and effective logical structure which helps the reader to find significant points.
	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> Can do <input type="checkbox"/> Not sure	<input type="checkbox"/> Can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure
Note Taking	No descriptors available. Please start reading from column number 4.	No descriptors available. Please start reading from column number 4.	No descriptors available. Please start reading from column number 4.	I can take notes as a list of key points during a straightforward lecture, provided the topic is familiar, and the talk is both formulated in simple language and delivered in clearly articulated standard speech.	I can take notes during a lecture which are precise enough for my own use at a later date, provided the topic is within my field of interest and the talk is clear and well-structured.	I can understand a clearly structured lecture on a familiar subject, and can take notes on points which strike me as important, even though I tend to concentrate on the words themselves and therefore to miss some information.		I can take detailed notes during a lecture on topics in my field of interest, recording the information so accurately and so close to the original that the notes could also be useful to other people.	I am aware of the implications and allusions of what is said and can make notes on them as well as on the actual words used by the speaker.
	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure	<input type="checkbox"/> I can do <input type="checkbox"/> Not sure

Part Three:

Personal information:

1. University ID:	_____
2. Age	_____
3. Level	A [Elementary] B [Intermediate] C [Advanced]
5. Section No.	_____

Further contact and prize draw:

- If you are interested in being contacted for a follow-up interview or to be part of a focus group meeting, please give your name and your preferred email address:

Name _____

Email _____

Contact number _____

- Finishing and submitting the three parts of the questionnaire will automatically enter you into a prize draw unless you inform us accordingly. Please provide us with the email address you wish to be used for the prize draw. The email will be immediately deleted after the prize draw.

Email

Privacy statement:

All data collected is solely for research purposes; no data will be used for any other purpose. All data will be kept confidential by the researcher. Your personal information will be anonymised. It will not be possible to trace your identity.

Thank you very much for your participation.

A2: Phase II Students' Questionnaire

General Information:

You are invited to participate in a study designed to identify the medical/healthcare writing requirements that students are required to have during their first academic year in college. The purpose of the study is to identify the minimum writing requirements the students need in order to be able to meet first-year college demands. At the same time, the study aims to explore the students' experience of the Preparatory Year Programme and to evaluate whether or not it enabled them to meet those requirements.

I would be very grateful if you would respond to all the questions. There will be an immediate prize draw for participants (at the time of participation) and a further prize draw at a later point in time (further details are provided on the information sheet you have).

Part One: Background Information

1. Personal Information (please tick or fill in relevant information):

1.1 Age:

- a. 18-20
- b. 21-23
- c. 24 and above

1.2 Nationality: _____

1.3 I finished the Preparatory Year Programme:

- a. Last year
- b. Two years ago
- c. Others (please specify)_____

1.4 My college is:

- a. College of Medicine
- b. College of Pharmacy
- c. College of Nursing
- d. College of Dentistry
- e. College of Applied Medical Sciences

2. English Background Information (please tick or fill in relevant information):

- 2.1 When did you start learning English?
- a. Early elementary school
 - b. Late elementary school
 - c. Intermediate school
 - d. Other _____
- 2.2 Have you taken an English language course in any English-speaking country?
- a. Yes
 - b. No
- If YES, for how long (in total)? _____
- 2.3 What was your level of English in the Preparatory Year Programme?
- a. A (Elementary)
 - b. B (Intermediate)
 - c. C (Advanced)
 - d. Other _____

Part Two: Required Writing Skills

The purpose of this section is to identify what the medical/healthcare colleges require students to be able to undertake in terms of writing.

1- In the following table:

(a) Please **list** the **writing skills and/or writing tasks** you needed in your college (for example, what kind of writing tasks were you required to do? What did you have to write about? What kind of writing skills did you need in order to write? ...etc.)

(b) Please **rate how well** the Preparatory Year Programme has prepared you to master those listed skills/tasks/activities listed in (a).

	Writing skills/activities and tasks required in your first year of college	How well has the PYP prepared you to develop those skills?			
		Very well prepared	Well prepared	Slightly prepared	Not prepared at all
1					
2					
3					
4					
5					
6					
7					
8					

2- The following section has ten writing scales. The purpose of these scales is **NOT** about identifying your writing abilities and what you can do in terms of writing. The purpose here is to identify the **writing skills the students in their first year of college generally require in order** to meet their colleges' writing demands.

- (a) In each scale (each row), please read all the descriptors provided about the different writing needs for the first year of medical college.
- (b) Then, put a (√) next to the one that best describes the students' needs in their first year at the college.
- (c) Please select only ONE option in each scale (each row).

	1	2	3	4	5	6	7	8	9
1- Overall written production	First year college students need to be able to write simple isolated phrases and sentences.	First year college students need to be able to write a series of simple phrases and sentences linked with simple connectors like 'and', 'but' and 'because'.		First year college students need to be able to write straightforward connected texts on a range of familiar subjects within field of interest, by linking a series of shorter discrete elements into a linear sequence.		First year college students need to be able to write clear, detailed texts on a variety of subjects related to my field of interest, synthesising and evaluating information and arguments from a number of sources.		First year college students need to be able to write clear, well-structured texts of complex subjects, underlining the relevant salient issues, expanding and supporting points of view at some length with subsidiary points, reasons and relevant examples, and rounding off with an appropriate conclusion.	First year college students need to be able to write clear, smoothly flowing, complex texts in an appropriate and effective style and a logical structure, which helps the reader to find significant points.
	■	■		■		■		■	■
2- Overall written interaction	First year college students need to be able to ask for or pass on personal details in written form.	First year college students need to be able to write short, simple formulaic notes relating to matters in areas of immediate need.		First year college students need to be able to convey information and ideas on abstract as well as concrete topics, to be able to check information and ask about or explain problems with reasonable precision. The students need to be able to write personal letters and notes asking for or conveying simple information of immediate relevance, getting across the point they feel to be important.		First year college students need to be able to express news and views effectively in writing, and relate to those of others		First year college students need to be able to express themselves in writing with clarity and precision, relating to the addressee flexibly and effectively.	As mentioned in 8
	■	■		■		■		■	■

3- Types of texts students need to be able to write	First year college students need to be able to write very short pieces of writing: isolated words and very short, basic sentences. For example, simple messages, notes and postcards.	First year college students need to usually be able to (write) short, simple pieces of writing. For example, simple personal letters, postcards, messages, notes, forms.		First year college students need to be able to write a continuous, intelligible text in which elements are connected.		First year college students need to be able to write a variety of different texts.		First year college students need to be able to write a variety of different texts. The students need to be able to express themselves with clarity and precision, using language flexibly and effectively.	First year college students need to be able to write a variety of different texts. The students need to be able to convey finer shade of meaning precisely. They need to be able to write persuasively.
	■	■		■		■		■	■
4- What Students need to be able to write	First year college students need to be able to write numbers and dates, their names, nationality address and other personal details required to fill in simple forms when travelling. The students need to be able to write short simple sentences linked with connectors such as 'and' and 'then'.	First year college students need to be able to write texts typically describe immediate needs, personal events, familiar places, hobbies, work, etc. The students need to be able to write texts typically consist of short, basic sentences. They need to be able to use the most frequent connectors (e.g. and, but because) to link the sentences in order to write a story or describe something as a list of points.		First year college students need to be able to convey simple information to friends, service people, etc. who feature in everyday life. They need to be able to give (write) news, express thoughts about abstract or cultural topics such as films, etc. They need to be able to describe experiences, feelings and events in some details.		First year college students need to be able to express news and views (in writing) effectively, and relate to those of others. The students need to be able to use a variety of linking words to make clearly the relationships between ideas. The students' spelling and punctuation need to be reasonably accurate.		First year college students need to be able to produce clear, smoothly flowing, well-structured writing, showing controlled use of organisational patterns, connectors and cohesive devices. They need to be able to qualify opinions and statements precisely in relation to degrees of, for example, certainty/uncertainty, beliefs/doubts, and likelihood. Their layout, paragraphing and punctuation need to be consistent and helpful. Their spelling needs to be accurate apart from occasional slips.	First year college students need to be able to create coherent and cohesive text marking full and appropriate use of variety of organizational patterns and a wide range of cohesive devices. The students' writing needs to be free of spelling errors.
	■	■		■		■		■	■

5- Vocabulary range and control	First year college students need to have a basic vocabulary repertoire of isolated words and phrases related to particular concrete situations.	First year college students need to have a sufficient vocabulary for the expression of basic communicative needs. The students need to have a sufficient vocabulary for coping with simple survival needs. They need to be able to control a narrow repertoire dealing with concrete everyday needs.	First year college students need to have sufficient vocabulary to conduct routine, everyday transactions involving familiar situations and topics.	First year college students need to have a sufficient vocabulary to express themselves with some circumlocutions on most topics pertinent to their everyday life such as family, hobbies and interests, work, travel, and current events. They need to have good control of elementary vocabulary but it is ok if major errors still occur when expressing more complex thoughts or handling unfamiliar topics and situations.		First year college students need to have a good range of vocabulary for matters connected to their field and most general topics. They need to be able to vary formulation to avoid frequent repetition. The students need a generally high lexical accuracy, and it is ok to have incorrect word choice without hindering communication.		First year college students need to have a good command of a broad lexical repertoire allowing gaps to be readily overcome with circumlocutions. The students need a good command of idiomatic expressions and colloquialisms. It is OK if occasionally make minor slips, but with no significant vocabulary errors.	First year college students need to have a good command of a very broad lexical repertoire including idiomatic expressions and colloquialisms; the students need to show awareness of connotative levels of meaning. They need to have a consistently correct and appropriate use of vocabulary.
	■	■	■	■		■		■	■
6- Grammatical accuracy	First year college students need to only have limited control of a few simple grammatical structures and sentence patterns in a learnt repertoire.	First year college students need to use some simple structures correctly, but it is ok to mix up tenses and to forget to mark agreement; nevertheless, the students usually need to be clear what they are trying to say.		First year college students need to have a reasonably accurate repertoire of frequently used 'routines' and patterns associated with more predictable situations.	First year college students need to communicate with reasonable accuracy in familiar contexts; with generally good control but it is OK to have noticeable mother tongue influence. It is important when errors occur, it is clear what they are trying to express.	First year college students need to have a relatively high degree of grammatical control. They should not make mistakes, which lead to misunderstanding.	First year college students need to have good grammatical control; it is OK to have occasional 'slips' or non-systematic errors in sentence structure but they are rare and can often be corrected in retrospect.	The First year college students need to consistently maintain a high degree of grammatical accuracy; errors should be rare and difficult to spot	First year college students need to maintain consistent grammatical control of complex language, even while attention is otherwise engaged (e.g. in forward planning, in monitoring others' reactions).
	■	■		■	■	■	■	■	■

7- Orthographic control	First year college students need to be able to copy familiar words and short phrases e.g. simple signs or instructions, names of everyday objects, names of shops and set phrases used regularly. They need to be able to spell their address, nationality and other personal details.	First year college students need to be able to copy short sentences on everyday subjects – e.g. directions how to get somewhere. They need to be able to write with reasonable phonetic accuracy (but not necessarily fully standard spelling) short words that are in my oral vocabulary.		First year college students need to be able to produce continuous writing, which is generally intelligible throughout. Spelling, punctuation and layout need to be accurate enough to be followed most of the time.		First year college students need to be able to produce clearly intelligible continuous writing, which follows standard layout and paragraphing conventions. Spelling and punctuation need to be reasonably accurate but may show signs of mother tongue influence.		First year college student' layout, paragraphing and punctuation need to be consistent and helpful. Spelling needs to be accurate, apart from occasional slips of the pen.	First year college students' writing in the first year of college needs to be orthographically free of error.
	■	■		■		■		■	■
8- Processing texts	First year college students need to be able to copy out single words and short texts presented in standard printed format.	First year college students need to be able to copy out short texts in printed or clearly handwritten format.	First year college students need to be able to pick out and reproduce key words and phrases or short sentences from a short text within the learner's limited competence and experience.	First year college students need to be able to collate short pieces of information from several sources and summarise them for somebody else. They need to be able to paraphrase short written passages in a simple fashion, using the original text wording and ordering.		First year college students need to be able to summarise a wide range of factual and imaginative texts, commenting on and discussing contrasting points of view and the main themes. They need to be able to summarise extracts from news items, interviews or documentaries containing opinions, argument and discussion. They need to be able to summarise the plot and sequence of events		First year college students need to be able to summarise long, demanding texts.	First year college students need to be able to summarise information from different sources, reconstructing arguments and accounts in a coherent presentation of the overall result.
	■	■	■	■		■		■	■

9- Reports and essays	No descriptors available. Start from number 4	No descriptors available. Start from number 4	No descriptors available. Start from number 4	First year college students need to be able to write very brief reports to a standard conventionalised format, which pass on routine factual information and state reasons for actions.	First year college students need to be able to write short, simple essays on topics of interest. The students need to be able to summarise, report and give their opinion about accumulated factual information on familiar routine and non-routine matters within their field with some confidence.	First year college students need to be able to write an essay or report which develops an argument, giving reasons in support of or against a particular point of view and explaining the advantages and disadvantages of various options. The students need be able to synthesize information and arguments from a number of sources.	First year college students need to be able to write an essay or report which develops an argument systematically with significant points and relevant supporting detail. The students need to be able to evaluate different ideas or solutions to a problem.	First year college students need to be able to write clear, well-structured expositions of complex subjects, underlining the relevant salient issues. They need to be able to expand and support points of view at some length with subsidiary points, reasons and relevant examples.	First year college students need to be able to produce clear, smoothly flowing, complex reports, articles or essays which present a case, or give critical appreciation of proposals or literacy works. I need to provide an appropriate and effective logical structure, which helps the reader to find significant points.
	■	■	■	■	■	■	■	■	■
10- Note Taking	No descriptors available. Please start reading from column number 4.	No descriptors available. Please start reading from column number 4.	No descriptors available. Please start reading from column number 4.	First year college students need to be able to take notes as a list of key points during a straightforward lecture, provided the topic is familiar, and the talk is both formulated in simple language and delivered in clearly articulated standard speech.	First year college students need to be able to take notes during a lecture which are precise enough for their own use at a later date especially that the topic is within their field of interest and the talk is clear and well-structured.	First year college students need to be able to understand a clearly structured lecture on a familiar subject, and to be able to take notes on points, which strike them as important, and it is OK that sometimes they miss some information.		First year college students need to be able to take detailed notes during a lecture on topics in their field of interest, recording the information so accurately and so close to the original that the notes could also be useful to other people.	First year college students need to be able to be aware of the implications and allusions of what is said and they need to be able to make notes on them as well as on the actual words used by the speaker.
	■	■	■	■	■	■		■	■

(1) Of all the scales that you have just read, are there any item(s) that are particularly important for first-year college students (for example, vocabulary, grammar, etc.)? Please explain:

(2) Of all the scales that you have just read, are there any that are less important or not required at all for first-year college students? Please explain:

(3) If you had the opportunity to modify the Preparatory Year Programme's writing course what changes would you make?

(4) Do you think you still need additional English writing support after the preparatory year programme? Why/why not?

Part Three: Personal Information

Your name: _____

University ID: _____

Contact Information for the Prize Draw (optional):

Email _____

Thank you for your participation ☺

A3: Evaluation Form

EVALUATION FORM

Name (Optional) _____ Track _____

Date: ___/___/2015 Category _____

Please take a moment to rate (from 1 to 5) and evaluate how you feel about the materials of the questionnaire you have just used.

Please tick "✓" the box that most represents your rating and fill in relevant details where required.

General Evaluation:

1. Overall, how would you rate the clarity of the given **instructions** of the questionnaire?

Very clear	1	2	3	4	5	Not clear at all

2. Overall, how **user-friendly** would you rate the questionnaire?

Very user-friendly	1	2	3	4	5	Not user-friendly at all

3. Overall, how would you rate the usefulness of the **layout** of the questionnaire?

Very useful	1	2	3	4	5	Not useful at all

4. Overall, how would you rate the **topic** of the questionnaire?

Very important	1	2	3	4	5	Not important at all

Letter writing:

5. How would you rate the clarity of the **instructions** given to write the *letter to a friend*?

Very clear	1	2	3	4	5	Not clear at all

6. How would you rate the **allocated time** given to write the letter (30 minutes)?

Enough	1	2	3	4	5	Not Enough at all

7. How would you rate the **number of words** required to write the letter (200 words)?

Enough	1	2	3	4	5	Not enough at all

8. How would you rate the merit of being given the **option** to write either in Arabic or English?

	1	2	3	4	5	
--	---	---	---	---	---	--

A very good idea							Not a Good idea at all
------------------	--	--	--	--	--	--	------------------------

9. How would you rate the **content of the letter**?

Very Important	1	2	3	4	5	Not important at all
----------------	---	---	---	---	---	----------------------

10. How would you rate the **questions** used to encourage you write the letter about your English writing?

Enough	1	2	3	4	5	Not Enough at all
--------	---	---	---	---	---	-------------------

I would appreciate any comments/suggestions on the letter part that might help to improve it:

Assessment grid:

11. How would you rate the **instructions** given on how to use the assessment grid?

Very Clear	1	2	3	4	5	Not Clear at all
------------	---	---	---	---	---	------------------

12. How would you rate the assessment grid in terms of **ease of use**?

Very Easy	1	2	3	4	5	Not Easy at all
-----------	---	---	---	---	---	-----------------

13. How would you rate the **time** given to work on the assessment grid (30 minutes)?

Very Short	1	2	3	4	5	Very long
------------	---	---	---	---	---	-----------

14. How would you rate the **layout** of the assessment grid?

Very Useful	1	2	3	4	5	Not Useful at all
-------------	---	---	---	---	---	-------------------

15. How would you rate the **descriptions** of the different given writing skills in the grid?

Very Easy to understand	1	2	3	4	5	Very Difficult to understand
-------------------------	---	---	---	---	---	------------------------------

I would appreciate any comments/suggestions on the assessment grid part that might help to improve it:

A4: The CEFR rating scale

	A1	A2	A2 +	B1	B1 +	B2	B2 +	C1	C2
Range	Has a very basic repertoire of words and simple phrases related to personal details and particular concrete situations	Uses basic sentence patterns with memorized phrases, groups of a few words and formulae in order to communicate limited information mainly in everyday situations.		Has enough language to get by, with sufficient vocabulary to express him/herself with some circumlocutions on topics such as family, hobbies and interests, work, travel, and current events.		Has a sufficient range of language to be able to give clear descriptions, express viewpoints on most general topics, using some complex sentence forms to do so. Language lacks, however, expressiveness and idiomaticity and use of more complex forms is still stereotypic.		Has a good command of a broad range of language allowing him/her to select a formulation to express him/herself clearly in an appropriate style on a wide range of general, academic, professional or leisure topics without having to restrict what he/she wants to say. The flexibility in style and tone is somewhat limited	Shows great flexibility in formulating ideas in differing linguistic forms to convey finer shades of meaning precisely, to give emphasis and to eliminate ambiguity. Also has a good command of idiomatic expressions and colloquialisms.
Coherence	Can link words or groups of words with very basic linear connectors like “and” and “then”.	Can link groups of words with simple connectors like “and”, “but” and “because”.		Can link a series of shorter discrete elements into a connected, linear text		Can use a number of cohesive devices to link his/her sentences into clear, coherent text, though there may be some “jumpiness” in a longer text.		Can produce clear, smoothly flowing, well-structured text, showing controlled use of organisational patterns, connectors and cohesive devices.	Can create coherent and cohesive texts making full and appropriate use of a variety of organisational patterns and a wide range of connectors and other cohesive devices.

Accuracy	Shows only limited control of a few simple grammatical structures and sentence patterns in a memorized repertoire. Errors may cause misunderstandings.	Uses simple structures correctly, but still systematically makes basic mistakes. Errors may sometimes cause misunderstandings.		Uses reasonably accurately a repertoire of frequently used “routines” and patterns associated with more common situations. Occasionally makes errors that the reader usually can interpret correctly on the basis of the context.		Shows a relatively high degree of grammatical control. Does not make errors which cause misunderstandings.		Consistently maintains a high degree of grammatical accuracy; occasional errors in grammar, collocations and idioms.	Maintains consistent and highly accurate grammatical control of even the most complex language forms. Errors are rare and concern rarely used forms
	A1	A2	A2 +	B1	B1 +	B2	B2 +	C1	C2
Description	Can write simple phrases and sentences about themselves and imaginary people, where they live and what they do, etc.	Can write very short, basic descriptions of events, past activities and personal experiences Can write short simple imaginary biographies and simple poems about people		Can write accounts of experiences, describing feelings and reactions in simple connected text. Can write a description of an event, a recent trip – real or imagined. Can narrate a story. Can write straightforward, detailed descriptions on a range of familiar subjects within his field of interest.		Can write clear, detailed descriptions of real or imaginary events and experiences marking the relationship between ideas in clear connected text, and following established conventions of the genre concerned. Can write clear, detailed descriptions on a variety of subjects related to his/her field of interest. Can write a review of a film, book or play.		Can write clear, detailed, well-structured and developed descriptions and imaginative texts in a mostly assured, personal, natural style appropriate to the reader in mind.	Can write clear, smoothly flowing and fully engrossing stories and descriptions of experience in a style appropriate to the genre adopted.
	Overall	Can write simple isolated phrases and sentences. Longer texts contain expressions and show coherence problems which make the text very hard or impossible to understand.	Can write a series of simple phrases and sentences linked with simple connectors like “and”, “but” and “because”. Longer texts may contain expressions and show coherence problems which makes the text hard to understand.	Can write straightforward connected texts on a range of familiar subjects within his field of interest, by linking a series of shorter discrete elements into a linear sequence. The texts are understandable but occasional unclear expressions and/or inconsistencies may cause a break-up in reading.	Can write clear, detailed official and semi-official texts on a variety of subjects related to his field of interest, synthesising and evaluating information and arguments from a number of sources. Can make a distinction between formal and informal language with occasional less appropriate expressions.	Can write clear, well-structured and mostly accurate texts of complex subjects. Can underline the relevant salient issues, expand and support points of view at some length with subsidiary points, reasons and relevant examples, and round off with an appropriate conclusion.	Can write clear, highly accurate and smoothly flowing complex texts in an appropriate and effective personal style conveying finer shades of meaning. Can use a logical structure which helps the reader to find significant points.		

A5: Sample of students' written texts

Semester 2 Final Writing Exam – Sci/Med (2014/2015)

45

"My University Life"

I study at King Saud University. I'm medical student. My study is very difficult so I work hard. My subjects are: English, Biostatistic, Chemistry, IT, Math and Biology.

Before university my daily routine is a bit full, but now it's very bad. I go to university at 6:00 a.m., but some days I go to university at 10:00 a.m. In 3 days in a week I have 4 classes and in 2 days in a week I have 2 classes. Sometimes I have late classes, I eat my breakfast at 11:00 a.m., then I don't eat anything because I don't have any time. Some days I have full day. I'm back to my house at 3:00 p.m.

The studying is important for everyone. We need studying because we need a good career in the future. We need a good life to live.

431 Semester 2 Final Writing Exam – Sci/Med (2014/2015)

Write at least 120 words for the writing task.

Make sure you write about all the parts of the writing task.

I study at King Saud University many subjects that will benefit me later in life. These subjects include: Biology, Biostatistics, Physics, English, and Chemistry. Each subject is very important and beneficial in life. For example, studying English in preparatory year is crucial for a student who wants to go into medical school; because all the subjects are in English. Biology is also very important for me to learn, because it teaches me the basic science of organisms.

My life at King Saud University includes many activities apart from learning beneficial subjects. It includes a daily routine that I never get bored from. First, I attend English class in the morning. Second, I go to a café with my friends to buy coffee and relax. Then, I attend the rest of my classes until 1:20 P.M. After that my friends and I go to pray. Finally, I attend my last class before going home. Some people may find my routine quite boring, but I love it!

Studying at university is actually very important for many reasons. The first reason is that it raises a person's educational level. Second, it helps in getting a better job in the future. Last but not least, studying university improves a person's social and academic skills.

A6: Screenshots of the email and rating form sent to raters

Dear...

I would like to thank you for showing an interest in participating in a pilot study of a small-scale data collection process that is necessary for my research. Your participation will help me select the samples that I will use for my training session with other groups of English teachers, which will help maintain the reliability of my data collection.

Again, thank you so much; your participation is very much appreciated!
I have attached to your email the following:

- 20 scanned written texts from different proficiency levels.
- The rating grid that you are kindly asked to use when rating the texts (please note that there are two formats of the same rating scale; please use the format that you prefer the most, and let me know which one should be used for my actual data collection).
- An excel file, in which you are kindly asked to fill in your ratings of each text.

I need your help to go through each text, one at a time. I would ask that you use the attached rating scale (based on the *Common European Framework of references CEFR*) to rate each text, and fill in the attached excel grid with your evaluation of each text. Each text has a number on the top of the page that has an equivalent number in the excel file. Please match the number on the top of the page with the one in the excel file when you complete your evaluation.

You also can choose from the plus levels if you feel that a participant is higher than a specific level but is still not high enough to reach the level above, and provide your justifications in the comments column.

In the rating scale, there are five categories (*overall, range, accuracy, coherence description*). You are kindly asked to rate each text in these categories, assigning a CEFR level to the text based on the students' performance.

Then, you are asked to give an overall judgement of which level you generally think the text should be at (based on the CEFR) with your comments and justifications of why you have reached this decision (in the comment column).

Kindly note that there is a drop menu in the excel sheet where you can select the level assigned to the text.

At the end, kindly return to me the completed excel file with your evaluation, comments and feedback.

I would really appreciate it a lot if you can send your evaluation by Friday, 22 September, 2017 or before (if possible).

Many thanks again for all your help and support!

Best regards,
Ebtesam

357

A7: Phase I Focus Group Guide

Questioning Route				
<p>Opening Questions: The researcher will ask some ice-breaking questions at the beginning to make participants, as mentioned by Krueger & Casey (2000), feel more comfortable and to encourage them to participate. These questions should require short and simple answers, and should not take a lot of time.</p>				
Students	Tutors	CU	AU/CAU	PDU
5 Min.	5 Min.	5 Min.	5 Min.	5 Min.
Tell us who you are? And which college you are interested in joining after leaving the PYP?	Tell us who you are? And how long have you been working at the PYP? And how long have you been teaching medical track students?	Could you briefly introduce yourself & tell me how long have you been working at PYP and how long have you been working on the Curriculum, assessment and Continuous assessment/PD?		
<p>Introductory Questions: These questions are used to introduce participants to the main topic of the focus group meeting and “get people to start thinking about their connection with the topic” (ibid., 2000: p. 44)</p>				
Students	Tutors	CU	AU/CAU	PDU
8 Min.	8 Min.	8 Min.	8 Min.	8 Min.
As a student in the medical track, how do you feel about your English writing? (Please explain/elaborate)	How do you feel about your students’ writing? (Please explain/elaborate)	How do you feel about what is offered to the medical track students in the PYP curriculum regarding English writing?	How do you feel about what is being offered in terms of assessment for PYP medical students regarding their English writing?	How do you feel about what is being offered in the classroom with regard to writing for the medical track students?
<p>Transition Questions: These questions “move the conversation into the key questions that drive the study. They serve as the logical link between the introductory questions and the key questions” (ibid., 2000: p. 45)</p>				
	Tutors	CU	AU/CAU	PDU

Students				
9 Min.	9 Min.	10 Min.	10 Min.	10 Min.
Tell me about your experience at the PYP learning English writing skills? What did you do to learn and improve your English writing?	Tell me about your experience teaching writing skills to medical students at the PYP. What do you usually do? How do you teach?	Tell me about your experience when designing the curriculum for writing to medical students? What do you usually do to set goals, objectives & learning outcomes? What is the role of CEFR when doing this?	Tell me about your experience when designing assessment/CA writing tasks. What do you usually do? Can you tell me about the process to design the task?	Tell me about your experience developing teachers in terms of teaching writing to medical students? How do you usually improve teachers' teaching of writing?
<p>Key Questions: These are considered the main questions that serve the main purpose of the study. There will be three questions/activities in this section.</p>				
Students	Teachers	CU	AU/CAU	PDU
First Activity/Question	First Activity/Question	First Activity/Question	First Activity/Question	First Activity/Question
25 Mins		30 Mins		
Using the writing descriptors (cut into six levels) of the common reference levels of self-assessment grid (CEFR: 26-27), the students/tutors will identify the descriptors that they believe the students can do and stick them on the left column of the blank paper; they will be asked to give explanations and justifications for their choices. They will stick the descriptors that they are not sure they are good at on the other column and they will be asked to give explanations and justifications. Then, they will highlight what they believe is		On a coloured piece of paper, please list the writing activities, tasks or skills your unit (CU, AU, CAU, PDU) is focusing on, and then, on another piece of paper (different colour), list what other writing skills you think your unit needs to add/deal with, especially for the MT students. Please discuss both lists.		

important from the two columns. Then, they will be asked to add what they think is missing and what they can do/need to be able to do in writing to be added to the framework that is missing from these descriptors.				
Second Activity/Question	Second Activity/Question	Second Activity/Question	Second Activity/Question	Second Activity/Question
45 Mins		40 Mins		
<p>Participants in the focus group will be divided into groups of twos or threes (depending on the number of participants in each group). They will be provided with cards (see appendix A8 for an example). Each card has one of the categories in the CEFR scales used for self- and tutor assessment (i.e. <i>Overall written production, interactions, what students can write, vocabulary range and control...etc.</i>). Participants were asked to talk and reflect on those skills and about what students can/cannot do, the difficulties they faced, and to specify the limitations & constraining factors that prevent students from mastering these skills. The following are further questions prepared to encourage more discussion:</p> <ul style="list-style-type: none"> • What can you (the students) write? • What are the writing skills you (the students) were involved in during the PYP? • What are the types of texts the students were asked to write? • Can you talk about grammar/spelling/vocabulary/use of punctuation at the PYP, and what you (the students) feel you can do? • Can you talk about writing reports and essays at the PYP, what you have learned and what you can do? • What are the skills in writing that are considered challenging to you (to the students)? • Are you satisfied with your current proficiency in writing after the PYP? 				
Third Activity/Question	Third Activity/Question	Third Activity/Question	Third Activity/Question	Third Activity/Question
20 Mins		19 Mins.		
How well do you feel (your students were) prepared in terms of your (their) writing? Are you satisfied with your (your students') current proficiency in writing by the end of the PYP? How has PYP changed your (your students') writing skills? Is there anything that could be done to improve your (the students') writing?		How well you feel your unit has contributed to the preparation of the students in terms of their writing? Is there anything you feel needs to be done to improve the curriculum/assessment/teaching of writing in terms of writing to MT students? Are there any constraining factors/issues that stop the development of writing in the PYP?		

Are there any constraining factors that mitigate against improving the students' writing?				
<p>Closing Questions: "These questions bring closure to the discussion. They allow participants to reflect on their comments, and are critical to analysis" (ibid., 2000: p. 45)</p>				
Students	Teachers	CU	AU/CAU	PDU
8 Mins				
<p>Of all that we have discussed, what is the most important point to you? What do you think about the importance of writing for MT students? Then, the researcher gives an overview of the purpose of the study and summary of the points mentioned in the meeting, before asking if anything has been missed, or if there is anything else that others wish to add.</p>				

CU= curriculum unit; AU= Assessment Unit; CAU= Continuous Assessment Unit; PDU= Professional Development Unit

A8: Sample of Cards used in FG activities

Overall written production الإنتاج الكتابي بشكل عام	Overall written interaction التكامل الكتابية بشكل عام
<p>ان يكون لها القدرة على :</p> <p>(1) نقل مضمون ما قامت بكتابته</p> <p>(2) تهيئة البدء في موضوع معين بأسلوبها الخاص</p> <p>(3) ان تكون المقدمة ملفتة وملحة للموضوع الذي سوف يتكلم اليه القارئ ولتستطيع تطبيقه</p> <p>(4) معرفة كيفية استخدام كلمات الترتيب</p> <p>(5) ابتداء الموضوع المناسب</p> <p>(6) كيفية انهاء ما تم كتابته</p> <p>(7) البدء بالمفهم فالأهم</p> <p>(8) كيفية ابتداء الجملة واختتامها بشكل صحيح لا يدخل بالمعنى</p>	<p>ان يكون لها القدرة على :</p> <p>(1) كتابة أي مقالة في اللغة الإنجليزية</p> <p>(2) كيفية صياغة المواضيع الكتابية من واقع التواصل الاجتماعي</p> <p>(3) المساهمة والافتتاح واعطاء البراهين</p> <p>(4) كيفية كتابة استبيان</p>

Group 2

Conditions and limitations الشروط والقيود عند الكتابة
<p>لا</p> <p>- ان لا يكون لدى الطلبة الكافية من الشرائح لهذا الموضوع .</p> <p>- عدم القدرة على الكتابة على موضوع يتقاعده صديقه .</p> <p>- عدم القدرة على كتابه مقدمه و خاتمه مترابله غير مرتبط مع الموضوع .</p> <p>- ان يكون المرفقة كافي</p>

A9: Phase Two Focus Group Guide

Opening Questions:	
Students	Faculty Staff
5 Minutes	
<ul style="list-style-type: none"> • Can you please introduce yourself? • In general, how did you find your (your students') experience of their first year at college, after PYP? 	
Introductory Questions:	
8 Minutes	
<ul style="list-style-type: none"> • What are the courses you (your students) have studied during this year at your college? • In these courses, do you (your students) need to write in English? • Do you (your students) need to write in English to answer exam questions? Please explain. • Do you (your students) have to submit assignments in English? Please explain. 	
Transition Questions:	
8 Minutes	
<p>How often do you (your students) write in English? And for what purposes? Are you satisfied with your (your students') level of English writing? Please explain.</p> <p>Do you think you (your students) have the writing skills required to be able to cope with the different demands of your course as a first-year student? Please explain.</p> <p>Are you (your students) having difficulties in your college because of your English writing? Please explain.</p>	
Key Questions:	
<p>In this section, there will be three activities for the discussion group. The purpose of the first one is to identify when the students are required to write, what they should write about, and for whom they are required to write, where writing appears in the curriculum and why they use writing.</p> <p>In the second activity (CEFR-lead approach), the researcher needs to identify the minimum writing requirements in each of the CEFR categories (these may appear in the CEFR descriptors or not), and how well the students were prepared in relation to each of those categories.</p>	
60 Minutes	
Activity ONE:	
<p>For students (FGs), using a blank sheet of A3 paper, participants (in pairs or threes) will be asked to reflect on this academic year at the MHCCs and list/brainstorm in response to the following questions (the same questions are to be asked in staff interviews):</p> <ul style="list-style-type: none"> • What are the topics you were asked (you asked your students) to write about? • What are the situations where you were asked (you asked students) to write in English and why (i.e. when do you usually ask your students to write in English?) 	

- What types of writing were you asked to employ (you asked your students to write)?

What are the writing skills, activities and tasks you were asked (you asked your students) to be involved in during this year?

Activity TWO:

Participants were provided with an envelope, which has the CEFR general writing descriptors (taken from the self-assessment grid, P: 26-27). Each descriptor is cut separately in a box. Participants will be asked to choose only one descriptor that they believe reflect the minimum writing requirement during their first year at college. Then they should discuss among themselves why they choose that descriptor.

Activity THREE:

In this activity, participants will be provided with cards which contain the categories of the CEFR used in the study (see the attachment below for an example). Participants will be asked to read each category and discuss the minimum writing needs in each of these categories. Participants will also be invited to talk about how well the PYP has prepared them to satisfy those needs, and to offer their views and comments.

Closing Questions:

14 Minutes

Students' Questions:

- How useful were the English writing classes at the PYP? Please explain.
- To what extent was the writing at the PYP relevant to the writing you needed at your college? Please explain.
- Was the time allocated to the English writing training at the PYP enough to use it effectively at your college? How? Why?
- If you had the opportunity to modify the PYP writing course, what changes would you make and why?
- What do you need most to help you improve your writing to reach your college requirements?

Staff Qs:

- Do you think that the PYP has prepared the students for the writing demands at your college? How?
- What do you need the most from PYP to help improve your students' writing to reach your college requirements?

The researcher, at the end, summarizes the main points discussed throughout the discussion and then asks the following:

- Have I missed anything? Or is there anything I misunderstood?
- Of all that we have discussed today, what is the most important point to you?
- Is there anything you want to add?

A11: Sample of Ethical approval letters

جامعة الملك سعود (034)
هاتف +966 11 4694006
فكس +966 11 4694555

المملكة العربية السعودية
ص.ب 2454 الرياض 11451
www.ksu.edu.sa

جامعة
الملك سعود
King Saud University



السنة التحضيرية

المكرمة/ إبتسام عبدالعليم عبدالحليم
محاضر قسم اللغة الإنجليزية بعمادة السنة التحضيرية
السلام عليكم ورحمة الله وبركاته
إشارةً إلى طلبكم الموافقة على إجراء بحث (المعايير التعليمية الأكاديمية الأساسية لإتقان
الكتابة الأكاديمية باللغة الإنجليزية ضمن الإطار الأوربي المرجعي المشترك للغات) بجامعة
Warwick بالمملكة المتحدة .
أفيدكم أن طلبكم تم مناقشته بإجتماع لجنة دعم الأبحاث العلمية بالعمادة و تم الموافقة
على إجراء البحث بمرحلتيه الأولى و الثانية .

و تقبلوا وافر التقدير

وكيلة عمادة السنة التحضيرية للشؤون الإدارية
ورئيسة لجنة دعم الأبحاث العلمية قسم الطالبات

د.سيناء بنت عبدالمحسن العقيل

ع.القطاني

٢/٤/١٤٣٧ هـ
١٤٣٦ / ٦ / ٤ هـ

A12: Information sheet

Information Sheet



Setting Minimum Academic Educational Writing Standards for Medical Students Benchmarked against the *Common European Framework of Reference for Languages: CEFR*

Researcher:

Ebtesam Abdulhaleem, PhD student at the Centre for Applied Linguistics, University of Warwick.

Participation and purpose of the study:

As a tutor/student in the medical track (MT) of the Preparatory Year, you are invited to participate in a study that seeks to identify PYP MT students' writing levels by assessing students using a questionnaire and follow-up interviews/focus-group meetings aligned to the *Common European Framework of Reference for Languages (CEFR)*. This research is part of my PhD project that I am undertaking at the University of Warwick. The purpose of the study is to identify the minimal standards medical/healthcare college students need to acquire before joining these colleges, and how well PYP students are prepared for these standards before leaving the PYP.

What your participation will involve:

The study consists of different stages, as outlined below in the table. If you are willing to participate, you will be invited to complete a questionnaire. You have the right to participate in one, or all, of the stages of the study. You do not have to commit yourself now to all stages of the study. I will be inviting you separately to each stage of the study, and you can then decide which parts you would like to participate in. You have the right to withdraw at any time without being subjected to any negative consequences.

The following table gives a general overview of the different steps of the study:

End of Academic Year 2014-2015 At the PYP	End of Academic Year 2015-2016 At MHCCs
Students	
CEFR-based self-assessment questionnaire Two-hour focus group meetings Student's final writing score	CEFR-based needs analysis questionnaire One-hour focus group meetings
PY Tutors	
CEFR-based tutor-assessment questionnaire Two-hour focus group meetings	
PY Coordinators	
Two-hour focus group meetings	
MHCC Teaching Staff	
	CEFR-based needs analysis questionnaire Interviews (30 Minutes)

There will be a prize draw for all questionnaires, focus-group meetings and interview participants.

Confidentiality and disclosure of information:

All data obtained will be anonymized. Any information that is obtained in connection with this study will remain absolutely confidential and will be used solely for research purposes. In my thesis publication or any other research reports or publications, information will be provided in such a way that no participating tutor can be identified by name or ID.

Benefits for participants:

By participating in this study, you can benefit by being part of an important study that is of relevance to your career/study. It will also benefit you by raising your awareness of the students' writing proficiency levels, and the difficulties and/or needs they may experience when joining the university MHCCs. At the end of the study, all participants will be welcome to consult the research findings.

Contact:

Should you have any questions related to the study, you can contact me at: E.Abdulhaleem@warwick.ac.uk. Should you have any complaints relating to this study, you are advised to contact the Deputy Registrar (see contact details below). You are also advised to see the [University's Complaints and Feedback Procedure](http://www2.warwick.ac.uk/services/gov/complaintsandfeedback/) at <http://www2.warwick.ac.uk/services/gov/complaintsandfeedback/>.

<p><u>Jo Horsburgh</u> Deputy Registrar Deputy Registrar's Office University of Warwick Coventry CV4 8UW Email: J.Horsburgh@warwick.ac.uk</p>	<p>PA – Natasha Lynch Tel: 024 765 22706 Email: n.lynch@warwick.ac.uk</p>
--	---

A13: Consent form for participation

THE UNIVERSITY OF
WARWICK

CONSENT FORM

Project Title: Setting Minimum Academic Educational Writing Standards for Medical Students Benchmarked against the *Common European Framework of Reference for Languages: CEFR*

Name of Researcher: Ebtesam Abdulhaleem

Participant consent:

I confirm that I have read and understood the information sheet for the above project.

I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason and without being penalised or disadvantaged in any way.

I understand that any information that is obtained in connection with this study will remain absolutely confidential and will be used solely for research purposes.

I confirm that I am willing to be contacted and invited to participate in different phases of the research.

I confirm that I have had the opportunity to ask any questions, and that I may keep the information sheet for my records.

Name of Participant Date Signature

Name of person taking Date Signature
consent if different
from researcher

Researcher Date Signature

A14: consent form for recordings

PARTICIPATION AND RECORDING CONSENT FORM

Project Title: Setting Minimum Academic Educational Writing Standards for Medical Students Benchmarked against the *Common European Framework of Reference for Languages: CEFR*

Name of Researcher: Ebtesam Abdulhaleem
Name of Supervisor: Prof Claudia Harsch & Dr Neil Murray

Participant consent:

I confirm that I have read and understood the Information Sheet for the above project and I agree to take part in this study. I understand that my participation is voluntary and that I am free to withdraw at any time without being penalised or disadvantaged in any way. I understand that any information that is obtained in connection with this study will remain absolutely confidential and will be used solely for research purposes. I confirm that I have had the opportunity to ask any questions, and that I may keep the Information Sheet for my records.

As part of this study, we would like to audio record your participation in the focus-group meetings and use it for the research purposes.

Please indicate your consent to what is mentioned in the following table:

I give my consent to:	Yes	No
Audio record of my participations with complete anonymity.		
A written transcript of the audio recordings.		
The use of data for academic and professional publications with complete anonymity.		
The use of extracts from the data for different research purposes.		
The use of extracts from the data for training and assessment purposes with complete anonymity.		

I have read the above descriptions and give my consent to the use of the data as indicated above.

Name: _____ (Please Print)

Signature _____

Email: _____ Date: _____

Contact Person: Ebtesam Abdulhaleem, Centre for Applied Linguistics, University of Warwick, Coventry CV4 8AL, UK.
Email: abdulhaleem@warwick.ac

A15: An individual report summarising each student's proficiency level

The following table has general global definitions and descriptions of the different CEFR levels. Refer back to the email to identify your level and then read the following short descriptions about it.

Please note that **A1** is considered the lowest level in the CEFR scale and **C2** is the highest. It is also important to remember that the result you have received was solely **based on your own judgment** while using the self-assessment grid.

	CEFR Levels	Global scale	Writing	Description of the levels
Proficient User	C2	Can understand with ease everything heard or read. Can summarize information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situation.	Can write letters on any subject and full notes of meetings or seminars with good expression and accuracy.	Level C2 refers to language users with near native-speaker competence. Users in this level are characterized with a degree of precision, appropriateness and ease with the language.
	C1	Can understand a wide range of demanding, longer texts, and recognize implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organizational patterns, connectors and cohesive devices.	Can prepare/draft professional correspondence, take reasonably accurate notes in meetings or write an essay, which shows the ability to communicate.	Level C1 refers to language users with a broad range of language use, which allows fluent, spontaneous communication.

Independent user	B2+	Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialization. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for their party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.	Can make notes while someone is talking or write a letter including non-standard request.	Level B2+ refers to language users similar to the ones in level B2. However, the focus on argument and social discourse can also be interpreted as a new focus on discourse skills. These skills will be clear in more conversational management, more advanced coherence and cohesion and the ability to use more negotiating language e.g. using persuasive language.
	B2			Level B2 refers to language users who have more control on their social discourse e.g. converse naturally, fluently and effectively and have more effective language awareness.
	B1+	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school leisure, etc. can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics, which are familiar, or of personal interest. Can describe experiences and events, dreams, hopes and ambitions and briefly give reasons and explanations for opinions and plans.	Can write letters or make notes on familiar or predictable matters.	Level B1+ refers to language users who can maintain interaction and get across what they want to in a range of contexts. In addition, users have the ability to exchange more information, for example, explaining problems and summarizing their opinion about an article, documentary ...etc.
	B1			Level B1 refers to language users who can maintain interaction and get across what they want to in a range of contexts.

Basic User	A2+	Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.	Can complete forms and write short simple letters or postcards related to personal information.	Level A2+ refers to more active participation in the use of the foreign language with some need of assistance and with certain limitations.
	A2			Level A2 is the level where language users start to use more social functions and getting out and about like using simple everyday forms, simple transactions in shops ...etc.

A16: Sample of students' letter to a friend

Dear Friend,

I'm writing you this letter in hope that you'll benefit from my words. PY is an essential year to prepare for the upcoming years in university, so whatever you learn here will sure remain in your mind till you graduate, especially writing.

The university has an amazing program in developing the girls' skills in writing. Personally I think lower levels' students benefited more from it than the highest level, because they worked hard to learn unlike level 3 students who saw it so easily that they didn't practice much. The CA project in writing helped me alot! I learned to look up for the subject online, read about it and gain information, and paraphrasing it in my own words without anyone's help. My advice for you my friend is to work hard and practice more at home even when you already know how to, because nothing will be stuck in your mind for the long term more than practicing. I hope you'll have a great year in PY with great writing skills.

Your loving friend,

Appendix B

Table B1 Post-Hoc Tukey HSD of Students' Self-assessment of their CEFR Levels across the three PYP Levels

Dependent Variable	(I) PYP Levels	(J) PYP Levels	Mean Difference (I-J)	Std. Error	P-value	95% Confidence Interval	
						Lower Bound	Upper Bound
What Students Can Write	Elementary	Intermediate	-0.48	0.29	0.22	-1.15	0.19
		Advanced	-2.40*	0.30	<0.001	-3.10	-1.69
	Intermediate	Elementary	0.48	0.29	0.22	-0.19	1.15
		Advanced	-1.92*	0.21	<0.001	-2.41	-1.43
	Advanced	Elementary	2.40*	0.30	<0.001	1.69	3.10
		Intermediate	1.92*	0.21	<0.001	1.43	2.41
Reports and Essays	Elementary	Intermediate	-0.36	0.31	0.46	-1.08	0.35
		Advanced	-2.61*	0.32	<0.001	-3.37	-1.86
	Intermediate	Elementary	0.36	0.31	0.46	-0.35	1.08
		Advanced	-2.25*	0.22	<0.001	-2.77	-1.73
	Advanced	Elementary	2.61*	0.32	<0.001	1.86	3.37
		Intermediate	2.25*	0.22	<0.001	1.73	2.77
Note Taking	Elementary	Intermediate	-0.22	0.30	0.74	-0.93	0.49
		Advanced	-1.72*	0.32	<0.001	-2.47	-0.98
	Intermediate	Elementary	0.22	0.30	0.74	-0.49	0.93
		Advanced	-1.50*	0.22	<0.001	-2.02	-0.98
	Advanced	Elementary	1.72*	0.32	<0.001	0.98	2.47
		Intermediate	1.50*	0.22	<0.001	0.98	2.02
Average levels	Elementary	Intermediate	-0.44	0.20	0.067	-0.91	0.02
		Advanced	-2.25*	0.21	<0.001	-2.74	-1.76
	Intermediate	Elementary	0.44	0.20	0.067	-0.02	0.91
		Advanced	-1.82*	0.15	<0.001	-2.15	-1.46
	Advanced	Elementary	2.25*	0.21	<0.001	1.76	2.74
		Intermediate	1.82*	0.15	<0.001	1.46	2.15

Bold with *=significant results

Table B2 Post-Hoc Tamhane of Students' Self-assessment of their CEFR Levels across the three PYP Levels

Dependent Variable	(I) PYP Levels	(J) PYP Levels	Mean Difference (I-J)	Std. Error	P-value	95% Confidence Interval	
						Lower Bound	Upper Bound
Overall Written Production	Elementary	Intermediate	-0.67	0.31	0.092	-1.41	-1.41
		Advanced	-2.34*	0.30	<0.001	-3.08	-3.08
	Intermediate	Elementary	0.67	0.31	0.092	-0.08	-0.08
		Advanced	-1.67*	0.18	<0.001	-2.11	-2.11
	Advanced	Elementary	2.34*	0.30	<0.001	1.60	1.60
		Intermediate	1.67*	0.18	<0.001	1.23	1.23
Overall Written Interaction	Elementary	Intermediate	-0.30	0.28	0.66	-0.98	-0.98
		Advanced	-2.74*	0.31	<0.001	-3.50	-3.50
	Intermediate	Elementary	0.30	0.28	0.66	-0.39	-0.39
		Advanced	-2.44*	0.24	<0.001	-3.01	-3.01
	Advanced	Elementary	2.74*	0.31	<0.001	1.98	1.98
		Intermediate	2.44*	0.24	<0.001	1.87	1.87
Types of Texts the Students can write	Elementary	Intermediate	-0.33	0.28	0.55	-1.00	-1.00
		Advanced	-2.33*	0.31	<0.001	-3.07	-3.07
	Intermediate	Elementary	0.33	0.28	0.55	-0.34	-0.34
		Advanced	-2.00*	0.23	<0.001	-2.55	-2.55
	Advanced	Elementary	2.33*	0.31	<0.001	1.59	1.59
		Intermediate	2.00*	0.23	<0.001	1.44	1.44
Vocabulary Range & Control	Elementary	Intermediate	-0.40	0.26	0.34	-1.04	-1.04
		Advanced	-2.30*	0.30	<0.001	-3.01	-3.01
	Intermediate	Elementary	0.40	0.26	0.34	-0.24	-0.24
		Advanced	-1.90*	0.22	<0.001	-2.42	-2.42
	Advanced	Elementary	2.30*	0.30	<0.001	1.59	1.59
		Intermediate	1.90*	0.22	<0.001	1.38	1.38
Grammatical Accuracy	Elementary	Intermediate	-0.77	0.35	0.083	-1.61	-1.61
		Advanced	-1.81*	0.38	<0.001	-2.73	-2.73
	Intermediate	Elementary	0.77	0.35	0.083	-0.07	-0.07
		Advanced	-1.04*	0.26	<0.001	-1.67	-1.67
	Advanced	Elementary	1.81*	0.38	<0.001	0.89	0.89
		Intermediate	1.04*	0.26	<0.001	0.41	0.41
Orthographic Control	Elementary	Intermediate	-0.36	0.36	0.70	-1.24	-1.24
		Advanced	-1.93*	0.36	<0.001	-2.81	-2.81
	Intermediate	Elementary	0.36	0.36	0.70	-0.52	-0.52
		Advanced	-1.57*	0.23	<0.001	-2.12	-2.12
	Advanced	Elementary	1.93*	0.36	<0.001	1.05	1.05
		Intermediate	1.57*	0.23	<0.001	1.02	1.02
Processing Texts	Elementary	Intermediate	-0.58*	0.21	0.020	-1.09	-1.09
		Advanced	-2.31*	0.25	<0.001	-2.91	-2.91
	Intermediate	Elementary	0.58*	0.21	0.020	0.07	0.07
		Advanced	-1.73*	0.20	<0.001	-2.21	-2.21
	Advanced	Elementary	2.31*	0.25	<0.001	1.72	1.72
		Intermediate	1.73*	0.20	<0.001	1.25	1.25

Bold with *=significant results

Table B3 Tukey HSD of Tutors' Assessment across the three PYP levels

Dependent Variable	(I) PYP Levels	(J) PYP Levels	Std. Error	P-value	95% Confidence Interval
--------------------	----------------	----------------	------------	---------	-------------------------

			Mean Difference (I-J)			Lower Bound	Upper Bound
Overall Written Production	Elementary	Intermediate	-1.61*	0.24	<0.001	-2.18	-1.04
		Advanced	-3.18*	0.25	<0.001	-3.77	-2.59
	Intermediate	Elementary	1.61*	0.24	<0.001	1.04	2.18
		Advanced	-1.57*	0.19	<0.001	-2.01	-1.12
	Advanced	Elementary	3.18*	0.25	<0.001	2.59	3.77
		Intermediate	1.57*	0.19	<0.001	1.12	2.01
Types of Texts Students can write	Elementary	Intermediate	-1.35*	0.26	<0.001	-1.95	-0.74
		Advanced	-2.82*	0.27	<0.001	-3.45	-2.20
	Intermediate	Elementary	1.35*	0.26	<0.001	0.74	1.95
		Advanced	-1.48*	0.20	<0.001	-1.95	-1.01
	Advanced	Elementary	2.82*	0.27	<0.001	2.20	3.45
		Intermediate	1.48*	0.20	<0.001	1.01	1.95
Reports and Essays	Elementary	Intermediate	-1.22*	0.31	<0.001	-1.94	-0.50
		Advanced	-2.21*	0.31	<0.001	-2.94	-1.47
	Intermediate	Elementary	1.22*	0.31	<0.001	0.50	1.94
		Advanced	-0.99*	0.21	<0.001	-1.48	-0.49
	Advanced	Elementary	2.21*	0.31	<0.001	1.47	2.94
		Intermediate	0.99*	0.21	<0.001	0.49	1.48
Note Taking	Elementary	Intermediate	-1.10*	0.33	<0.001	-1.88	-0.31
		Advanced	-2.15*	0.34	<0.001	-2.95	-1.35
	Intermediate	Elementary	1.10*	0.33	<0.001	0.31	1.88
		Advanced	-1.05*	0.23	<0.001	-1.59	-0.51
	Advanced	Elementary	2.15*	0.34	<0.001	1.35	2.95
		Intermediate	1.05*	0.23	<0.001	0.51	1.59
Average of all scales	Elementary	Intermediate	-1.33*	0.19	<0.001	-1.78	-0.88
		Advanced	-2.86*	0.20	<0.001	-3.32	-2.39
	Intermediate	Elementary	1.33*	0.19	<0.001	0.88	1.78
		Advanced	-1.53*	0.15	<0.001	-1.87	-1.17
	Advanced	Elementary	2.86*	0.20	<0.001	2.39	3.32
		Intermediate	1.53*	0.15	<0.001	1.17	1.88

Bold with *=significant results

Table B4 Post-Hoc Tamhane of Tutors' Assessment across the three PYP levels

Dependent Variable	(III) (IV)	PYP Levels	(J) PYP Levels	Mean Difference (I-J)	Std. Error	P-value	95% Confidence Interval	
							Lower Bound	Lower Bound
Overall Written Interaction	Elementary	Intermediate		-1.51*	0.25	<0.001	-2.12	-0.89
		Advanced		-2.76*	0.25	<0.001	-3.37	-2.15
	Intermediate	Elementary		1.51*	0.25	<0.001	.89	2.12
		Advanced		-1.26*	0.19	<0.001	-1.71	-0.80
	Advanced	Elementary		2.76*	0.25	<0.001	2.15	3.37
		Intermediate		1.26*	0.19	<0.001	.80	1.71
What students Can Write	Elementary	Intermediate		-1.46*	0.23	<0.001	-2.01	-0.91
		Advanced		-2.99*	0.25	<0.001	-3.58	-2.40
	Intermediate	Elementary		1.46*	0.23	<0.001	.91	2.01
		Advanced		-1.54*	0.19	<0.001	-1.99	-1.08
	Advanced	Elementary		2.99*	0.25	<0.001	2.40	3.58
		Intermediate		1.54*	0.19	<0.001	1.08	1.99
Vocabulary Range and Control	Elementary	Intermediate		-1.16*	0.20	<0.001	-1.65	-0.67
		Advanced		-2.52*	0.23	<0.001	-3.06	-1.95
	Intermediate	Elementary		1.16*	0.20	<0.001	.67	1.65
		Advanced		-1.35*	0.20	<0.001	-1.82	-0.88
	Advanced	Elementary		2.51*	0.23	<0.001	1.95	3.06
		Intermediate						

Grammatical Accuracy	Elementary	Intermediate	1.35*	0.20	<0.001	.88	1.82
		Intermediate	-1.11*	0.23	<0.001	-1.66	-.56
		Advanced	-2.29*	0.26	<0.001	-2.91	-1.67
	Intermediate	Elementary	1.11*	0.23	<0.001	.56	1.66
		Advanced	-1.18*	0.20	<0.001	-1.65	-.71
		Advanced	2.29*	0.26	<0.001	1.67	2.91
Orthographic Control	Elementary	Intermediate	1.18*	0.20	<0.001	.71	1.65
		Intermediate	-0.67	0.34	0.147	-1.50	.16
		Advanced	-2.75*	0.35	<0.001	-3.59	-1.91
	Intermediate	Elementary	0.67	0.34	0.147	-.16	1.50
		Advanced	-2.08*	0.18	<0.001	-2.51	-1.65
		Advanced	2.75*	0.35	<0.001	1.91	3.59
Processing Texts	Elementary	Intermediate	2.08*	0.18	<0.001	1.65	2.51
		Intermediate	-1.51*	0.25	<0.001	-1.43	-.59
		Advanced	-2.76*	0.25	<0.001	-3.63	-2.54
	Intermediate	Elementary	1.51*	0.25	<0.001	.59	1.43
		Advanced	-1.26*	0.19	<0.001	-2.54	-1.60
		Advanced	2.76*	0.25	<0.001	2.54	3.63
	Advanced	Intermediate	1.256*	0.19	<0.001	1.60	2.54

Bold with *=significant results

Table B5 Post Hoc Tukey analysis of range, coherence, accuracy, description, and overall grouped by PYP levels

Dependent Variable	(I) PYP Levels	(J) PYP Levels	Mean Difference (I-J)	Std. Error	P-value	95% Confidence Interval	
						Lower Bound	Upper Bound
Range	Elementary	Intermediate	-0.33	0.25	0.40	-0.93	0.27
		Advanced	-1.53	0.27	<0.001	-2.15	-0.89
	Intermediate	Elementary	0.33	0.25	0.4	-0.27	0.93
		Advanced	-1.19	0.18	<0.001	-1.61	-0.77
	Advanced	Elementary	1.53	0.27	<0.001	0.89	2.15
		Intermediate	1.19	0.18	<0.001	0.77	1.61
Coherence	Elementary	Intermediate	-0.42	0.25	0.227	-1.01	0.18
		Advanced	-1.33	0.26	<0.001	-1.97	-0.70
	Intermediate	Elementary	0.42	0.25	0.227	-0.18	1.01
		Advanced	-0.92	0.18	<0.001	-1.33	-0.50
	Advanced	Elementary	1.33	0.26	<0.001	0.70	1.96
		Intermediate	0.92	0.18	<0.001	0.50	1.33
Accuracy	Elementary	Intermediate	-0.20	0.24	0.674	-0.78	0.37
		Advanced	-1.40	0.25	<0.001	-2.00	-0.79
	Intermediate	Elementary	0.20	0.24	0.674	-0.37	0.78
		Advanced	-1.19	0.17	<0.001	-1.60	-0.79
	Advanced	Elementary	1.40	0.25	<0.001	0.79	2.01
		Intermediate	1.19	0.16	<0.001	0.79	1.59
Description	Elementary	Intermediate	-0.27	0.24	.0507	-0.85	0.31
		Advanced	-1.34	0.26	<0.001	-1.95	-0.74
	Intermediate	Elementary	0.27	0.24	0.507	-0.31	0.85

		Advanced	-1.07	0.17	<0.001	-1.48	-0.67
	Advanced	Elementary	1.34	0.26	<0.001	0.74	1.95
		Intermediate	1.07	0.17	<0.001	0.67	1.48
Overall	Elementary	Intermediate	-0.31	0.25	0.423	-0.90	0.28
		Advanced	-1.44	0.26	<0.001	-2.06	-0.81
	Intermediate	Elementary	0.31	0.25	0.423	-0.28	0.90
		Advanced	-1.12	0.17	<0.001	-1.54	-0.71
	Advanced	Elementary	1.44	0.26	<0.001	0.81	2.06
		Intermediate	1.12	0.17	<0.001	0.71	1.54

Bold with *=significant results

Table B6 Descriptive statistics and Kolmogorov Smirnov Test for normality of average students' self, tutors' and raters' ratings of the student's level

Type	Normal Parameters ^{a,b}		Test Statistic	p-value
	M	SD		
Self-	5.45	1.63	0.07	.200 ^{c,d}
Teachers	5.49	1.73	0.08	.095 ^c
Raters	4.18	0.96	0.07	.200 ^{c,d}

a. Test distribution is Normal, b. Calculated from data, c. Lilliefors Significance Correction, d. This is a lower bound of the true significance.

M=Mean, SD=Standard deviation

Table B7 Tukeys Post Hoc analysis for ratings grouped as to the Type of raters

(I) Type	(J) Type					
	Students self-assessment		Teachers' assessment		Raters' assessment	
	Mean Difference (I-J)	p-value	Mean Difference (I-J)	p-value	Mean Difference (I-J)	p-value
Students self-assessment			-0.031	0.99	1.28*	<0.001
Teachers' assessment	0.031	0.99			1.31*	<0.001
Raters' assessment	-1.28*	<.001	-1.31*	<0.001		

* The mean difference is significant at the 0.05 level.

Table B8 Post Hoc Tukey Analysis of PYP level grouped by assessor

PYP Levels	(I) Type	(J) Type	Mean Difference (I-J)	p-value
Elementary	Self-	TA	-0.03	0.998
		Raters	1.44*	0.004
	Tutors	SA	0.03	0.998
		Raters	1.47*	0.004
	Raters	SA	-1.44*	0.004
		TA	-1.47*	0.004
Intermediate	Self-	TA	-0.22	0.67
		Raters	0.92*	0.001
	Tutors	SA	0.22	0.67
		Raters	1.14*	<0.001
	Raters	SA	-0.92*	0.001
		TA	-1.14*	<0.001
Advanced	Self-	TA	0.26	0.67
		Raters	1.77*	<0.001
	Tutors	SA	-0.26	0.67
		Raters	1.52*	<0.001
	Raters	SA	-1.77*	<0.001
		TA	-1.52*	<0.001

Appendix C

Table C1 Post-Hoc Tukey HSD of levels perceived to be required by students across five MHCCs

Dependent Variable	(I) College	(J) College	Mean Difference (I-J)	Std. Error	P-value	95% Confidence Interval	
						Lower Bound	Upper Bound
Overall written interaction	Medicine	Pharmacy	0.91	0.35	0.064	-0.03	1.86
		Nursing	1.07	0.36	0.026	0.08	2.06
		Dentistry	-0.05	0.38	1.00	-1.09	1.00
		AMS	0.33	0.32	0.84	-0.56	1.22
		Nursing	0.16	0.36	0.99	-0.82	1.13
		Dentistry	-0.96	0.38	.082	-1.99	0.07
	Pharmacy	AMS	-0.58	0.32	0.36	-1.45	0.29
		Dentistry	-1.12	0.39	0.036	-2.19	-0.05
		AMS	-0.74	0.34	0.18	-1.66	0.18
		Dentistry	0.38	0.36	0.83	-0.60	1.36
		Pharmacy	1.10	0.36	0.021	0.11	2.08
		Nursing	1.69	0.38	0.000	0.66	2.72
Orthographic Control	Medicine	Dentistry	0.003	0.40	1.00	-1.09	1.10
		AMS	0.60	0.34	0.40	-0.33	1.53
		Nursing	0.60	0.37	0.50	-0.43	1.62
	Pharmacy	Dentistry	-1.09	0.40	0.047	-2.18	-0.01
		AMS	-0.50	0.34	0.58	-1.42	0.43
		Dentistry	-1.69	0.41	0.00	-2.82	-0.56
	Nursing	AMS	-1.09	0.36	0.019	-2.07	-0.12
		Dentistry	0.6-	0.38	0.52	-0.44	1.63

Reports and Essays	Medicine	Pharmacy	0.61	0.22	0.039	0.02	1.21
		Nursing	0.74	0.23	0.011	0.12	1.36
		Dentistry	0.05	0.24	1.00	-0.61	0.71
		AMS	0.31	0.20	0.56	-0.25	0.87
		Nursing	0.12	0.22	0.98	-0.49	0.74
	Pharmacy	Dentistry	-0.56	0.24	0.13	-1.21	0.09
		AMS	-0.31	0.20	0.55	-0.86	0.25
		Dentistry	-0.69	0.25	0.045	-1.36	-0.01
		AMS	-0.43	0.21	0.26	-1.01	0.15
		Dentistry	AMS	0.26	0.23	0.79	-0.37
Note Taking	Medicine	Pharmacy	0.79	0.26	0.020	0.08	1.49
		Nursing	1.06	0.27	0.001	0.33	1.79
		Dentistry	0.32	0.29	0.81	-0.47	1.10
		AMS	0.88	0.24	0.003	0.22	1.55
		Nursing	0.27	0.27	0.84	-0.46	1.00
	Pharmacy	Dentistry	-0.47	0.28	0.46	-1.25	0.31
		AMS	0.010	0.24	0.99	-0.57	0.76
		Dentistry	-0.74	0.29	0.084	-1.55	0.06
		AMS	-0.18	0.25	0.96	-0.87	0.51
		Dentistry	AMS	0.57	0.27	0.23	-0.18
Average CEFR Levels	Medicine	Pharmacy	0.74	0.22	0.006	0.15	1.32
		Nursing	1.33	0.22	<0.001	0.72	1.94
		Dentistry	0.17	0.24	0.95	-0.48	0.82
		AMS	0.62	0.20	0.018	0.07	1.17
		Nursing	0.59	0.22	0.063	-0.02	1.20
	Pharmacy	Dentistry	-0.56	0.24	0.13	-1.22	0.09
		AMS	-0.11	0.20	0.98	-0.66	0.44
		Dentistry	-1.16	0.25	<0.001	-1.83	-0.48
		AMS	-0.70	0.21	0.008	-1.28	-0.13
		Dentistry	AMS	0.45	0.23	0.27	-0.17

Table C2 Post-Hoc Tamhane test of levels perceived to be required by students across five MHCCs

Dependent Variable	(I) College	(J) College	Mean Difference (I-J)	Std. Error	P-value	95% Confidence Interval	
						Lower Bound	Upper Bound
Overall written Production	Medicine	Pharmacy	0.22	0.32	0.999	-0.67	1.12
		Nursing	1.86	0.39	<0.001	0.74	2.99
		Dentistry	0.30	0.32	0.97	-0.61	1.22
		AMS	0.43	0.30	0.82	-0.43	1.29
		Nursing	1.64	0.39	<0.001	0.53	2.74
	Pharmacy	Dentistry	0.08	0.31	1.00	-0.81	0.97
		AMS	0.21	0.29	0.999	-0.62	1.04
		Dentistry	-1.56	0.39	0.001	-2.68	-0.44
		AMS	-1.43	0.38	0.002	-2.51	-0.35
		Dentistry	AMS	0.13	0.30	1.00	-0.73
Type of Texts	Medicine	Pharmacy	1.07	0.36	0.037	0.04	2.10
		Nursing	1.30	0.43	0.033	0.06	2.54
		Dentistry	-0.19	0.39	1.000	-1.29	0.92
	Pharmacy	AMS	0.69	0.35	0.41	-0.31	1.69
		Nursing	0.23	0.43	1.000	-0.98	1.45
		Dentistry	-1.26	0.38	0.012	-2.34	-0.18
		AMS	-0.38	0.34	0.96	-1.35	0.60

What Students Can Write	Nursing	Dentistry	-1.49	0.45	0.011	-2.76	-0.22	
		AMS	-0.61	0.42	0.80	-1.80	0.58	
	Dentistry	AMS	0.88	0.37	0.17	-0.17	1.93	
		Pharmacy	1.09	0.34	0.018	0.11	2.06	
	Medicine	Nursing	1.98	0.39	<0.001	0.87	3.08	
		Dentistry	0.55	0.36	0.75	-0.49	1.59	
	Vocabulary Range & Control	AMS	Dentistry	0.91	0.36	0.13	-0.13	1.94
			Nursing	0.89	0.37	.17	-0.17	1.95
		Pharmacy	Dentistry	-0.53	0.35	.74	-1.52	0.46
			AMS	-0.18	0.35	1.000	-1.16	0.80
		Nursing	Dentistry	-1.42	0.39	0.004	-2.54	-0.30
			AMS	-1.07	0.39	0.069	-2.18	0.05
Dentistry		AMS	0.35	0.37	0.98	-0.69	1.40	
		Pharmacy	0.89	0.31	0.047	0.01	1.77	
Medicine		Nursing	1.35	0.36	0.003	0.32	2.39	
		Dentistry	0.74	0.31	0.17	-0.14	1.63	
Grammatical Accuracy		AMS	Dentistry	0.77	0.30	0.10	-0.08	1.61
			Nursing	0.46	0.37	0.91	-0.60	1.52
	Pharmacy	Dentistry	-0.15	0.32	1.00	-1.06	0.77	
		AMS	-0.13	0.31	1.00	-1.00	0.75	
	Nursing	Dentistry	-0.61	0.37	0.66	-1.67	0.45	
		AMS	-0.59	0.36	0.67	-1.62	0.44	
	Dentistry	AMS	0.02	0.31	1.00	-0.85	0.90	
		Pharmacy	0.72	0.29	0.12	-0.09	1.53	
	Medicine	Nursing	1.51	0.33	<0.001	0.56	2.46	
		Dentistry	0.71	0.27	0.092	-0.06	1.49	
	Processing Texts	AMS	AMS	0.90	0.27	0.011	0.13	1.67
			Nursing	0.79	0.36	0.25	-0.23	1.81
Pharmacy		Dentistry	-0.01	0.30	1.00	-0.87	0.85	
		AMS	0.18	0.30	1.00	-0.68	1.04	
Nursing		Dentistry	-0.80	0.35	0.21	-1.79	0.19	
		AMS	-0.61	0.35	0.57	-1.60	0.38	
Dentistry		AMS	0.19	0.29	0.999	-0.64	1.01	
		Pharmacy	0.54	0.33	0.65	-0.39	1.47	
Medicine		Nursing	1.34	0.33	0.001	0.40	2.29	
		Dentistry	-0.21	0.39	1.00	-1.31	0.89	
Processing Texts		AMS	Dentistry	0.65	0.32	0.35	-0.25	1.56
			Nursing	0.80	0.29	0.063	-0.02	1.62
	Pharmacy	Dentistry	-0.75	0.35	0.30	-1.75	0.25	
		AMS	0.11	0.28	1.00	-0.67	0.89	
	Nursing	Dentistry	-1.55	0.36	<0.001	-2.57	-0.53	
		AMS	-0.69	0.28	0.15	-1.49	0.11	
	Dentistry	AMS	0.86	0.34	0.13	-0.12	1.85	

Table C3 Post-Hoc Tukey HSD of levels students perceive as being needed across the across the three PYP Levels

Dependent Variable	(I) PYP Levels	(J) PYP Levels	Mean Difference (I-J)	Std. Error	P-value	95% Confidence Interval	
						Lower Bound	Upper Bound
Type of Texts	Elementary	Intermediate	-0.84	0.38	0.073	-1.74	0.06
		Advanced	-1.75	0.40	<0.001	-2.68	-0.82

Vocabulary Range and Control	Intermediate	Elementary	0.84	0.38	0.073	-0.06	1.74
		Advanced	-0.91	0.26	0.002	-1.53	-0.29
	Advanced	Elementary	1.75	0.40	<0.001	0.82	2.68
		Intermediate	0.91	0.26	0.002	0.29	1.53
	Elementary	Intermediate	-0.48	0.33	0.309	-1.26	0.29
		Advanced	-0.96	0.34	0.014	-1.75	-0.16
Intermediate	Elementary	0.48	0.33	0.309	-0.29	1.26	
	Advanced	-0.47	0.23	0.100	-1.01	0.07	
Advanced	Elementary	0.96	0.34	0.014	0.16	1.75	
	Intermediate	0.47	0.23	0.100	-0.07	1.01	
Orthographic Control	Elementary	Intermediate	-0.67	0.37	0.162	-1.53	0.19
		Advanced	-1.57	0.38	<0.001	-2.46	-0.69
Intermediate	Elementary	Advanced	0.67	0.37	0.162	-0.19	1.53
		Advanced	-0.90	0.25	0.001	-1.49	-0.31
Advanced	Elementary	Advanced	1.57	0.38	<0.001	0.69	2.46
		Intermediate	0.90	0.25	0.001	0.31	1.49
Note Taking	Elementary	Intermediate	-0.30	0.26	0.480	-0.90	0.31
		Advanced	-0.95	0.26	0.001	-1.57	-0.32
Intermediate	Elementary	Advanced	0.30	0.26	0.480	-0.31	0.90
		Advanced	-0.65	0.18	0.001	-1.07	-0.23
Advanced	Elementary	Advanced	0.95	0.26	0.001	0.32	1.57
		Intermediate	0.65	0.18	0.001	0.23	1.07
Average levels	Elementary	Intermediate	-0.50	0.22	0.051	-1.01	0.00
		Advanced	-1.24	0.22	<0.001	-1.76	-0.72
Intermediate	Elementary	Advanced	0.50	0.22	0.051	0.00	1.01
		Advanced	-0.74	0.15	<0.001	-1.09	-0.39
Advanced	Elementary	Advanced	1.24	0.22	<0.001	0.72	1.76
		Intermediate	0.74	0.15	<0.001	0.39	1.09

Table C4 Post-Hoc Tamhane test of levels students perceived as being needed across the three PYP Levels

Dependent Variable	(I) PYP	(J) PYP	Mean Difference (I-J)	Std. Error	P-value	95% Confidence Interval	
	(II) Levels	Levels				Lower Bound	Lower Bound
Overall Written Production	Elementary	Intermediate	-0.66	0.41	0.29	-1.65	0.33
		Advanced	-1.56	0.40	0.001	-2.53	-0.59

	Intermediate	Elementary	0.66	0.41	0.29	-0.33	1.65
		Advanced	-0.90	0.21	<0.001	-1.41	-0.39
	Advanced	Elementary	1.56	0.40	0.001	0.59	2.53
		Intermediate	0.90	0.21	<0.001	0.39	1.41
Overall Written Interaction	Elementary	Intermediate	-0.98	0.33	0.011	-1.78	-0.18
		Advanced	-1.69	0.34	<0.001	-2.51	-0.87
	Intermediate	Elementary	0.98	0.33	0.011	0.18	1.78
		Advanced	-0.71	0.24	0.008	-1.28	-0.15
What Students Can Write	Elementary	Intermediate	-1.13	0.36	0.007	-2.01	-0.26
		Advanced	-1.84	0.36	<0.001	-2.72	-0.96
	Intermediate	Elementary	1.13	0.36	0.007	0.26	2.01
		Advanced	-0.71	0.26	0.019	-1.33	-0.09
Grammatical Accuracy	Elementary	Intermediate	1.84	0.36	<0.001	0.96	2.72
		Advanced	0.71	0.24	0.008	0.15	1.28
	Intermediate	Elementary	-1.13	0.36	0.007	-2.01	-0.26
		Advanced	-1.84	0.36	<0.001	-2.72	-0.96
Processing Texts	Elementary	Intermediate	1.13	0.36	0.007	0.26	2.01
		Advanced	-0.71	0.26	0.019	-1.33	-0.09
	Intermediate	Elementary	1.84	0.36	<0.001	0.96	2.72
		Advanced	0.71	0.26	0.019	0.09	1.33
Reports and Essays	Elementary	Intermediate	-0.19	0.31	0.90	-0.96	0.57
		Advanced	-1.06	0.31	0.003	-1.81	-0.31
	Intermediate	Elementary	0.19	0.31	0.90	-0.57	0.96
		Advanced	-0.87	0.21	<0.001	-1.37	-0.36
Grammatical Accuracy	Advanced	Elementary	1.06	0.31	0.003	0.31	1.81
		Intermediate	0.87	0.21	<0.001	0.36	1.37
	Elementary	Intermediate	-0.60	0.28	0.099	-1.28	0.08
		Advanced	-1.45	0.30	<0.001	-2.16	-0.73
Processing Texts	Intermediate	Elementary	0.60	0.28	0.099	-0.08	1.28
		Advanced	-0.85	0.22	0.001	-1.38	-0.31
	Advanced	Elementary	1.45	0.30	<0.001	0.73	2.16
		Intermediate	0.85	0.22	0.001	0.31	1.38
Reports and Essays	Elementary	Intermediate	-0.32	0.18	0.22	-0.76	0.12
		Advanced	-0.91	0.19	<0.001	-1.37	-0.46
	Intermediate	Elementary	0.32	0.18	0.22	-0.12	0.76
		Advanced	-0.59	0.15	<0.001	-0.96	-0.23
Grammatical Accuracy	Advanced	Elementary	0.91	0.19	<0.001	0.46	1.37
		Intermediate	0.59	0.15	<0.001	0.23	0.96

Table C5: a summary of findings from the focus group brainstorming/mind -maps activity

<u>Topics</u>			
<u>College of Nursing</u>	<u>College of AMS</u>	<u>College of Pharmacy</u>	<u>College of Medicine</u>
- Kidney failure	- Assessment of sound articulations	- Aloe and its origin and medical use	- Medical related topics (for each block in the curriculum)
- Hypertension	- Origins of the language	- Prescriptions of different medicine	- Time management
- Unprofessional nursing	- Autism	- Prophetic medicine	- Doctor-patient relationship
- Nursing theory	- Thyroid gland	- Depression	- Professionalism
- Nursing care for patients with cardio	- Broca's articles		

<ul style="list-style-type: none"> - Nursing care for patients with respiratory program 	<ul style="list-style-type: none"> - Rule of ethics - Translation of child's voice sample - ADHD - Learning disabilities - pregnancy - Down syndrome - Diabetes - Asthma - Nuclear receptor - Parts of the eye - Optics - Clinical practice management - endocrine system 		<ul style="list-style-type: none"> - How to deal with stress - How to deal with patients
<u>The Purpose of Writing</u>			
<ul style="list-style-type: none"> - Exams (SAQ, list, define) - Assignments (situation-based assignments, analysing scenario, analysing a problem (what is the effect of a problem, what will be the nurse responsibility in a certain situation) - Class activities - Emails - Research - Homework 	<ul style="list-style-type: none"> - Exams - Assignments - Presentations - Emails - Surveys - Research - Homework 	<ul style="list-style-type: none"> - Exams (complete the answer, short answer questions, give reasons) - Assignments - Emails - Note-taking - Project - Letters - Homework - Research (mini research about a plant) 	<ul style="list-style-type: none"> - Exams (SAQ) - OSPE (practical exam) - Assignments - Notes - Research - Filling out forms - Emails (communication with faculty staff) - Presentations
<u>Writing Genre</u>			
<ul style="list-style-type: none"> - Note taking - Documentation - Diagnosis - Short answer questions - Short essays - Case study (very limited essays and reports) 	<ul style="list-style-type: none"> - Summary of articles - Presentations - Reports - Brochures - Notes - Essays - SAQ - Diagnostic reports 	<ul style="list-style-type: none"> - Reports (sometimes) - Posters (mini research e.g. depression) - Brochures - Short essays - Paragraphs 	<ul style="list-style-type: none"> - Essays - Summary - Reports (clinical skills reports to medical education) - Mind maps - Write about a case - Posters
<u>Rhetorical Modes</u>			
<ul style="list-style-type: none"> - Analysis - Description - Reflection 	<ul style="list-style-type: none"> - Rephrasing - Summarising - Analysing 	<ul style="list-style-type: none"> - Description - Comparison - Reasoning - Analysis (analysing patient's disease) 	<ul style="list-style-type: none"> - Reflection on a situation - Analysis of a situation - Description - Explanation

Appendix D

Table D1 One-Way Analysis of Variance ANOVA of differences between Students' assessment at phase I and II across medical MHCCs

CEFR Scales	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	η^2
Overall written Production						
Between Groups	39.12	4	9.78	1.23	0.30	0.020
Within Group	1966.31	247	7.96			
Total	2005.43	251				
Overall written interaction						
Between Groups	18.29	4	4.57	0.44	0.78	0.007
Within Group	2579.28	247	10.44			
Total	2597.57	251				
Type of Texts						
Between Groups	37.02	4	9.26	0.91	0.46	0.015
Within Group	2509.41	247	10.16			
Total	2546.43	251				
Vocabulary Range and Control						
Between Groups	5.98	4	1.49	0.20	0.94	0.003
Within Group	1825.23	245	7.45			
Total	1831.20	249				
Grammatical Accuracy						
Between Groups	27.44	4	6.86	0.86	0.49	0.014
Within Group	1968.37	246	8.00			
Total	1995.81	250				
Orthographic Control						
Between Groups	9.51	4	2.38	0.27	0.90	0.004
Within Group	2132.71	244	8.74			
Total	2142.22	248				
Processing Texts						
Between Groups	15.71	4	3.93	0.51	0.73	0.008
Within Group	1895.82	245	7.74			
Total	1911.52	249				
Reports and Essays						
Between Groups	83.04	4	20.76	3.18	0.014	0.049
Within Group	1599.20	245	6.53			
Total	1682.24	249				
Note Taking						
Between Groups	33.38	4	8.35	1.19	0.32	0.019
Within Group	1719.20	245	7.02			
Total	1752.58	249				
Average CEFR Levels						
Between Groups	1.94	4	0.48	0.15	0.96	0.002
Within Group	812.21	247	3.29			
Total	814.14	251				

SS=Sum of squares, df=degrees of freedom, MS=mean square, F=F ratio, η^2 =Effect size

Table D2 Robust Test of Equality of Mean of differences between Students' assessment at phase I and II across medical MHCCs

CEFR Scales	Statistic	df1	df2	P-value
What Students Can Write				
Welch	0.76	4	89.65	0.56
Brown-Forsythe	0.78	4	141.81	0.54

df=degrees of freedom

Table D3 Post-hoc Tukey HSD of student misalignment by medical MHCCs

Dependent Variable	(I) College	(J) College	Mean Difference (I-J)	Std. Error	P-value	95% Confidence Interval	
						Lower Bound	Upper Bound
Reports and Essays	Medicine	Pharmacy	-0.70	0.48	0.59	-2.02	0.61
		Nursing	-1.99	0.66	0.024	-3.81	-0.17
		Dentistry	-0.71	0.51	0.64	-2.12	0.70
	Pharmacy	AMS	-1.27	0.45	0.039	-2.50	-0.04
		Nursing	-1.29	0.67	0.31	-3.12	0.55
		Dentistry	-0.01	0.52	1.00	-1.44	1.42
	Nursing	AMS	-0.57	0.46	0.73	-1.82	0.69
		Dentistry	1.28	0.69	0.35	-0.62	3.18
		AMS	0.72	0.65	0.80	-1.05	2.49
	Dentistry	AMS	-0.56	0.49	0.79	-1.91	0.79

Table D4 One-Way Analysis of Variance of Students misalignment across the three PYP Levels

CEFR Scales	SS	df	MS	F	P-value	η^2
Type of Texts						
Between Groups	45.55	2	22.78	2.27	0.11	0.018
Within Group	2500.88	249	10.04			
Total	2546.43	251				
What Students Can Write						
Between Groups	55.59	2	27.79	2.99	0.052	0.024
Within Group	2295.60	247	9.29			
Total	2351.18	249				
Vocabulary Range and Control						
Between Groups	93.01	2	46.51	6.61	0.002	0.051
Within Group	1738.19	247	7.04			
Total	1831.20	249				
Grammatical Accuracy						
Between Groups	0.28	2	0.14	0.02	0.98	0.000
Within Group	1995.53	248	8.05			
Total	1995.81	250				
Processing Texts						
Between Groups	47.83	2	23.92	3.17	0.044	0.025
Within Group	1863.70	247	7.55			
Total	1911.52	249				
Note Taking						

Between Groups	43.07	2	21.54	3.11	0.046	0.025
Within Group	1709.50	247	6.92			
Total	1752.58	249				
Average CEFR Levels						
Between Groups	47.95	2	23.97	7.79	0.001	0.059
Within Group	766.20	249	3.08			
Total	814.14	251				

SS=Sum of squares, df=degrees of freedom, MS=mean square, F=F ratio, η^2 =Effect size

Table D5 Robust Test of Equality of Mean of Student misalignment across the three PYP Levels

CEFR Scales	Statistic	df1	df2	P-value
Overall Written Production				
Welch	4.61	2	55.58	0.014
Brown-Forsythe	4.40	2	86.62	0.015
Overall Written Interaction				
Welch	5.96	2	67.46	0.004
Brown-Forsythe	8.10	2	174.76	<0.001
Orthographic Control				
Welch	0.49	2	52.40	0.62
Brown-Forsythe	0.45	2	67.93	0.64
Reports and Essays				
Welch	18.00	2	57.50	<0.001
Brown-Forsythe	18.56	2	95.52	<0.001

df=degrees of freedom

Table D6 Post-Hoc Tukey HSD of student misalignment across the three PYP Levels

Dependent Variable	(I) PYP Levels	(J) PYP Levels	Mean Difference (I-J)	Std. Error	P-value	95% Confidence Interval	
						Lower Bound	Upper Bound
Vocabulary Range and Control	Elementary	Intermediate	-0.20	0.63	0.95	-1.28	1.67
		Advanced	1.07	0.64	0.22	0.43	2.10
	Intermediate	Elementary	0.20	0.63	0.95	-2.57	0.43
		Advanced	1.27	0.35	0.001	-2.10	-0.43
Processing Texts	Advanced	Elementary	-1.07	0.64	0.22	-1.29	1.77
		Intermediate	-1.27	0.35	0.001	-0.47	2.63
	Elementary	Intermediate	0.24	0.65	0.93	-1.77	1.29
		Advanced	1.08	0.66	0.23	-0.02	1.71
Note Taking	Intermediate	Elementary	-0.24	0.65	0.92	-2.63	0.47
		Advanced	0.85	0.37	0.055	-1.71	0.02
	Advanced	Elementary	-1.08	0.66	0.23	-2.08	0.84
		Intermediate	-0.85	0.37	0.055	-1.24	1.73
Average levels	Elementary	Intermediate	-0.62	0.62	0.58	-0.84	2.08
		Advanced	0.24	0.63	0.92	0.04	1.69
	Intermediate	Elementary	0.62	0.62	0.58	-1.73	1.24
		Advanced	0.86	0.35	0.038	-1.69	-0.04
Average levels	Advanced	Elementary	-0.24	0.63	0.92	-1.21	0.74
		Intermediate	-0.86	0.35	0.038	-0.32	1.66
	Elementary	Intermediate	-0.24	0.41	0.83	-0.74	1.21
		Advanced	0.67	0.42	0.25	0.36	1.46
	Intermediate	Elementary	0.24	0.41	0.83	-1.66	0.32

	Advanced	0.91	0.23	<0.001	-1.46	-0.36
Advanced	Elementary	-0.67	0.42	0.25	-1.28	1.67
	Intermediate	-0.91	0.23	<0.001	0.43	2.10

Table D7 Post-Hoc Tamhane of student misalignment across the three PYP Levels

Dependent Variable	(I)	PYP	(J) PYP	Mean Difference (I-J)	Std. Error	P-value	95% Confidence Interval	
	(II)	Levels	Levels				Lower Bound	Lower Bound
Overall Written Production	Elementary		Intermediate	-0.15	0.65	0.99	-1.78	1.48
			Advanced	0.88	0.61	0.41	-0.68	2.44
	Intermediate		Elementary	0.15	0.65	0.99	-1.48	1.78
			Advanced	1.03	0.35	0.012	0.18	1.88
	Advanced		Elementary	-0.88	0.61	0.41	-2.44	0.68
			Intermediate	-1.03	0.35	0.012	-1.88	-0.18
Overall Written Interaction	Elementary		Intermediate	-0.98	0.53	0.19	-2.30	0.33
			Advanced	0.47	0.55	0.78	-0.89	1.84
	Intermediate		Elementary	0.98	0.53	0.19	-0.33	2.30
			Advanced	1.46	0.43	0.002	0.42	2.49
	Advanced		Elementary	-0.47	0.55	0.78	-1.84	0.89
			Intermediate	-1.46	0.43	0.002	-2.49	-0.42
Reports and essays	Elementary		Intermediate	-0.31	0.55	0.92	-1.69	1.07
			Advanced	1.55*	0.53	0.020	0.20	2.91
	Intermediate		Elementary	0.31	0.55	0.92	-1.07	1.69
			Advanced	1.86	0.32	<0.001	1.10	2.63
	Advanced		Elementary	-1.55	0.53	0.020	-2.91	-0.20
			Intermediate	-1.86	0.32	<0.001	-2.63	-1.10

Table D8 One-Way Analysis of Variance of tutors' and students' misalignment by MHCCs

CEFR Scales	SS	df	MS	F	P-value	η^2
Overall written Production						
Between Groups	14.72	4	3.68	0.50	0.73	0.008
Within Group	1754.40	240	7.31			
Total	1769.13	244				
Overall written interaction						
Between Groups	39.01	4	9.75	1.28	0.28	0.021
Within Group	1835.66	240	7.65			
Total	1874.67	244				
Type of Texts						
Between Groups	28.32	4	7.08	0.91	0.46	0.015
Within Group	1863.66	240	7.77			
Total	1891.98	244				
What Students Can Write						
Between Groups	30.75	4	7.69	1.05	0.38	0.017
Within Group	1752.21	239	7.33			
Total	1782.96	243				

Vocabulary Range and Control						
Between Groups	6.87	4	1.72	0.28	0.89	0.005
Within Group	1451.78	238	6.10			
Total	1458.65	242				
Grammatical Accuracy						
Between Groups	5.71	4	1.43	0.22	0.93	0.004
Within Group	1580.30	239	6.61			
Total	1586.01	243				
Orthographic Control						
Between Groups	29.19	4	7.30	0.96	0.43	0.017
Within Group	1714.66	225	7.62			
Total	1743.84	229				
Processing Texts						
Between Groups	60.35	4	15.09	2.37	0.054	0.040
Within Group	1441.21	226	6.38			
Total	1501.56	230				
Reports and Essays						
Between Groups	6.23	4	1.56	0.27	0.90	0.005
Within Group	1315.32	225	5.85			
Total	1321.55	229				
Note Taking						
Between Groups	5.94	4	1.48	0.19	0.94	0.003
Within Group	1727.18	224	7.71			
Total	1733.12	228				
Average CEFR Levels						
Between Groups	8.27	4	2.07	0.62	0.65	0.011
Within Group	747.31	224	3.34			
Total	755.58	228				

SS=Sum of squares, df=degrees of freedom, MS=mean square, F=F ratio, η^2 =Effect size

Table D9 One-Way Analysis of Variance ANOVA of tutors' misalignment across the three PYP Levels

CEFR Scales	SS	df	MS	F	P-value	η^2
Overall written Production						
Between Groups	74.58	2	37.29	5.33	0.005	0.042
Within Group	1694.55	242	7.00			
Total	1769.13	244				
Overall written interaction						
Between Groups	41.06	2	20.53	2.71	0.069	0.022
Within Group	1833.61	242	7.58			
Total	1874.67	244				
Type of Texts						
Between Groups	26.90	2	13.45	1.75	0.18	0.014
Within Group	1865.08	242	7.71			
Total	1891.98	244				
What Students Can Write						
Between Groups	50.68	2	25.34	3.53	0.031	0.028
Within Group	1732.28	241	7.12			
Total	1782.96	243				
Vocabulary Range and Control						
Between Groups	57.83	2	28.92	4.95	0.008	0.040
Within Group	1400.82	240	5.84			
Total	1458.65	242				
Grammatical Accuracy						

Between Groups	8.67	2	4.33	0.66	0.52	0.005
Within Group	1577.35	241	6.55			
Total	1586.01	243				
Orthographic Control						
Between Groups	45.97	2	22.99	3.07	0.048	0.026
Within Group	1697.87	227	7.48			
Total	1743.84	229				
Reports and Essays						
Between Groups	22.81	2	11.41	1.99	0.14	0.017
Within Group	1298.73	227	5.72			
Total	1321.55	229				
Note Taking						
Between Groups	23.62	2	11.81	1.56	0.21	0.014
Within Group	1709.50	226	7.56			
Total	1733.12	228				
Average CEFR Levels						
Between Groups	35.77	2	17.89	5.62	0.004	0.047
Within Group	719.80	226	3.19			
Total	755.58	228				

SS=Sum of squares, df=degrees of freedom, MS=mean square, F=F ratio, η^2 =Effect size

Table D10 Robust Test of Equality of Mean of teacher misalignment across the three PYP Levels

CEFR Scales	Statistic	df1	df2	P-value.
Processing Texts				
Welch	9.58	2	48.71	<0.001
Brown-Forsythe	10.37	2	84.57	<0.001

df=degrees of freedom

Table D11 Post-hoc Tukey HSD of misalignments between tutors and students across PYP

Dependent Variable	(I) PYP Levels	(J) PYP Levels	Mean Difference (I-J)	Std. Error	P-value	95% Confidence Interval	
						Lower Bound	Upper Bound
Overall written Production	Elementary	Intermediate	1.67	0.58	0.013	0.29	3.05
		Advanced	1.91	0.59	0.004	0.52	3.30
	Intermediate	Elementary	-1.67	0.58	0.013	-3.05	-0.29
		Advanced	0.24	0.36	0.78	-0.61	1.08
	Advanced	Elementary	-1.91	0.59	0.004	-3.30	-0.52
		Intermediate	-0.24	0.36	0.78	-1.08	0.61
What Students Can Write	Elementary	Intermediate	0.85	0.59	0.33	-0.55	2.24
		Advanced	1.47	0.60	0.038	0.06	2.88
	Intermediate	Elementary	-0.85	0.59	0.33	-2.24	0.55
		Advanced	0.62	0.36	0.20	-0.23	1.48
	Advanced	Elementary	-1.47	0.60	0.038	-2.88	-0.06
		Intermediate	-0.62	0.36	0.20	-1.48	0.23
Vocabulary Range and Control	Elementary	Intermediate	0.76	0.54	0.34	-0.52	2.04
		Advanced	1.51	0.55	0.017	0.22	2.80
	Intermediate	Elementary	-0.76	0.54	0.34	-2.04	0.52

		Advanced	0.75	0.33	0.058	-0.02	1.52
	Advanced	Elementary	-1.51	0.55	0.017	-2.80	-0.22
		Intermediate	-0.75	0.33	0.058	-1.52	0.02
Orthographic Control	Elementary	Intermediate	0.76	0.71	0.54	-0.92	2.44
		Advanced	1.47	0.72	0.10	-0.22	3.15
	Intermediate	Elementary	-0.76	0.71	0.54	-2.44	0.92
		Advanced	0.71	0.38	0.15	-0.18	1.59
	Advanced	Elementary	-1.47	0.72	0.10	-3.15	0.22
		Intermediate	-0.71	0.38	0.15	-1.59	0.18
Average levels	Elementary	Intermediate	0.93	0.46	0.10	-0.14	2.01
		Advanced	1.43	0.46	0.006	0.35	2.50
	Intermediate	Elementary	-0.93	0.46	0.10	-2.01	0.14
		Advanced	0.49	0.25	0.11	-0.09	1.07
	Advanced	Elementary	-1.43	0.46	0.006	-2.50	-0.35
		Intermediate	-0.49	0.25	0.11	-1.07	0.09

Table D12 Post-Hoc Tamhane of tutors' and students' misalignment across the three PYP Levels

Dependent Variable	(I) PYP Levels	(J) PYP Levels	Mean Difference (I-J)	Std. Error	p-value	95% Confidence Interval	
						Lower Bound	Upper Bound
Processing Texts	Elementary	Intermediate	1.22	0.57	.127	-0.256	2.7
		Advanced	2.34	0.60	.002	0.815	3.855
	Intermediate	Elementary	-1.22	0.57	.127	-2.7	0.256
		Advanced	1.11	0.34	.004	0.292	1.933
	Advanced	Elementary	-2.34	0.60	.002	-3.855	-0.815
		Intermediate	-1.11	0.34	.004	-1.933	-0.292

Table D13 Differences Between PYP Tutors (Phase I) and College Staff' assessment at phase II using independent t-test

CEFR Scales	Phase I (Levels Achieved)		N	Phase II (Levels Needed)		Differences					
	M	SD		M	SD	M	SD	t	df	P-value	Cohen's d
Overall Written Production	6.64	2.17	19	5.89	1.41	-0.74	0.51	-1.46	262	.144	-0.35
Overall Written Interaction	6.11	2.12	19	4.95	1.68	-1.16	0.50	-2.24	262	.021	-0.53
Type of Texts Need to be able to Write	6.36	2.24	19	6.47	2.44	0.11	0.54	0.21	262	.837	0.05

What Can They Write	5.46	2.12	19	6.32	1.89	0.86	0.50	1.71	262	.089	0.41
Vocabulary	5.53	2.07	19	5.63	1.89	0.10	0.49	0.21	261	.834	0.05
Grammatical Control	5.44	2.08	19	5.47	1.65	0.04	0.49	0.07	261	.943	0.02
Orthographic Control	4.91	2.16	19	5.63	1.67	-0.24	0.49	-0.50	248	.617	-0.12
Processing Texts	4.91	2.16	19	4.47	1.35	-0.44	0.50	-0.87	248	.384	-0.21
Writing Essays	5.61	2.23	19	5.84	1.50	0.23	0.52	0.44	247	.661	0.10
Note Taking	5.37	2.34	19	5.53	1.39	0.16	0.55	0.29	247	.770	0.07
Average CEFR Scales	5.71	1.75	19	5.62	1.00	-0.09	0.41	-0.22	262	.826	-0.05

M=Mean, SD=Standard deviation, df=degrees of freedom

Coding scheme for CERF Scales: 1 (A1); 2 (A2); 3 (A2+); 4 (B1); 5 (B1+); 6 (B2); 7 (B2+); 8 (C1); 9 (C2)

Misalignment score calculated as phase II level MINUS phase I level.

Cohen's-d calculated as $2 \times t / \sqrt{df}$

Table D14 Differences Between Teachers' assessment at phase I and II using Welch's two sample t-test

CEFR Scales	Phase I (Levels Achieved)		Phase II (Levels Needed)		Differences					
	M	SD	M	SD	M	SD	t	df	P-value	Cohen's d
Overall Written Production	6.64	2.17	5.89	1.41	-0.74	0.35	-2.11	25	.045	-0.35
Processing Texts	4.91	2.16	4.47	1.35	-0.44	0.34	-1.29	26	.21	-0.21
Average CEFR Scales	5.71	1.75	5.62	1.00	-0.089	0.25	-0.35	27	.73	-0.05

M=Mean, SD=Standard deviation, df=degrees of freedom

Coding scheme for CERF Scales: 1 (A1); 2 (A2); 3 (A2+); 4 (B1); 5 (B1+); 6 (B2); 7 (B2+); 8 (C1); 9 (C2)

Misalignment score calculated as phase II level MINUS phase I level.

Cohen's-d calculated as $2 \times t / \sqrt{df}$