Title: Differences in individual speech when learning through PBL discussion



A PhD thesis submitted by

Catherine Haines

March 2017

MEDICAL EDUCATION UNIT
SCHOOL OF MEDICINE
UNIVERSITY OF NOTTINGHAM UK

Acknowledgements

My special thanks go to Professor Reg Dennick for his continued interest and support for my project and for the many fascinating discussions we have shared about learning, communication, science and art.

I also thank my friends for their forbearance over the length of my preoccupation with this part-time study. I thank my sister for her loyalty, my father, the late Dr Robert Haines, for instilling me with a love of learning and wonder at the Universe, and my mother, the late Elizabeth Haines, who helped me to seek to share this, simply and creatively.

Table of Contents

Title: Differences in individual speech when learning through PBL discussion 1
Acknowledgements 2
Table of Contents 3
List of Tables 5
List of Figures 11
ABSTRACT: Differences in individual speech when learning through PBL discussion. 13
Chapter 1 – Introduction to the thesis 17
1.1 Introduction to the chapter 17
1.2 Overview of PBL education 20
1.3 The GEMS PBL Curriculum entry to medical schools 21
1.4 Discussion as a means of learning 22
1.5 The role of the group in the work of discussion 23
1.6 The discussion group as a means of learning specialist vocabulary 24
1.7 The discussion group as a protective social unit 24
1.8 The role of the individual in discussion learning: talking and listening, talkativeness 24
1.9 Psychological factors in the PBL setting 25
1.10 Corpus analysis as a tool for researching discussion 28
1.11 Summary 29
1.12 Conclusion 29
1.13 Chapter Outline for the study 30
2 Literature Review 32
2.1 Introduction to the chapter 32
2.2 PBL in medical education 37
2.3 Groups and learning 49
2.4 Introduction to Learning styles 85
2.5 Introduction to the literature relating Personality to the PBL setting 102
2.6 Applied linguistics and Corpus analysis 126
2.7 Literature review conclusions 151
2.8 Research questions 152
3 Methods 158
3.1 Introduction 158
3.2 Quintax Questionnaires: personality typing 189
3.3 Kolb Learning Styles Inventory (LSI) 200

3.4 Statistical Methods 209
4 Corpus Analysis Results 214
4.1 Introduction to the chapter 214
4.2 Study of the characteristics of the PBL spoken corpus in comparison with BNC Spoken (Educational) 229
4.3 A comparison using corpus analysis of the Female and Male Corpora 244
4.4 Corpora divided according to academic performance 253
4.5 Personality: analysis by FFM 275
4.6 Quintax temperament Types: Four-way corpus analysis 298
4.7 Kolb Learning Styles analysis 310
5 Statistical Results 322
5.1 Introduction to Chapter 322
5.2 Descriptive Statistics 323
5.3 Associations between entry measures (dependent variables) 335
5.4 Knowledge test at year end as dependent variable 339
5.5 Total Speech 347
6 Discussion 351
6.1 Introduction 351
6.2 General Comment on Corpus Analysis 357
6.3 Speech behaviour and gender 358
6.4 Performance and significant differences 363
6.5 Personality type, learning styles and speech behaviour 370
6.6 The main implications of this study for learners and facilitators of PBL discussions 379
6.7 Re-statement of the known limitations of the study 381
6.8 Future research questions arising as a result of this study 385
6.9 Conclusions: 387
6.10 Appendices 403
Appendix B: Ethics Addendum 407
Appendix C: Sample study consent form 410
Appendix D: Sample individual Quintax report 413
Appendix E Spreadsheet showing summary and averages for each PBL group's participation 415
Appendix F: SPSS variables 418
Appendix G Poster pilot 2: (Haines 2010) 420
Appendix H: Blank Kolb inventory 421

List of Tables

Table 1 Theoretical Perspectives for understanding groups adapted from Poole &Hollingshead, 2005 52
Table 2 Adapting Brookefield's conversational roles to relate to PBL discussions 59
Table 3 Assessment categories used by teachers to assess student contributions Huang et al. Grades A – D for contribution, within the four categories, speaking, tone, attitude, announcement 69
Table 4 Summarising instruments, with date of instrument version, devised to assess learning styles according to the perceived stability of the learning style 87
Table 5 Summarising the KLSI types 93
Table 6 Nicknames for the Four Kolb Learning Styles 94
Table 7 KLSI polarities and Honey and Mumford's renaming 94
Table 8 A summary of the FFM traits and sub facets: NEO PI-Revised 106
Table 9 A brief summary of McCrae and Costa's FFM continua 107
Table 10 Summary table for four dominant Quintax temperaments 109
Table 11 FFM Dimensions: distribution for medical students n=449 123
Table 12 Personality and language use adapted from (Iacobelli, Gill et al. 2011) 144
Table 13 Research Questions posed 154
Table 14 Table describing Quintax scales, in preferred order for Quintax 193
Table 15 Dependent Variables: capable of being used in analyses 211
Table 16 Independent variables capable of being considered in analysis 212
Table 17 Summary of the Corpora and Sub corpora and their relative sizes 219
Table 18 Frequency of nouns in PBL corpus 223
Table 19 Analysis of parts of speech (POS) in the PBL corpus 224
Table 20 Frequency of conjunction use, CS Parts of Speech, (POS) category 225
Table 21 Question word Use 226
Table 22 Overall semantic categorisation of the PBL spoken Corpus 227
Table 23 B3: Most frequent terms within B3 category 228
Table 24 Terms used significantly more frequently in the PBL corpus 234
Table 25 Terms used significantly less frequently in the PBL corpus in comparison to the BNC spoken 235
Table 26 Parts of speech (POS) used significantly more frequently in the PBL corpus 236
Table 27 Parts of speech (POS) used significantly less frequently in the PBL corpus 237
Table 28 Semantic categories: PBL corpus v BNC Ed Spoken: Subjects which occurred more frequently in the PBL corpus 238
Table 29 PBL corpus v BNC Ed Spoken: Semantic analysis. Less likely to occur in PBL corpus 239
Table 30 List of Z99 word frequency occurrences 240
Table 31 Table of occurrences and relative frequencies for Semantic category, Z4 241
Table 32 Terms used significantly more frequently in the female corpus and significantly less frequently in the male corpus 244

Table 33 Terms used significantly less frequently in the female corpus and significantly more frequently in the male corpus 245
Table 34 Parts of speech terms used significantly more frequently in the female corpus and significantly less frequently in the male corpus 246
Table 35 Parts of Speech analysis. A comparison between Female and Male Interjections – (UH) 248
Table 36 Parts of speech terms used significantly less frequently in the female corpus and significantly more frequently in the male corpus 248
Table 37: A comparison between female and male conjunctive statement (CS) use 249
Table 38 Semantic categories used significantly more frequently in the female corpus and significantly less frequently in the male corpus 249
Table 39 Comparison of pronoun use: female to male 250
Table 40 Semantic categories used significantly less frequently in the female corpus and significantly more frequently in the male corpus 251
Table 41 An analysis of significant differences between occurrence of some highly technical terms, over occurring in the male corpus and under occurring in the female corpus 252
Table 42 Frequency analysis showing categories used significantly more frequently in the low GAMSAT score corpus and significantly less frequently in the high GAMSAT score corpus 253
Table 43 Frequency analysis showing categories used significantly less frequently in the low GAMSAT score corpus and significantly more frequently in the high GAMSAT score corpus 254
Table 44 POS analysis showing categories used significantly more frequently in the low GAMSAT score corpus and significantly less frequently in the high GAMSAT score corpus 255
Table 45 POS analysis showing categories used significantly less frequently in the low GAMSAT score corpus and significantly more frequently in the high GAMSAT score corpus 256
Table 46 Semantic analysis showing categories used significantly more frequently in the low GAMSAT score corpus and significantly less frequently in the high GAMSAT score corpus 256
Table 47 Most frequently used words in cause and effect category 257
Table 48 A comparison of the occurrence of question words between the GAMSAT corpora 258
Table 49 Frequency analysis: nouns occurring more frequently in the low PPD corpus 260
Table 50 Frequency of words used significantly less frequently in the low PPD score corpus and significantly more frequently in the high PPD score corpus 261
Table 51 POS used significantly more frequently in the low PPD score corpus and significantly less frequently in the high PPD score corpus 262
Table 52 Frequency of Parts of Speech used significantly less frequently in the low PPD score corpus and significantly more frequently in the high PPD score corpus 263
Table 53 Frequency of semantic categories which are used significantly more frequently in the low PPD score corpus and significantly less frequently in the high PPD score corpus 263

Table 54 Frequency of semantic categories which are used significantly less frequently in the low PPD score corpus and significantly more frequently in the high PPD score corpus 264
Table 55 Frequency of words used significantly more frequently in the low MCQ score corpus and significantly less frequently in the high MCQ score corpus 266
Table 56 Frequency of words used significantly less frequently in the low MCQ score corpus and significantly more frequently in the high MCQ score corpus 268
Table 57 POS analysis showing POS used significantly more frequently in the low MCQ score corpus and significantly less frequently in the high MCQ score corpus 268
Table 58 POS analysis used significantly less frequently in the low MCQ score corpus and significantly more frequently in the high MCQ score corpus 269
Table 59 Low to high comparison of question use in MCQ 270
Table 60 Question words – occurrence in high to low MCQ scorers 271
Table 61 Semantic analysis showing categories used significantly more frequently in the low MCQ score corpus and significantly less frequently in the high MCQ score corpus 271
Table 62 Semantic analysis showing categories used significantly less frequently in the low MCQ score corpus and significantly more frequently in the high MCQ score corpus 272
Table 63 Most frequently used words in Anatomy category: low MCQ scorers v high MCQ scorers 273
Table 64 Frequency of words used significantly more frequently in the introvert corpus and significantly less frequently in the extravert corpus 275
Table 65 Frequency of words used significantly less frequently in the introvert corpus and significantly more frequently in the extravert corpus 276
Table 66 Parts of Speech: categories used significantly less frequently in the introvert corpus and significantly more frequently in the extravert corpus 276
Table 67 Question word Use introvert v extravert 277
Table 68 Semantic analysis: categories used significantly more frequently in the introvert corpus and significantly less frequently in the extravert corpus 278
Table 69 Semantic analysis: categories used significantly less frequently in the introvert corpus and significantly more frequently in the extravert corpus 279
Table 70 Differences in use of B3 category words: a comparison between introverts and extraverts 281
Table 71 Semantic analysis: categories used significantly more frequently in the introvert corpus and significantly less frequently in the extravert corpus 281
Table 72 Frequency analysis: terms used significantly less frequently in the adaptable corpus and significantly more frequently in the structured corpus 284
Table 73 Parts of Speech: categories used significantly less frequently in the adaptable corpus and significantly more frequently in the structured corpus 284
Table 74 Parts of Speech: categories used significantly more frequently in the adaptable corpus and significantly less frequently in the structured corpus 285
Table 75 Semantic analysis: categories used significantly less frequently in the adaptable corpus and significantly more frequently in the structured corpus 286
Table 76 Frequency analysis: terms used significantly more frequently in the personable corpus and significantly less frequently in the logical corpus 287

Table 77 Frequency analysis: terms used significantly less frequently in the personable corpus and significantly more frequently in the logical corpus 287
Table 78 Parts of Speech: categories used significantly more frequently in the personable corpus and significantly less frequently in the logical corpus 288
Table 79 Semantic analysis: categories used significantly less frequently in the personable corpus and significantly more frequently in the logical corpus 288
Table 80 Frequency analysis: terms used significantly more frequently in the grounded corpus and significantly less frequently in the theoretical corpus 289
Table 81 Semantic analysis: categories used significantly less frequently in the grounded corpus and significantly more frequently in the theoretical corpus 290
Table 82 Frequency of nouns used significantly more frequently in the calm corpus and significantly less frequently in the volatile corpus 291
Table 83 Frequency of nouns used significantly less frequently in the calm corpus and significantly more frequently in the volatile corpus 292
Table 84 Frequency of nouns used significantly more frequently in the calm corpus and significantly less frequently in the volatile corpus 292
Table 85 Examination of most frequently occurring general adjectives JJ category, from POS analysis 293
Table 86 Frequency of nouns used significantly less frequently in the calm corpus and significantly more frequently in the volatile corpus 294
Table 87 Semantic analysis: categories used significantly more frequently in the calm corpus and significantly less frequently in the volatile corpus 294
Table 88 A comparison of terms most frequently mentioned in the Disease category 295
Table 89 Semantic analysis: categories used significantly less frequently in the calm corpus and significantly more frequently in the volatile corpus 296
Table 90 A Comparison of semantic category S6+; obligation between the calm and volatile speaker 296
Table 91 Words over-used by AG as opposed to other types 298
Table 92 Words under-used by AG as opposed to other types 298
Table 93 Semantic categories over-used by AG as opposed to other temperaments combined 299
Table 94 Semantic categories under-used by AG as opposed to other temperaments combined 299
Table 95 Words over-used by SG as opposed to other types 301
Table 96 Words under-used by SG as opposed to other types 301
Table 97 Parts of speech under-used by SG as opposed to other temperaments combined 302
Table 98 Semantic categories over-used by SG as opposed to other temperaments combined 302
Table 99 Semantic categories under-used by SG as opposed to other temperaments combined 302
Table 100 Words over-used by LT as opposed to other types 304
Table 101 Words under-used by LT as opposed to other types 304
Table 102 Parts of speech under-used by LT as opposed to other temperaments combined 304
Table 103 Semantic categories over-used by LT as opposed to other temperaments combined 305

Table 104 Semantic categories under-used by LT as opposed to other temperaments combined 305
Table 105 Words over-used by PT as opposed to other types 307
Table 106 Words under-used by PT as opposed to other types 307
Table 107 Words over-used by PT as opposed to other types 308
Table 108 Semantic categories over-used by PT as opposed to other temperaments combined 308
Table 109 Semantic categories under-used by PT as opposed to other temperaments combined 309
Table 110 Words over-used by convergent types as opposed to other types 310
Table 111 Words under-used by convergent types as opposed to other types 311
Table 112 Parts of Speech, over-used by convergent types as opposed to other types 311
Table 113 Parts of Speech, under-used by convergent types as opposed to other types 311
Table 114 Semantic categories, over-used by convergent types as opposed to other types 312
Table 115 Terms over-used by accommodator types as opposed to other types 313
Table 116 Terms under-used by accommodator types as opposed to other types 313
Table 117 Parts of Speech analysis, showing under-use by accommodator types as opposed to other types 314
Table 118 Semantic analysis, showing over-use by accommodator types as opposed to other types 314
Table 119 Word frequency analysis, showing over-use by accommodator types as opposed to other types 316
Table 120 Word frequency analysis, showing under-use by accommodator types as opposed to other types 316
Table 121 Parts of Speech analysis, showing under-use by assimilator types as opposed to other types 317
Table 122 Semantic analysis, showing over-use by assimilator types as opposed to other types 317
Table 123 Semantic analysis, showing under-use by accommodator types as opposed to other types 318
Table 124 Word frequency, terms over-used by divergent types as opposed to other types 319
Table 125 Word frequency, terms under-used by divergent types as opposed to other types 319
Table 126 Parts of Speech; terms under-used by divergent types as opposed to other types 319
Table 127 Semantic categories: terms over-used by divergent types as opposed to other types 320
Table 128 Semantic categories: terms under-used by divergent types as opposed to other types 320
Table 129 Basic descriptive statistics for dispersion of key continuous variables 329
Table 130 showing acceptable Collinearity Statistics for multiple regression analysis model 4 340
Table 131 Summarising table of results in relation to the research questions posed 353

List of Figures

Figure 1 Early clinical experience and educational theories summarised (da Silva, 2014) ... 42
Figure 2 Heron's facilitator behaviours ... 61
Figure 3 Schmidt and Moust, 2000 Model demonstrating interdependence of PBL elements ... 63
Figure 4 Dolmens and Schmidt's Theme 1: Influence of the discussion in the tutorial group ... 65
Figure 5 Ringelmann's horse power findings ... 68
Figure 6 Visscher's et al. interactivity categories and themes, discussed in the focus group interviews ... 74
Figure 7 Kolb's Learning Styles Quadrants ... 90
Figure 8 Kolb's Learning Cycle ... 91
Figure 9 A typical sample from an educational speech corpus ... 131
Figure 10 Presenting a summary of meta analyses of gender differences in health care consultations ... 149
Figure 11 Sample of corpus text: Speaker 7 from PBL group 11, a selection of contributions ... 218
Figure 12 Concordance showing examples of B3 ... 228
Figure 13- BNC Sampler Corpus ... 230
Figure 14 Log likelihood significance level (Rayson 2003) ... 233
Figure 15 Summary of Bayes' effect size interpretation (Wilson 2013) ... 234
Figure 16 Concordance illustrating use of 'You' in female corpus ... 247
Figure 17 Concordance illustrating use of 'You' in male corpus ... 247
Figure 18 Concordance illustrating use of pronouns in low GAMSAT scorers ... 255
Figure 19 Concordance illustrating pronoun use in low score PPD corpus ... 261
Figure 20 Concordance illustrating pronoun use in low MCQ scorers ... 267
Figure 21 Concordance illustrating use of 'so' from low scorer MCQ corpus ... 270
Figure 22 Concordance illustrating use of 'so' from high scorer MCQ corpus ... 270
Figure 23 Interjections by extraverts ... 277
Figure 24 Concordance illustrating use of letters in quiz ... 285
Figure 25: Concordance illustrating interjections in personable speech ... 288
Figure 26 Grounded speech concordances showing awareness of time 'at the moment' ... 290
Figure 27 Gender Balance for PBL student sample ... 324
Figure 28 Introversion v extraversion for PBL student sample ... 324
Figure 29 Personable v logical (agreeableness) for PBL student sample ... 325
Figure 30 structured v adaptable (conscientiousness) for PBL student sample ... 325
Figure 31 grounded v theoretical (openness to experience) for PBL student sample ... 326
Figure 32 Calm and Volatile frequency (neurosis) for the PBL student sample ... 326
Figure 33 Quintax Temperament type frequency for PBL student sample ... 327
Figure 34 Kolb's Four learning styles for PBL student sample ... 328
Figure 35 GAMSAT scores were suitable for performing linear regression ... 337
Figure 36 GAMSAT scores were suitable for performing linear regression ... 338
Figure 37 MCQ scores were suitable for performing multiple linear regression ... 339
Figure 38 SPSS Coefficients table for Model 4, multiple linear regression predicting MCQ scores ... 341
Figure 39 Simplicity index – Flesch complexity measures were suitable for performing linear regression ... 343

Figure 40 Simplicity index – Flesch complexity measures were suitable for performing linear regression.....345
Figure 41 Simplicity index –percentage total speech and gender were suitable for performing linear regression 347
Figure 42 Introversion/extraversion and total proportion of turns to speak measures were suitable for performing linear regression 349

ABSTRACT: Differences in individual speech when learning through PBL discussion.

BACKGROUND: Problem Based Learning (PBL) facilitators have sometimes sought to motivate learners to 'talk more' in order to 'learn more' but there has been little work to differentiate through differences evident in their speech behaviour during PBL discussions and to identify which differences are associated with successful learning. Corpus analysis is a useful tool for probing students' speech in PBL, providing a quantitative basis for examining student talk.

AIMS AND OBJECTIVES: The aim of this study was to create a speech corpus of at least 100,000 words, which identifies each individual student's contribution to a PBL discussion of the same case at the same time within six randomly selected, gender-balanced PBL groups. The study then uses corpus analysis to explore the nature of the speech of individuals and their performance in learning. A statistical analysis was designed to triangulate findings through measuring and analysing underlying factors thought to influence the speech behaviour of individual learners, such as gender and personality type and to identify correlations with educational attainment.

METHODS: Wmatrix3 was used to identify significant differences in frequency, parts of speech and semantic category usage from verbatim transcripts of the same three session cycle of PBL. SPSS analyses identified significant differences between learners' speech behaviour, personality type, gender and attainment in the first year.

RESULTS: Multiple and simple linear regression analysis demonstrates significant correlation to a moderate degree between certain data points. In summary, the calmer, more conscientious, lower openness to experience learner, with a high GAMSAT score, achieved significantly higher MCQ scores at the end of year one. Introverted students spoke less frequently, but with

longer utterances. Male students spoke significantly more frequently than female students. More emotionally reactive (neurotic) students achieved significantly lower scores on knowledge tests at the close of year one and their talk focussed their own performance. Corpus analysis demonstrated statistically significant differences between personality, gender, high and low test performers. Lower performers frequently dwelt on 'what' and 'why' and more general technical terminology. Higher performers used 'when' and 'how' and highly specific technical vocabulary more frequently.

CONCLUSION: This study demonstrates that corpus analysis of PBL transcripts provides rich evidence of significant differences in individual speech behaviour whilst learning through discussion. The results broadly replicate well-established findings about gender communication styles and personality difference. This suggests that there may be a utilitarian case for enhancing the approach of PBL facilitators and learners to individual differences without resorting to the overgeneralisations of stereotyping. For example, male and female learners often demonstrate different communication strategies, which may potentially limit the breadth of their future clinical communication skills. Facilitators and learners may use this awareness to explicitly encourage both male and female learners to adopt the full range of communication behaviours. It is important for facilitators to model best practice in their own communication in the initial sessions. By making such language patterns evident, it becomes possible to ensure that PBL groups become more skilled in initiating spontaneous contributions from introvert and female learners, in encouraging longer contributions from extraverts, more factual statements from female students and more varied types of speech from male learners. Learners who identify themselves as emotionally reactive (neurotic or volatile) could productively seek to moderate unproductive anxiety from an early stage through adopting excellent study habits and attitudes, particularly with respect to assessments.

Abbreviations and specialist terms

	Term in full	**Definition, if needed**
ASE	Association for Science Education	
BASE	British Academic Spoken English Corpus	
BNC	British National Corpus	
BPS	British Psychological Society	
CANBEC	Cambridge and Nottingham Business English Corpus	
CANCODE	Cambridge and Nottingham Corpus of Discourse in English	
d.v.	Dependent variable	
CSA	Clinical Skills Assessment	
EFPA	European Federation of Psychologists' Associations	
FFM	Five Factor Method for personality type measurement	
GAMSAT	Graduate Australian Medical School Aptitude test	GEMS aptitude test used at Derby
GEMS	Graduate Entry Medical School	
GMC	General Medical Council	
i.v.	Independent or predictor variable	
KWIC	Key Words in Context	
KLSI	Kolb's Learning Styles Inventory	
LL	Log Likelihood	
MBTI	Myers Briggs Type Inventory	
MCQ	Multiple Choice Questions	
MiCase	Michigan Corpus of Spoken English	
NARIC	National Association for the recognition and comparison of education and skills	
NEO-PI -R	Neuroticism- extraversion-openness to experience personality inventory, revised	

ORPHEUS	FFM personality test for use in occupational settings	
PBL	Problem Based Learning	
PG/s	Postgraduate/s	
PPD	Personal and Professional Development	
Quintax	Quintax Personality Type Test	
RCGP	Royal College of General Practitioners	
SPQ	Study Process Questionnaire	
STROBE	STrengthening the Reporting of OBservational studies in Epidemiology	Research guidelines to strengthen observational research
UCREL	University Centre for Computer Corpus Research on language, Lancaster University	
UG/s	Undergraduate/s	
USAS	University semantic analysis system	
Wmatrix3	A corpus analysis tool, version 3, University of Lancaster	

Chapter 1 – Introduction to the thesis

1.1 Introduction to the chapter

This section introduces the broad medical educational context for the study and discusses why it is an important area to research. It introduces the current knowledge about PBL prior to the literature review chapter and identifies where this research study can make a novel contribution to the field. It will introduce the key educational and psychological concepts relating to the individual in the Problem Based Learning setting. It will propose corpus analysis and correlational analysis as suitable approaches to conduct this mixed methods, observational study. The introduction concludes with an outline of the subsequent chapters of the thesis.

1.1.1 Background

Research into student-centred learning seeks to understand powerful individual differences and to optimise learning for all. It seeks to deepen our understanding of what it means to be an active learner, or for an educator to create student engagement. Yet it is still a challenge to move beyond simplistic measures of participation, engagement and attainment. PBL facilitators have sometimes sought to motivate learners to 'talk more' in order to learn more (Azer 2009) but there has been little work to differentiate between learners through the differences evident in their speech behaviour and to explore how this may relate to factors such as personality, gender and test scores. Corpus analysis has begun to prove a useful tool for probing students' speech in PBL (Da Silva and Dennick 2010), providing a quantitative basis in addition to qualitative approaches to examine student talk. By necessity, this study is limited in its generalisability by being restricted to 45 subjects in 6 PBL groups, comprising approximately one half of one first year cohort of Derby, Graduate Entry Medical students (GEMS), cohort beginning 2011. However, the study does provide a close examination of 'who' is doing the learning through analysing how speech

behaviours during PBL discussions differ and how this may be related to measures of learning, on entry and at the close of the first year of study.

This thesis examines the relationships between a learners' personality type and learning style, speech during a PBL cycle and results at the start and close of the first year. The intention is to better understand factors which may influence successful study behaviour which relies strongly on speech and discussion. It is hoped that this study will be able to inform future facilitators of PBL discussion of behaviours which may enhance study. It may also have relevance to learning through other discussion-intensive methods.

1.1.2 The individual in medical education

Medical Education, the means by which medicine is taught and learned, has become a thriving academic field, with its own literature and research ethos, with the UK Medical Education journal recently celebrating the publication of its fiftieth volume (Eva 2016). Misch (Misch 2002), contests that medical students are not at liberty to behave as free adult learners capable of intrinsically motivated andragogy as envisaged by Knowles (Knowles 1998), but are indentured trainees. Indeed, medicine has long been taught through an apprenticeship model, based on the traditional practice which legally tied an individual to a master for a sufficient length of time to acquire the skills and experience necessary to become an autonomous practitioner. Kolb's influential learning cycle sought to summarise theories relating how experience can be transformed into new behaviour which could account for how learning such as that during an apprenticeship occurs (Kolb 1984).

One benefit of this system was that the apprentice was bound in partnership with the master and to some extent his needs (and it was usually a he) would have been considered. This tradition lives on through the array of supervision and tutoring which seeks to match the needs of the current 'apprenticed medic' with the opportunities provided by the system throughout lengthy postgraduate specialist training. Learning through being in relationship with another, such as a parent, is humanity's way of moulding

behaviour, establishing learning through the powerful mechanisms of mimicry, role modelling and social interaction championed by Vygotsky (Vygotskiĭ 1978).

1.1.3 Choosing the right individual applicants for medical school

Certainly selection for medical study is taken very seriously and as with any oversubscribed resource, is required to discriminate between applicants and to monitor the fairness of this process (Mathers, Sitch et al. 2016). For undergraduate medical study in the UK, the key determinant has always been academic attainment at Advanced Level study, usually for scientific subjects, including Chemistry. The ability to achieve high grades in a narrow range of scientific subjects has been taken to indicate that the learner will continue to be effective at medical school. In addition, students have been required to demonstrate their suitability and motivation in increasingly systematic ways through interviews, references and personal statements. Competition to enter Medical School is fierce. Many candidates gain the highest possible 'A' Level scores and so present an equality of opportunity challenge to discriminate fairly between candidates.

Establishing new GEMS at the turn of the century offered a chance to widen participation (Cleland, Nicholson et al. 2015). However, identifying new and reliable methods for selection to medical school which would predict future performance in the clinical setting and remove inequalities continues to be challenging. In their Ottawa 2010 summarising statement on selection for the healthcare professions, Prideaux, Roberts *et al* strongly recommended that selection for entry be subject to the same rigour as the assessments within medical school, particularly with respect to reliability and validity (Prideaux, Roberts et al. 2011). The statement characterises the current recruitment of clinicians as being based on multiple sources of data about candidates but with little compelling evidence for predictive validity and concerns for reliability. This is particularly true beyond medical school where the authors identify a general cognitive – non-cognitive divide. Within medical schools there is reliance on qualifications measuring cognitive

knowledge, but on entering clinical practice as a trainee this diminishes (Department of Health 2004) Unfortunately, the practices which are then increasingly used to assess and promote clinicians have not yet been much studied.

Measures of academic attainment for school leavers, such as the US Grade Point Average (GPA), and in the UK, 'A' level scores, demonstrated the strongest evidence for predictive validity. The UK Clinical Aptitude Test (UKCAT) is also used by a consortium of medical and dental Schools to compare aspiring students (UKCAT 2017). The Medical College Admission Test (MCAT) was identified as an acceptable alternative for graduate entry students, most medical studies in the US beginning after a first degree. Multiple mini-interviews (MMI) were also identified as improving on the previous very low predictive power of interviews, which were still a widespread practice. Personality testing was singled out as meriting further work.

However, the Ottawa statement also called attention to the imperative of ensuring that future clinicians were more representative of society as a whole, where students had the potential, but for a number of reasons, may not have been able to demonstrate this in the highly competitive circumstances of Medical School selection. Garrud has shown that medical schools in the UK are persistently under representing disadvantaged sections of society (Garrud 2011). If individual differences in addition to cognitive measures and social class may be associated with success at GEMS, this may contribute to selection which is more representative of society.

1.2 Overview of PBL education

PBL can no longer be considered an innovation in medical education. It was first implemented at McMaster University, Canada, in 1969. Several other universities followed over the next decade as this new method gained popularity. Today, 45 years later, PBL has spread worldwide and into many different disciplines such as engineering and law.

Savin-Baden and Major describe a worldwide shift in the views on education during this time, relating to the recognition of the individual (Savin-Baden and Major 2004). Indeed, the authors observe that women and ethnic minorities are more suited to the self-directed learning method of problem-based learning. Women now make up the majority of entrants to UK medical schools. Perhaps PBL better meets the needs of individuals, but perhaps it suits some individuals better than others; just as traditional approaches such as lectures and knowledge exams might suit some better than others.

In a debate between Dolmans and Gijbels, they both agreed that the future of PBL research should move away from questioning whether or not PBL is effective and towards exploring what aspects of PBL make it effective and to consider under which circumstances it is effective and why (Dolmans and Gijbels 2013). They describe this as a micro-analytical approach. With information about which individual features of PBL work for which type of person in which circumstance, programmes may be tailored to capitalise on aspects that are known to work.

1.3 The GEMS PBL Curriculum entry to medical schools

Diversity and supply issues drove an expansion of medical school places in the UK and Ireland at the end of the twentieth century (Department of Health 2004). The intention was to provide a 'fast-track' for graduates to join a traditional medical school after an accelerated programme of learning. Derby GEMS students study for 18 months and join the 'mainstream' at the clinical practice stage when traditional students have completed an extra year of study. One of the previous barriers was possession of relevant science and life science background. However, studies have found that, in spite of an initial phase when life sciences graduates were shown to out-perform non-biological science graduates, by the time of the final assessment stage, there was no difference between the two groupings. (Finucane, Flannery et al. 2013). It is the position of the Medical Schools'

Council that the only things which should affect an applicant's chance of entering the medical degree are ability and motivation (Medical Schools' Council 2017).

Derby GEMS students belong to two PBL groups per year and study a programme of 75 PBL cases which are designed to contextualise the associated anatomy, clinical skills and theoretical learning activities which are programmed around two core PBL sessions per week. The method adopted is an adaptation of the Maastricht seven stage model, but which is not strictly adhered to, according to the ethnographic study of PBL in use at Derby GEMS by one of the PBL facilitators (Jennings 2013).

Students typically brainstorm their learning and identify learning objectives in response to the trigger problem in session one. They divide the work between the individuals in the group to cover through self-study and then report back to the group in session two. Further results and information may be available, which leads to elaboration of knowledge, before a shorter final session focuses on clinical management, ensuring that the learning objectives matched those for the case, and briefly reviewing the learning process itself.

1.4 Discussion as a means of learning

Higher education in the arts and humanities disciplines has traditionally been delivered through a small number of lectures from experts, with linked opportunities for learners to discuss the ideas and apply them to relevant contexts through planned small group teaching sessions, often called seminars. Learning through discussion is largely taken for granted. Dysfunctional discussion groups are usually the only ones where this mechanism is explicitly discussed. The educational literature contributes surprisingly few explicit studies of discussion as an educational method. The literature focuses on talking as a means of acquiring language itself, for children and speakers of other languages.

1.5 The role of the group in the work of discussion

In some ways, a learning group discussion is functioning as a team with a joint purpose. The business literature examines many behaviours which were labelled and considered necessary for effective operation of teams. For example, the Belbin Roles, first observed and labelled by Meredith Belbin and husband, in the early 1970s, were used to identify ways in which managers could better facilitate the contributions of their employees (Belbin, 2000). Certainly, a number of people working together can achieve more in the same time frame.

The tradition of personal development and counselling identified that a group can also provide important information about how one is perceived or behaves by providing a forum where a group can express their unique viewpoints. In interprofessional learning in healthcare, group working is increasingly viewed as a vital context for clinical practitioners. Multi-source feedback is commonly used in medical training to provide such input in the appraisal context (Donaldson 2006).

The basic mechanism in discussion is that one can uncover one's views and understanding through expressing them to a sympathetic audience to whom we offer the same courtesy. One can observe one's own fluency, gain feedback from others and, by listening to others, compare our achievements with those of the group. Peer group pressure is also an extremely powerful motivational force, particularly during the socialisation process from adolescence to full adulthood. Through hearing what we agree or disagree with, we can create more sophisticated definitions. The PBL group seeks to maximize this process in order to teach medical terminology, clinical reasoning and basic clinical knowledge through exploring the differences in understanding between the individuals within an established educational group.

1.6 The discussion group as a means of learning specialist vocabulary

The PBL group is an excellent setting for acquiring new language, such as the medical terminology which must be learned in the first years of studying medicine. Bridge quantified the scale of this endeavour as being equivalent to 13,000 terms in the first year alone, equivalent to mastering the vocabulary of three Advanced Level qualifications in foreign languages (Bridge 1962)

1.7 The discussion group as a protective social unit

It is also worth noting that learning is an arduous and sometimes stressful process and that the support of a group can contribute markedly to reserves of resilience and thereby promote effectiveness in learning (Garrud and Yates 2012). By contrast, if a group operates 'dysfunctionally' the experience can damage or certainly impair the quality of the learning experience and limit the support available to an individual (Koh, Khoo et al. 2008).

1.8 The role of the individual in discussion learning: talking and listening, talkativeness

It is important to recognise that every speaker requires at least one listener. Thus, in a PBL group of say eight members, as one person learns through speaking, seven other people have the potential to learn through listening. When there is silence, and nobody speaks, it may be that all eight students are actively learning.

The extent to which an individual is willing to speak in group settings varies. It may also vary for one individual person depending on the context, a person being more willing to speak within a small group of known, supportive colleagues who have a working relationship and common goals, or within a family environment. However, even when the context remains

the same, a talkative, or extravert person will show a relatively stable tendency to speak more often than an introverted person who may always prefer to think things through before speaking out (Cain 2013).

1.9 Psychological factors in the PBL setting

1.9.1 Measuring individual difference: elusive measures of personality, multiple intelligences and learning style

Psychology has long sought to clarify what makes us all human, whilst defining how individuals differ. In essence, observation led to a typological approach; modern statistical methods, such as the FFM is strongly supported and neurological science is still at an early stage.

No doubt humans have always attempted to understand and predict the behaviour of their fellows. Theophrastus proposed thirty types of men in the Greece of the 4th century BC, based on his observation. The literature review, (Section 2.5) will review significant typologically-based approaches, particularly Jung's archetypes and the link with sixteen Myers Briggs Type Inventory (MBTI) types and four temperaments (Keirsey and Bates 1984), (Furnham 1996) (Jung 1921). The methods (Section 3.2) will describe the balance of detail and overview which is justified for this study in the light of modern statistically-based approaches, chiefly the five factor method (FFM).

Modern personality theory begins in the world of language, with the lexical hypothesis. This proposes that humans encode their understanding through language and so all descriptions of human qualities must be contained within the dictionary, ready for researchers to refine (Allport and Odbert 1936). Advances in statistical analysis led to a unifying approach which emerged in the 1980s through the work of Costa and McCrae (Costa and McCrae 1985). This describes the clusters of human characteristics for all humans through what became known as 'the big five' human traits: extraversion,

agreeableness, conscientiousness, openness to experience, and neuroticism. The five factor method (FFM) continues to be supported for is ability to predict much about human behaviour.

During the 1980s and 1990s there was an increase in the literature seeking to understand effectiveness through studying those who succeed – particularly in business and sporting fields (Bandura 1986, Covey 1989, Ericsson 2016). In education, simple dichotomous typologies, such as distinguishing between students who habitually adopt deep or surface styles of learning continue to be widely discussed in association with effectiveness in teaching and learning (Marton and Saljo 1976).

In the 1980s Howard Gardner's multiple intelligences introduced a new vocabulary for the factors impacting on success. Again, although widely appealing, comparison with current evidence in neuroscience was unable to identify a strong case (Waterhouse 2006).

For a time, there was great enthusiasm for learning styles theory which seemed to predict the strengths and needs of an individual with accuracy so that learning experiences could be more effectively tailored. However, comprehensive meta-analysis of learning styles and their impact on learning have been shown to be at best a small factor in predicting success or efficacy in learning (Coffield, Moseley et al. 2004).

The upside to the failure of the reductive view that we could measure once, tailor learning and predict outcomes, is that human learning is more adaptable than we thought: for all individuals. Neuroscience has now demonstrated that plasticity continues far longer than we realized (Grabowski. 2002).

1.9.2 Gender differences and PBL

One of the key differences between individuals is their gender. Gender has become the focus of intense study over the past fifty years. The most recent edition of 'Women, Men and Language' (Coates 2015) speaks of four separate waves of feminism over the past four decades and of the current

challenge to the binary definition of the social construct of gender itself. Explanations to account for difference has varied as evidence and approaches have continually developed between favouring the genetic, or 'nature' explanation and the mechanisms for encouraging expression, through 'nurture'. Sociological discourse sought to account for gender as a primarily social construct; whereas an increase in the technological and scientific measurements that can now be made of biological differences between individuals, strengthens evidence for natural biological differences. For example, recent popular science has considered the average differences between the genders related to levels of ability to empathise versus systematise (Ruigrok, Salimi-Khorshidi et al. 2014). Some have speculated on the evidence for the role this may play in selecting study in higher education (Wheelwright, Baron-Cohen et al. 2006).

In the UK, medical school places have been increasingly taken up by females. Garrud reports the totals from 2003 – 2009 for UK medical school entry to be split, 43% male and 57% female overall. (Garrud 2011) However, the staff employed within medical schools may not yet match this change, particularly at senior levels, which reflects the profile of medical students when the majority were male, some decades ago. The medical register in the UK states the totals to be: 55% male and 45% female practitioners in May 2016 (GMC 2016).

However, gender, has been shown to have marked effect on behaviour relevant to this study, particularly conversational style, (Tannen 2001, Coates 2015). It has also been shown to impact on scores in certain medical exams, particularly related to True/False/Abstain questions which one study showed to advantage males significantly and advised caution in their use (Kelly and Dennick 2009).

Personality type has been shown to be broadly equally distributed between the genders, and to be relatively stable over time, although agreeableness and neuroticism scores tend to be and remain higher in female subjects over their lifetime (Chapman, Duberstein et al. 2007).

1.10 Corpus analysis as a tool for researching discussion

My first and higher degree in English Literature and Creative Writing both concern themselves with the description and depiction of individuals through their behaviour, primarily through speech. Literature concerns itself with representing the world and also conducting experiments with ways of seeing the world as experienced by different characters. Creative writing involves the conscious choice of adopting a viewpoint, for example, first or third person, omniscient, or unreliable. This choice determines the view of events which is shared directly with the reader, or leaves lacunae for meaning to be discerned by the reader in the way they are likely to do in real life, where surely, being able to predict the likely behaviour of others has conferred an evolutionary advantage. Corpus analysis provides a tool which can reveal many of these patterns in texts. For example, the entire works of Dickens are digitised and available for study, with one recent publication at the University of Nottingham, exploring the use of body imagery in the characterisation of the works of Dickens (Mahlberg 2013). Applied linguistics, although a discipline with specific techniques in its own right, relies on other disciplines to provide the subjects for its study; the text or corpus.

1.10.1 Corpus Analysis and applied or computational linguistics in education and medical education studies

Corpus analysis tools have made it possible to analyse the patterns in language which would have been most laborious in the past. It is now feasible to study the patterns of language without spending many hours counting by hand. The technique has been applied to discussions with health practitioners. For example, either transcribing telephone conversations between health advisers on the NHS Advice Helpline and their patients, or through online discussions with teenagers, which also has the advantage of being self-documenting (Harvey, Brown et al. 2007).

However, the creation and use of speech corpora is rare. The Nottingham department of applied linguistics has compiled a one million word speech corpus of business meetings in English, the Cambridge and Nottingham

spoken Business English Corpus (CANBEC) (Carter 1998). Carter discusses the considerable challenges to creating this corpus. For example, when building CANBEC the researcher was refused permission to record in a high proportion of cases. When given permission, the researcher attended meetings to operate the equipment, which may have caused participants to alter the nature of their recorded speech, one of the unintended consequences associated with the well-known Hawthorne effect (Parsons 1974). For this reason, spoken corpora are often relatively small. The study of helpline discussions was based on 61,000 words and the part of patients were played by researchers so the relevant part of the corpus, relating to the adviser, was considerably smaller (Harvey, Brown et al. 2007). A spoken word corpus of PBL discussions where individual characteristics were measured and linked to speech would present a novel opportunity to study associations between factors thought to influence learning, differences in speech behaviour and performance in tests.

1.11 Summary

In Summary, this introduction has provided a starting point for the examination of speech behaviours, employed through PBL group discussions which might be associated with gender, personality types, learning styles and attainment during the early years of medical school.

1.12 Conclusion

The Derby GEMS began graduating students in 2005. A new medical school with a PBL rapid-entry curriculum provided challenges and opportunities for educators. The GEMS Nottingham curriculum is innovative with a strong student-centred basis. However great emphasis and reliance is placed on the effective learning opportunities created within the eighteen month curriculum centred on PBL discussions. Any study which reviews such a key element in detail from an educational perspective and seeks to identify enhancements for faculty and students is to be welcomed.

Many disciplines rely heavily on learning through discussion, particularly in the arts and humanities, and yet the exact mechanisms for how this might suit individual learners is not yet well understood. This study hopes to go some way towards rectifying this situation.

This chapter has provided an overview of the history, issues and current state of the relevant fields of study: individual factors such as gender, personality type and learning style which are thought to impact on learning success in higher education and the selection and attainment for medical students in early medical school; PBL and group discussion as learning methodologies, and the contribution of corpus analysis to understanding qualitative issues with a quantitative basis.

1.13 Chapter Outline for the study

Chapter one has introduced the background to recent thinking in the fields of learning through discussion, PBL, personality, learning style, and corpus analysis research.

Chapter two will review the literature in order to establish the relationship between this study and the existing knowledge about PBL.

Considering:

- Educational concepts in the PBL setting
- Key concepts and definitions relating to individual difference, learning, especially in groups, and educational attainment
- Psychological concepts in the PBL setting, particularly relating to personality and learning styles
- Corpus analysis, the creation of spoken word corpora and application of linguistics in educational and medical education settings.

It will conclude with a statement of the research aim, objectives and research questions.

Chapter three will present the methodologies chosen for this study.

The results of the study will then be presented in **Chapter four** (the analysis using corpus analysis and applied linguistic techniques) and **Chapter five** (the statistical analysis of quantitative data).

The thesis concludes with **Chapter six**, discussing the results and possible implications for faculty development in selecting, preparing and supporting students for effective learning through a PBL or heavily discussion based curriculum. The limitations of this study will be summarised. The final section will present the overall conclusions and suggest possible future research building on this study.

2 Literature Review

2.1 Introduction to the chapter

Firstly, the broad sweep of the literature which might be included in a review for this study is considered. Next the methodology adopted for this literature review will be discussed with reference to best practice and the limitations common to literature reviews of this type.

The focus of the literature review has been limited to four related fields:

- Educational concepts in the PBL setting
- Key concepts and definitions relating to individual difference, learning, especially in groups, and educational attainment
- Psychological concepts in the PBL setting, personality and learning styles
- Corpus analysis, the creation of spoken word corpora and application of linguistics in educational and medical education settings

The results of the literature review will be presented by reviewing each of these sections in turn, establishing key definitions. Key studies will be briefly synthesised, analysed and evaluated with respect to this study according to the stated literature review methodology. Each section will conclude with a brief summary of the relevant issues for this study.

Finally, as a result of the overall literature review, the need for this research study will be established, the research questions will be stated, in the form of aims and objectives and research hypotheses, where possible, for this PhD study.

2.1.1 Introduction to the literature: overview

There is clearly a vast body of literature from different disciplinary traditions which has some relevance to aspects of this study. Indeed, all behavioural

sciences seek to understand in what way humans are the same, and in what ways they differ, both as individuals and as groups. Although behavioural sciences differ in their terminology, focus and approach, they are often concerned with similar territory and may use identical methodologies such as questionnaires, assessment data, observations and interviews to achieve their aims.

In addition, this study employs corpus and statistical analysis tools. Applied linguistics needs a text, a body of words or corpus, and so methods used in very different fields may have relevance. Statistical methods are also tools which help to make differences and similarities in behaviour evident and susceptible to study.

It is necessary to establish the relevant current position within each of these discourses, in order to demonstrate the nature of the gaps in the literature, to define the research questions and to establish a credible methodology for this research.

2.1.2 Literature review methodology

The recent BEME Guide to conducting best evidence systematic review in medical education (Hammick, Dornan et al. 2010) argues that full systematic review is a group process, which is not a methodology available to an individual researcher undertaking a PhD study. The choice of methodology for the literature review for this study therefore consists of a narrative review and synthesis, conducted by an individual researcher, which makes use of elements of best practice from systematic review processes where possible, within these limitations. The methodology for searching, inclusion and evaluation will now be briefly discussed in order to provide transparency.

Hammick et al (Hammick, Dornan et al. 2010) recommend identifying the theoretical or conceptual framework for the literature review, identifying clear aims, objectives or research questions (where possible), setting the scope for searches, setting the criteria for inclusion and evaluating impact,

and using appropriate systems for storing and handling data. These will briefly be considered.

2.1.3 Theoretical or conceptual framework for the literature review

Firstly, the theoretical or conceptual framework for the literature review is in line with the position of the researcher and is broadly constructivist. This paradigm takes into account the researcher's history, with degrees in language and literature and a professional career in technical support before becoming a practitioner within the community of educational developers. As an emerging occupation, educational developers have constructed their practice and theoretical frameworks from a wide range of previous disciplinary and professional standpoints, as articulated by Lave and Wenger, characterised as situated learning and legitimate peripheral learning (Lave 1991). For medical educators, communities have emerged from both clinical practitioners and non-clinical, educational and other contributors giving a rich nexus of theories and approaches. In his summary of constructivist approaches to education over the past twenty five years of medical education, Dennick, identifies and synthesises many of these strands and their place in a constructivist approach to medical education (Dennick 2016)

Thus, the literature has been reviewed from as wide a range of standpoints as possible, in order to draw on the relevant disciplines in order to inform this study.

2.1.4 Aims and objectives, and where possible research questions for literature reviews

Hammick et al, (Hammick, Dornan et al. 2010) also acknowledge that medical education reviews are often best tackled through formulating aims and objectives rather than tightly focussed research questions. Therefore, the aims and objectives for the literature review were to adopt a broad scope, to identify the key contributory papers in the published and also grey literature, in particular PhD theses in the substantive fields of interest. This

review seeks to synthesise and evaluate these findings. A summary of the search terms and search strategy follows.

2.1.5 Search Strategy

Over a period of several years, searches were conducted of the major online databases which were relevant (PubMed, Medline, PsychInfo, Cinahl, Ovid and ERIC; Web of Science, e-theses). Further searches used Google and Google Scholar; personal knowledge; backward and forward citations; specific searches of medical education journals and medical education conference abstracts. Various combinations of search terms were used, according to requirements. Early full combined searches of all terms of interest revealed clear gaps in the literature which was confirmed by subsequent extensions of the search.

For example, within Web of Science, a significant database for cross disciplinary searches across medicine and the social sciences, the search for topics including personality and corpus analysis and Problem based learning in English: (TS=(personality AND corpus analysis))AND (PBL OR problem based learning) AND LANGUAGE:(English), returned only one record, which failed to satisfy the inclusion criteria discussed below. Repeating the search without PBL as a search term identified 98 possible papers, however, only one paper satisfied the inclusion criteria, and is included in the literature review. A search for "learning through discussion" identified 23 records, of which 2 were included.

2.1.6 Criteria for inclusion within literature review

The criteria for inclusion were chiefly that of relevance to the current study in terms of Population, Intervention (Indicator), Comparator (Control) and Outcome (P.I.C.O.) as defined and discussed for example at the University of Oxford's site (Oxford 2017).

Using this approach, papers which concern an identical population would restrict the search as closely as possible to GEMS students in the first year of

study, in the UK, whose study comprises a substantial proportion of PBL learning through discussion, but near matches would also be considered. The indicators which would be sought would be entry and first year assessment scores, personality measures, especially using the FFM, Learning styles, especially using the Kolb Learning Styles Inventory, individual speech behaviour during PBL discussions, group learning effectiveness measures. There would be no control group available, as it is an observational study. The outcome would be statistically significant findings or links between these indicators, if any and evidence of significant differences in corpus analysis.

Literature will primarily be evaluated on its relevance to the educational indicators identified above. In view of the wide range of literature under review, the account will focus on the most recent publications and consider systematic and meta-analyses, particularly for the extensive areas of personality and PBL.

This observational study does not seek to measure the transfer of learning for the subjects beyond their first year of GEMS. However, all forms of medical education by definition, primarily seek to create the most effective clinical practitioners they can within the resources available. In the UK, a series of publications from the GMC, has culminated in January 2016 in the publication of: Promoting excellence: standards for medical education and training, which defines what outcomes all medical education in the UK need to create (GMC 2016). Longitudinal studies which explore the transfer of aspects of study into clinical practice would be included in order to seek evidence of the effectiveness of the educational process in the clinical setting.

2.1.7 Storing and handling literature review data

Endnote 7 has been used to store the results of literature searches, both citations and where possible the full pdf text. Endnote has then been used to include citations embedded in the dissertation in Harvard style, from a total of 1125 records, 339 have been directly referenced.

2.2 PBL in medical education

2.2.1 An introduction to the scope of this section

This section will provide a brief overview of PBL as a method of teaching. PBL will be defined as a learning methodology and considered in the light of wider papers and systematic reviews which give a picture of the field, now in its fifth decade of use as an 'innovation' in medical education.

Specific descriptions of how PBL is delivered at Derby GEMS will be included in the Methods chapter. Specific literature which relates PBL closely to the key areas of interest, will be re-stated below but, will then be considered in the relevant, focussed sections of this literature review.

- Educational concepts in the PBL setting
- Key concepts and definitions relating to individual difference, learning, especially in groups, and educational attainment
- Psychological concepts in the PBL setting, personality and learning styles
- Corpus analysis, the creation of spoken word corpora and application of linguistics in educational and medical education settings

PBL is one of the most heavily researched innovations of the medical curriculum and it was not felt practical or necessary to attempt an in depth review of the wider field. It is important to state at the outset that this review seeks to exclude much of the literature relating to evaluating the effectiveness of PBL in comparison with other methods, and also to focus on the behaviour and expertise of the individual students rather than the facilitator.

In summary, numerous studies have established that, in essence, PBL methods have a valuable contribution to make to medical education. Given the high motivation, intellectual qualifications on entry and effectiveness of medical student learners, with adequate resourcing and sequencing of

clinical and lecturing inputs balancing delivery, PBL can deliver comparable outcomes to more traditional approaches. This is particularly true for post graduate learners. PBL may also contribute some additional benefits, being enjoyed by many learners, providing social benefits and being linked with enhanced confidence in transition to clinical practice and more consistent approaches to career updating (Strobel and Van Barneveld 2009, Norman and Schmidt 2016). This study will focus on literature describing the role of this, at times, elusive individual learner and to better understand their behaviour during discussions within the PBL group.

2.2.2 Current spread of PBL within medical education

Schmidt (Schmidt, Van Der Molen et al. 2009) cites evidence for small group tutorials on clinical problems being adopted within a majority of the medical schools in the United States, with 20% considering themselves to be fully PBL based. He claims that most Australian and many European curricula are strongly based on PBL as their instructional method. Within the UK, of 33 current undergraduate medical schools, eight describe themselves as offering a PBL style curriculum, five as integrated, and the remainder as traditional. Graduate entry medical courses are available at 16 medical schools (Medical Schools' Council 2017). Clearly, many medical students are currently learning through PBL discussions of some sort.

2.2.3 PBL Definitions and points of distinction

However, the PBL literature suffers from a lack of agreement or rigour regarding the definition of the precise form of PBL which is being practiced and the extent to which it is employed within the curriculum. So it is important to clarify what is meant by PBL, and briefly to review the process by which PBL students learn and to consider how it differs from other case-orientated teaching methods.

Barrows and Tamblyn provide a simple working definition, stating that PBL is: 'the learning that results from the process of working towards the

understanding or resolution of a problem' (Barrows and Tamblyn 1980). Dolmans (Dolmans, De Grave et al. 2005) identify three key characteristics:

- In PBL the clinical problems act as a stimulus for learning
- In PBL the tutor acts as a facilitator
- In PBL small group work forms the basis of interaction

Many enquiry–based educational methods make use of clinical cases to give context and integrate knowledge: an effective lecture or a visit to a clinic could both offer this. Barrows describes a hierarchical taxonomy, ranging from lecture-based case presentations through to closed-loop PBL where a reflective phase is added to the standard PBL process (Barrows 1986). However, most case-based methods begin at Bloom's cognitive domain level of application and invite the learner to apply the knowledge they have already acquired in a new and realistic context (Bloom 1989).

But one key difference is that PBL is specifically designed to confront the learner with their own lack of knowledge in the face of a clinical situation. The PBL clinical problem itself then becomes the stimulus, the trigger for new learning. In psychological terms, this confrontation maximizes 'cognitive dissonance', (Festinger 1957).

2.2.4 Personal demands of the PBL curriculum on an individual

Already we have described crucial differences from other forms of learning on which we might predict individuals to have strong personal views. The learner is repeatedly confronted with their lack of clinical knowledge in the clinical context and is required to acknowledge and discuss this with their peers in front of a facilitator, a member of staff. This may be easier to tolerate if an individual has maturity and a strong sense of confidence and self-worth. The facilitator is not required to directly teach or explain, but to help the learners to do so. This may be frustrating where an individual has a more traditional expectation of staff and how to acquire knowledge. An individual spends 4 – 7 hours a week in the company of a small group of

peers, who may have a powerful influence on their effectiveness and enjoyment. This situation makes demands on an individual's appetite for sociability, tolerance of the differing needs and skills of others and commitment to inter-dependence, within an environment which strongly assesses and rewards competitiveness and autonomy. Where discussed in the literature, some of these issues will feature in future sections, particularly as individual difference is accounted for through personality theory.

2.2.5 PBL Cycle and steps: elaboration of the definition

The precise process and delivery of PBL varies widely. In general, a PBL 'cycle' is based around a single clinical case and consists of two or three group meetings with intervening self-directed study periods, where individuals study tasks agreed within the PBL group. Each PBL session is usually 1-2 hours in length.

2.2.5.1 The role of each meeting

Meeting 1:
Introduction of the case. Discussion leads to the group formulating learning objectives which are allocated within the members of the group to complete as individuals during an intervening self-study phase, typically over 3 – 5 days, during which related learning activities may be programmed.

Meeting 2:
Individuals deliver the learning objectives to the group

Meeting 3:
The case is resolved and the effectiveness of the group is briefly reviewed

There are a number of stages which the students work through to complete the cycle. Again, the exact stages can vary between institutions. Harvard

(Davis 1999) uses a six step process whereas perhaps the most well-known is the Maastricht seven-step process from the Dutch university, which is summarised by Wood as follows (Wood 2003):

2.2.6 The Seven Steps

1. Clarify the meaning of terms contained within the problem
2. Define the problem
3. Brainstorm ideas
4. Structure information from brainstorm results and form hypotheses
5. Formulate learning objectives through group consensus
6. Search for information in self-directed study and prepare to share this with the group
7. Synthesis of research information within the group to ensure all learning objectives have been addressed.

The key purpose of the steps within PBL is to maximise the educational benefit. If steps are missed or poorly completed, the learning will suffer as a result.

2.2.6.1 Brief overview of key Educational Theory and PBL: Two ways of conceiving of the PBL model contrasted: Dolmans, Schmidt

Dolmans (Dolmans, De Grave et al. 2005) provides us with a brief summary relating PBL to current educational theory reminding us that:

PBL is:

- *constructive* where learning is created through discussion and elaboration in a group activity,
- *self-directed* where individuals plan, monitor and evaluate their own learning process,
- *collaborative* as the group has a common goal and
- *contextual* where information is contextualised to make it relevant to clinical practice though the use of real cases.

These four fundamental aspects of PBL form part of the educational theory supporting its use. The following figure maps some of the relationships between learning theories which contribute to building early clinical knowledge and experience.

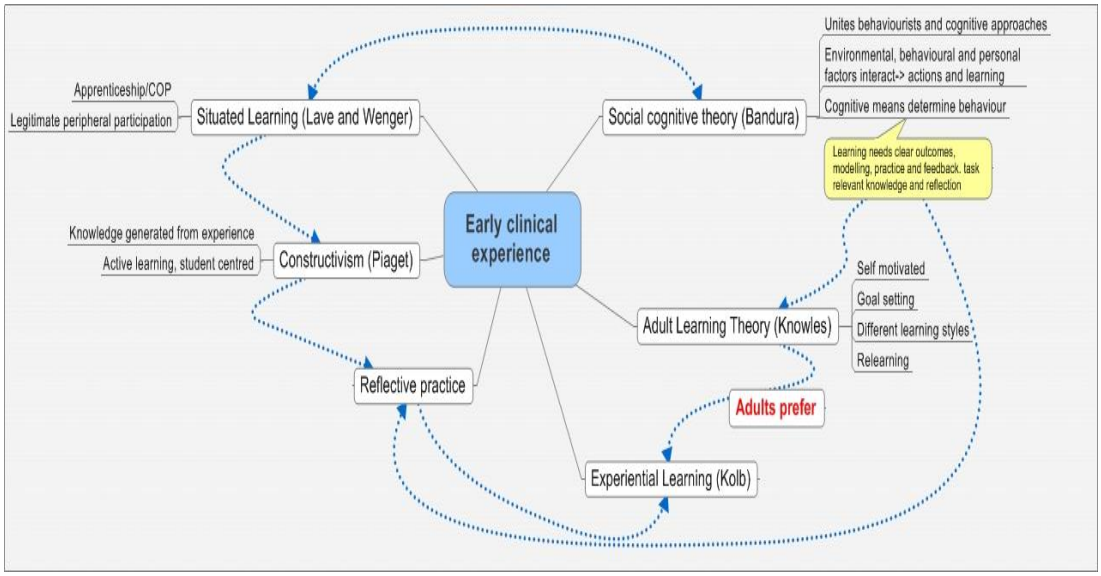


Figure *1* Early clinical experience and educational theories summarised (da Silva, 2014)

2.2.7 A brief overview of the social constructivist elements – i.e.: the construction of knowledge through group elaboration and discussion

There have been strong criticisms of approaches where learners have been left to 'construct' their own view of knowledge with insufficient guidance which has led to concern about gaps in knowledge and poorer results than 'full guidance' (Kirschner, Sweller et al. 2006, Clark, Kirschner et al. 2012). Such studies have drawn attention to the role of cognitive overload impairing the effectiveness of learning in novices. However, in the domain of highly motivated, adult post-graduate learners, there is a strong theoretical case for allowing the individual to determine more of their learning needs, and customise their learning experience to their personal profile of preference and existing expertise (Knowles 1998).

2.2.8 A brief overview of the social Dimension in PBL

One strand of educational theory links Vygotsky, (Vygotskiĭ 1978) with Bandura's ideas about learning through modelling and social influences, (Bandura 1986), and Lave and Wenger's (Lave 1991) notions of communities of practice where the PBL group is the 'family' and nascent professional community where the behaviours of becoming a doctor are rehearsed. Learners gain feedback from their peers and their effectiveness with learning through the PBL process. The social dimension is also stressed by Savery and Duffy, when they identify the role of other learners and the facilitator in developing understanding, which could not be undertaken alone. They select, 'encourage testing ideas against an alternative framework' and 'design the learning environment to support and challenge the learner's thinking' as two of their eight key 'instructional principles' (Savery and Duffy 1995). However, how individuals differ in their behaviour in groups, as demonstrated through speech, is not foregrounded.

2.2.9 PBL and evidence for enhancing social competencies

However, being an individual within the PBL group may enhance an individual's social competence. This will also be considered in the section on learning in groups. Koh et al (Koh, Khoo et al. 2008) conducted a systematic review of the evidence for the influence of PBL learning on 37 competencies after graduation. They concluded that PBL positively influenced both observed and self-reported assessments of four competences:

- coping with uncertainty
- appreciating legal and ethical issues in health care
- communication skills
- self-directed continuing learning

An individual's skill in coping with uncertainty and using their communication skills are both firmly within the remit of this study and will re-emerge through this review. For example, responses to uncertainty relate closely to the emotional volatility (neuroticism) dimension of the FFM of personality

measurement. Corpus analysis can provide detailed insight into communication skills based on sampling speech behaviour.

2.2.10 Schmidt – six fundamental educational processes

Schmidt et al (Schmidt, Rotgans et al. 2011) give a powerful summary of what they term, the sizable volume of research into the cognitive constructivist approach to PBL and summarise six fundamental educationally effective processes.

2.2.11 Schmidt's six fundamental educational processes in PBL

Schmidt et al's summary identifies key tenets of PBL as follows:

1. Activation of prior learning has a major impact on learning efficacy. (Ausubel, Novak et al. 1968)
2. Prior knowledge is necessary for subsequent knowledge
3. Knowledge is structured by the individual and in relation to the context
4. Storage and retrieval of knowledge improves after elaboration
5. Activation from long term memory depends on external cues
6. Anything which prolongs the study time is likely to enhance study.

The authors claim rich evidence for two powerful educational theories which combine to support the effectiveness of PBL. Firstly, PBL provides the opportunity for learners to identify and elaborate on their understanding, associated with tenets 1 – 5 above. But PBL also provides excellent conditions in which to develop more effective self-directed learning strategies which remain for a clinician's lifelong learning. Schmidt et al conclude that both combine to the benefit of PBL.

Regarding developing an individual's self-directed learning habits, one of social psychology's themes is to explore the mechanism and power of

motivation and social control. Belonging to a group of peers, is widely held to exert strong motivational force on an individual's behaviour, as explored by Maslow (Maslow 2013). When a PBL peer group has a vested interest in an individual contributing positively to the group's knowledge base and joint processes, there is strong social force on the individual to meet that expectation, and for the group to reward or punish an individual on the basis of their meeting that expected contribution. In addition, the human fear of being excluded from a group of one's peers, even aside from an individual's motivation to succeed at medical school, is a powerful unconscious motivator.

2.2.12 Competition and stress – PBL groups as emotionally protective

Competitiveness is one expression of strong social pressure, which can influence behaviour. Some UK medical schools believe that the competitive drive of their medical students may even be so high that it causes unhelpful focus on assessment rather than the learning itself. One recent small scale Brazilian, qualitative study of medical students, highlighted competitiveness among students as a top stressor (Pereira & Barbosa, 2013). However, students also identified time spent in class discussing their self-management with other students as very helpful. PBL groups are also likely to prompt students to talk about their study habits and expectations and to negotiate deadlines and workloads with their peers. This is likely to lead to learners gaining more insight into their own and other people's thinking strategies and their effectiveness, rather than simply knowing about a fellow student's final result.

2.2.13 PBL Learners and appropriate 'mind-sets'

When learners gain insight into the effort which others put into their learning it may enhance their mind set as discussed in the work of Carol Dweck, which has become widely adopted in school settings. Dweck (Dweck 1986) argues that if a learner believes that their abilities are fixed, they are less

able to motivate themselves to make efforts to achieve. With this mind-set, a learner fears failure and all negative feedback as this strongly reinforces helplessness. However, when a learner holds a belief that they can grow, they become more susceptible to learning itself. They are more likely to see feedback as information which helps them to more effectively direct their efforts to gain better results.

Dweck's work applies aspects of well–established psychological theory directly into the classroom. Theories of 'learned helplessness' (Maier and Seligman 1976) where an individual has learned to expect no reward from effort, and 'external locus of control' (MacDonald 1971) where an individual attributes success to random fate rather than focussing on aspects within their direct control, are both seen as impacting negatively on learning. Both of these unhelpful beliefs are discussed in the first longitudinal study of UK medical student stress (Firth 1986). The most common complaint about their medical training was 'feeling useless' perhaps a combination of 'learned helplessness' and attributing the cause to be outside their control. Feeling overworked was simply not mentioned.

2.2.13.1 PBL and Perry's taxonomy of intellectual and ethical development in College

William Perry's taxonomy of knowledge can be used to describe the stages of understanding with relation to knowledge which could apply to how students develop during their PBL study (Perry 1970). Perry posited that students begin with a basic duality; knowledge is either right or wrong. This position equips a student to memorise and apply received knowledge, but in order to reach higher levels of knowledge, further levels are required. Students may, with experience, acknowledge multiplicity, and the subjective nature of knowledge. They also come to realise that there are problems which have solutions which they can learn, and those for which there is no right solution yet known. This position can be uncomfortable, particularly for a field such as medicine, where so much hangs on the application of knowledge.

Students at higher levels of expertise, Perry suggests become able to develop procedural or relative knowledge which leads to a final stage of commitment to constructed knowledge. This procedure is iterative. Reflection is an important component of the final stage, uncovering what is not known.

The PBL process, through encouraging the contesting of knowledge by learners from the start as a way of building their understanding is well suited to developing students in this way. Personality theory has much to say about the differences in how people tolerate uncertainty, and how they choose to deal with ambiguity. For example in FFM, openness to experience measures are strongly related to desire for novelty and learning.

2.2.14 PBL and elaboration and reflection

The PBL cycle begins with recalling existing knowledge. This is considered by Ausubel (Ausubel, Novak et al. 1968) and other constructivists to be a crucial phase in learning: the activation of prior learning. Cognitive Psychologists also agree that practice in recalling learning is far more effective than passive review (Hogan and Kintsch 1971). The clinician also draws on their 'reflection in action' (Schön 1983), their store of expertise in order to practice their skill. A thorough training in this practice could establish effective study habits. In addition, it is thought that one aspect of acquiring diagnostic expertise and clinical reasoning is through building a mental bank of 'illness cases' and recalling these when the current case matches a stored pattern (Blissett, Cavalcanti et al. 2012). It appears reasonable to suggest that PBL assists in developing these 'illness cases' by exposing students to accounts of a range of cases covering the required syllabus. We might also expect that these first cases might be able to 'blueprint' or have a very strong impact, within which all subsequent exposure might be cognitively constructed.

2.2.15 Summary of Introduction to Education theory in the context of PBL

This section has provided a general overview of how PBL is currently regarded within Medical Education as a method of teaching and broadly located it within educational theory. It has considered some ways in which PBL situates learning within a social and simulated clinical context. The following section continues the discussion to look more closely at learning within a group through discussion.

2.3 Groups and learning

2.3.1 Introduction to learning in groups

This section of the literature review will define the aspects of learning in groups which have a close bearing on this study. The focus of this study is not on the specific composition of the PBL group itself or on who took a particular role for the case. However, all learning through discussion requires that individuals interact, both as speaker and also listener and participant within the group setting. There is clearly an impact on the individual from belonging to a PBL group. This review also acknowledges the relevant major exclusions of non-verbal data, facilitator and interactional data from this study. Next, it will present a summary of the current range of theoretical perspectives which influence behaviour in groups, particularly where the 'work' of the group is discussion and knowledge creation. Each approach will be briefly considered in relation to this study.

This section will then consider the implications of working in a group, considering group formation and numbers, and group roles and the impact, particularly on behaviour in discussions, within any group, then within the PBL group. Next productivity will be considered, first in any work group, considering process losses and the phenomenon of 'Social Loafing' and subsequently in PBL groups.

The review will then consider aspects of the process of learning through discussion, considering the role of listener and speaker, before considering what are legitimate 'products' of discussion. The review will consider issues in measuring their occurrence, particularly where it has been suggested that such behaviours are monitored and assessed in PBL.

This section will conclude in identifying the gaps where this study may contribute novel insight into how individuals learn through PBL discussions in groups before moving on to considering theories of learning style and

personality which seek to uncover more about these individual learners themselves.

2.3.2 Non-verbal analysis excluded from this study

The study is not collecting non-verbal data for PBL discussions. Darwin observed similarities in emotional expression between animals and man (Darwin 1998) and there have long been efforts to identify non-verbal skills in communication. Micro-expressions have even been identified (Ekman 2003) which convey our emotional states, even when we try to disguise them through our presented behaviour. However, students for this study did not consent to being videoed as well as recorded. Audio-recording-only consent precludes non-verbal analysis. Reading, interpreting and responding to non-verbal cues is of course a very important factor in discussion, but is simply not available to this study. Rich territory still remains; far more than is feasible to analyse within one PhD study.

2.3.3 Facilitator's speech and chronological interactions excluded from this analysis

The contribution of the facilitator is also excluded in favour of exploring the link between the individual learners and certain individual factors. In addition, due to the complexity of creating 45 individual corpora as the basis of analysis, it was decided to restrict this PhD study to analysing the products of a shared context student discussion. Sequential interaction between individual students will not be included in the analysis. Other analyses exist which focus on applying cognitive frameworks to discussions between PBL participants (Jamieson 2013), including a fellow PhD student, (Tokode 2016). Dr Tokode, is using a small number of student discussions from the same source for detailed case analysis. Future research projects may be able to analyse the interactions between individuals, in chronological sequence and cross relate cognitive categories to data from this study, such as personality type and measure of attainment.

2.3.4 Theoretical Perspectives for understanding groups

A rich diversity of perspectives exists on how we may consider a group. Below is a table giving just one summary of relevant perspectives, grouped according to their key approach and concerns adapted from Poole and Hollingshead (Poole and Hollingshead 2005).

Table 1 Theoretical Perspectives for understanding groups adapted from Poole &Hollingshead, 2005

Perspective	**Inputs**	**Processes**	**Outputs**	**Theories and concepts**
Functional: assumes groups are goal oriented and performance can be evaluated	Task, group structure, cohesion, environment	Interactions, information processing, conflict management	Group effectiveness, productivity. Quality. Satisfaction with group outcome	Function theory of group decision-making, groupthink, collective information processing, goal setting
Psychodynamic, Psychoanalytic and humanistic approaches focus on the emotional and unconscious processes and how they influence conscious processes	History of the group and member, unresolved problems, biological instincts	Facilitator – member dynamics, dependence and interdependence, group fantasies	Individual growth and development, group insight and understanding	Freudian psychoanalysis, humanistic psychology, psychodrama, field theory (Gestalt)
Social identity: explains group processes in terms of social groupings; in groups, out group	Structure of society, socio-economic class and social identity	Shared identity. Inter and intra group social comparisons. Development and maintenance of social norms	Conformity, loyalty. Collective action and protest	Social identity theory. Self-categorisation theory, group culture, social loafing.
Conflict-power-status – focuses on the dynamics of power and status in group interactions. Assumes that inequalities of power and resources generate conflict.	Individual status outside groups. Resources. Power and status within the group. Inter dependence of group members	Conflicts. Conflict management. Negotiation. Consensus building	Redistribution of resources. Achievement of member interests. Group performance, change in status	Political science. Social psychology. Social exchange theory, power – depended theory game theory

Symbolic-interpretive. An approach to understanding the social construction of groups and the use of symbols within groups to create meaning	Conditions or stimuli for symbolic interpretation	Fantasy. Dialectics. Sense-making. Metaphor, narratives, rites and rituals	Common vision. Group identity. Group boundaries. Group cohesion	Symbolic convergence theory, dialectical group theory. Group decision making. Groups as communities
Feminist: group dynamics and outcomes as a product of the difference in male and female power, privilege, motivations in social situations	Gender composition of group, individual gender perspectives, societal gender perspectives	Oppression and domination. Unequal participation. Male autocratic leadership. Male control-oriented communication	A focus on interpersonal relationships within groups. A focus on the division of power relations in groups and society	Oppression and domination. Gender construction. Individual difference theories. Structural inequality theories
Social network: groups are interlinked structures in larger social networks. Groups are seen of patterns of relationships (ties) among members (nodes)	Member characteristics. Pre-existing networks. Network structure and properties	Affiliation Exchange diffusion information flow	Task effectiveness and efficiency. Cohesiveness. Attitude and belief convergence	Graph theory and sociograms, social cliques, self-interest and social exchange, network ties, homophily
Temporal – groups change with time and group processes unfold in time as experiences accumulate	Group members' constructions of time. Amount of time available	Group development, better performance over time. Project completion, time management		Sequential, cyclical and punctuated equilibrium. Structuration theory. Action theory, complex systems
Evolutionary. Small groups are located between genes and culture. Natural selection has shaped evolution of groups	Individual-inherited tendencies for group behaviour	Co-operation altruism. Nurturing and teaching	Survival. Reproduction	Evolutionary theory. Socio-biology. Group selection. Reciprocal altruism. Pro-social behaviour

2.3.4 Considering each of these Theoretical Perspectives and their relevance to PBL discussions and groups

This study draws on all nine of these theoretical perspectives which will now briefly be discussed.

The PBL group is assembled to fulfil a function; the achievement of learning objectives for each case.

The psychodynamic perspective perhaps comes closest to suggesting how an individual may come to behave in a certain way during discussions. Personality theories and measures claim that different personality types behave differently in relation to others and show differences in their speech behaviour. Theories of language acquisition relate to how children develop language as conditioned and modelled strongly in their home environment, within their original small group, the family (Cashdan and Grugeon 1972). Indeed, many treatments of sub-optimal family conditioning, focus very strongly on talking and listening. For example, either within a strong re-parenting dyad with an individual therapist, or within a group setting, where the conflicts and lessons of a group member's previous family life may be re-explored in a new setting in order to identify patterns of behaviour with others which no longer serve a constructive purpose in adult life (Anderson and Goolishian 1988).

Theoretical perspectives relating to the social identity of groups helps in examining how they are formed and maintained and the roles individuals take. This includes considering how individuals contribute, and the phenomenon termed 'social loafing'.

Through counting the number of words spoken by individual students and the facilitator, within a group over the same time span, issues relating to power and dominance or surgency can be quantified and analysed.

Individuals convey their meaning literally through language, but also, symbolically and often sub-consciously, through the sub-text; that which is 'beneath' what they

say but which may be interpreted, from what they say. Language makes use of common tropes or underlying ideas which contribute greatly to our understanding (Lakoff and Johnson 1999). For example, it is commonplace to talk of speaking in relation to common bodily experiences, such as eating. We chew things over, we spit words out, and we swallow our words. By analysing the semantic content of student talk, it may be possible to identify literal talk both inter-personal and cognitive, but also to uncover hidden interpretations.

Through examining corpora divided by gender, the feminist perspective will be of great use in providing a theoretical framework within which to interpret findings.

Each PBL group is one group, working for a period of time within the overall cohort. Students and facilitators are interlinked over time in patterns which tie them together, as individual 'nodes' within a system. Methodologies, such as sociograms and increasingly computerised ways of mapping data can uncover patterns relating to interaction, including talk. However, as already noted, this has been identified as a future research possibility.

In considering the progress of the PBL discussion through one cycle, a temporal perspective becomes important. In considering the selection of the suitable time frame to select, issues related to group formation and social identity are key.

Human groups are fundamental to our evolution. Genes and their expression through culture and opportunity are crucial determinants of how individuals behave in groups. Personality and learning styles are one way in which behaviour can be theorised which will form an important part of this study.

2.3.5 Phases in group formation

It is well-established that groups need to grow to maturity by successfully completing phases of working together, described by Tuckman and Jensen as forming, norming, storming and performing (Tuckman 1965, Tuckman and Jensen 1977). The forming stage is necessary for getting to know the individuals within the

group, establishing social bonding. A norming and storming cycle ensues, where norms for behaviour and performance are identified and perhaps challenged until differences in these domains are resolved and become established. Secure in its identity, a group can then enter a performing phase until new challenges or misunderstandings arise which require them to attend to group processes and re-storm and re-norm before continuing effectively as a work group. Bales and Johnson and Johnson (Bales 1970), (Johnson and Johnson 1987) long established that cooperative learning where groups reflect on their learning processes are likely to be more successful. The discussions of a group in the 'performing' phase is characterised by having relatively low periods of conflict and misunderstanding.

The sampling of PBL discussion for this study was designed to take account of this cycle, by ensuring that the first year groups were experienced and well-established before the recording took place. Within a GEMS cohort this is likely to be particularly important as learners are from diverse educational and age backgrounds in comparison with the usual homogeneity observed in undergraduate medical cohorts.

2.3.6 Group size

Recommended group sizes vary according to purpose and study. Five are considered sufficient to constitute a cohesive group (Bruffee 1999), with up to ten participants providing opportunities to develop communication skills, knowledge and collaboration, (Benson, Noesgaard et al. 2001) with (Barrows 1988) giving an upper limit of 10 for effective functioning in the PBL context. Of course, it is inevitable that as group size increases, participation by individuals in the group decreases (Pavitt 1994). When the work of a group is discussion, each individual is in competition with the rest in order to participate through being the individual who is speaking, although all may contribute at the same time by listening.

2.3.7 Group roles and tasks

People adopt roles in groups which gain certain social rewards, categorised according to Thibault and Kelley (Thibaut and Kelley 1959) as follows:

- Prescribed roles – conforming to societal norms: ie. a good student, a suitable future doctor
- Subjective roles – the individual's perception: ie. I'm a helpful and supportive person
- Enacted role – behaviour displayed: ie. offers help to fellow group members
- Functionally required: ie. somebody has to do it

For example, it may be that by taking the role of Chair or Scribe, a student may make more interventions or take more turns in conversation than if they are simiply in the role of themselves as student.

Penland and Fine (Penland and Fine 1974) identify three broad categories of role, identifying group nurturing, personally gratifying and task oriented focus for roles people may choose, as follows:

- Group building and maintaining roles
- Individual roles
- Group task roles relate to the task the group is working to complete.

2.3.8 Brookfield's conversational roles in the educational setting

Brookfield extends an analysis of roles in the educational context of small group discussions. He describes a longer list of roles which may at times be helpful to the efficacy of learning through discussion. He advises the learning facilitator acts as a role model, who demonstrates, comments upon, and then inceasingly expects the group to adopt these behaviours(Brookfield and Preskill 1999, Brookfield and Preskill 2012).

Table 2 Adapting Brookefield's conversational roles to relate to PBL discussions

Brookfield's role	**Description**	**Pbl equivalent**
Problem, dilemma or theme poser	Introducing the topic from own perspective	All
Reflective analyst	Keeps record and gives summary every 20 minutes, issues and common themes	Chair
Scrounger	Listens for helpful resources, suggestions and tips and summarises at end	Scribe
Devil's advocate	Listens for emerging consensus and explicitly challenges to challenge group-think and encourage alternative considerations	Facilitator/Chair
Detective	Unchecked, unacknowledged biases, particularly prejudicial, power related blindness	Any
Theme spotter	Identifies what remains to be explored for next session	Chair/Facilitator
Umpire	Listens for judgmental comments or contraventions of groundrules	Chair/Facilitator
Textual focuser	Identifies when point lacks grounding in evidence and seeks clarification	Any
Connector	Shows connections between participant's contributions	Any
Summariser	Observations drawing together several people's contributions	Any
Appreciator	Says when contributions are useful or interesting	Any
Questioner	Draws out or extends what others say	Any
Speculator	Introduces new possible interpretations and ideas: what if?	Any
Active listener	Paraphrases other rcontributions	Any
Underscorer	Emphasizes relevance, accuracy or resonance of another person's comments and says why	Any

2.3.9 Brookfield's roles and their impact on learning through discussion

The descriptions and categories above are all presented as potentially beneficial behaviours when performed with great skill and emotional intelligence and which individuals may offer within the group learning setting. As will be discussed within the section on personality type in the PBL setting, some of the roles are much more likely to be preferred by some personality types or learning styles than others. For example, the appreciator role is likely to be demonstrated by learners who score more highly on what FFM terms agreeableness. Of course, any of these roles could be inimical to learning if performed clumsily. For example, persistent questioning and interuption of a fellow learner when they are delivering the results of their self-derected learning sessions, could lead to ill-feeling, even where the knowledge content needs to be critiqued to assist group understanding.

2.3.10 Derby GEMS: PBL roles decribed

In Derby GEMS, there are typically five roles which learners may adopt during one cycle. Students rotate the role of Chair, who takes the lead in ensuring that the group collaboration progresses during the time available. Learners also take it in turns to be scribe – with whiteboards and coloured pens available around the PBL room where the scribe records salient points of discussion and decision to support the progress through that case. These notes remain throughout the case and students often photograph them and share for their notetaking information. These are examples of the more formal, functional roles described by Thibault and Kelly (Thibaut and Kelley 1959).

During session two, Derby GEMS students often role play a consultation between the patient and doctor, in order to fully explore the clinical scenario and its bio-psycho-social aspects. In some groups the facilitator may take one of these roles. In other groups, learners play both roles. In other groups, all learners can make suggestions as to what would be of assistance in the role-played consultation. However,

whatever role a learner is taking, it is likely that their particular personality will be evidenced in the way that they speak in the PBL group.

2.3.11 Facilitators and group discussion

Heron (Heron 1999), identified three different modes of facilitation. A hierarchical mode is characterised by the facilitator taking a more authoritative stance, whereas a cooperative mode operates more in partnership with the group. In the autonomous mode, the facilitator chooses to support the group more indirectly, choosing to hold back and allow them to solve their own issues. Heron categorised facilitator interventions in terms of the following:

Prescriptive:	seeks to direct behaviour
Informative:	gives information, teaches something the facilitator believes the recipient didn't know
Confronting:	identifies something in order to heighten awareness
Cathartic:	enables the expression of strong emotions
Catalytic:	encourages questioning
Supportive:	affirms the worth and value of clients (particularly praise)

Figure 2 Heron's facilitator behaviours

2.3.12 Facilitators and group discussion in PBL

Applying Heron's model to PBL facilitation might lead to lower levels of spoken contribution by PBL facilitators when compared to other discussion group models. Some comparisons are considered in the corpus section of this literature review. Wilkie discusses a range of approaches, broadly recommending that the PBL faciitator serves the group best when considered to progress swiftly through the styles identified by Heron (Savin Baden and Wilkie 2004). In authorative style the

facilitator may model the PBL process in order to establish a cooperative learning partnership, where the facilitator may spend extended periods with an experienced group where they are adopting a strongly autonomous style, intervening seldom.

It might be expected that student discourse in PBL discussion may take on some of the characteristics of facilitation identified by Heron, particularly when in the role of Chair, or in the role of 'teacher' presenting knowledge which they have researched on behalf of the group. Corpus analysis of semantic categories and parts of speech may be helpful in identifying some of these aspects in the student corpus.

However, this study redacts the utterances of the facilitators and so whilst acknowledging that the skills of the facilitator may have an impact on how student discussion plays out, it is not the focus of this study. The total proportion of facilitator speech will be reported for each group in the appendix, but the data results and analysis will focus on the student utterances alone. As noted elsewhere, other studies focus on the interaction between facilitator and learners in constructing knowledge through PBL discussions (Jamieson 2013) (Tokode 2016).

2.3.13 One theoretical model for quantifying the influence of PBL processes on achievement

Schmidt and Moust summarise a body of research from Maastricht, where the factors which affect small group PBL processes have been studied in medicine and other curricula (Schmidt and Moust 2000). A number of these studies have a close bearing on this PhD study.

The authors build on the earlier theoretical model of PBL proposed by Schmidt and Gijselaers, (Schmidt 1990) in order to create a theoretical model which quantifies the elements within the PBL process in order to show their interdependence. Schmidt and Moust then seek to quantitatively measure the strength of these influences between elements. Their figure is adapted below (Schmidt and Moust 2000) p. 29.

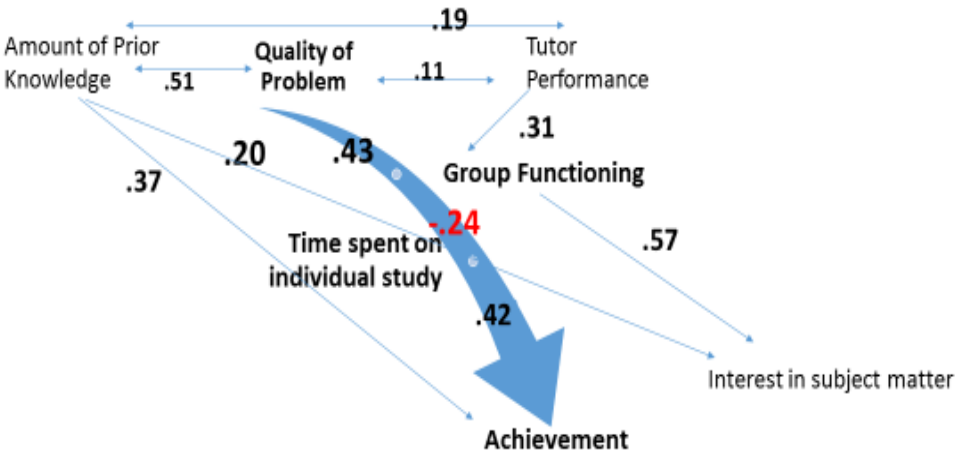


Figure 3 Schmidt and Moust, 2000 Model demonstrating interdependence of PBL elements

The model theorises that the crucial elements of PBL are:

- The amount of a learner's prior knowledge
- The quality of the PBL problem itself
- The performance of the tutor/facilitator
- The way the PBL group functions
- Time spent on individual study by the learner

The model suggests that the effectiveness of the learning may be identified by the learner's achievement in an end of course test, but also in an independent measure of the learner interest in the topic under study.

Data were gathered by combining results from the end of course test (achievement) with the results of combined Likert scales, such as the average ratings given by

students to a number of statements about the quality of the PBL problem. The reporting mechanism for group functioning is not given. In their calculations of influence, the authors used structural equations modeling (Bentler 1989) to identify the extent to which data fitted the model. They identified positive influence for each dependent relationship, save for one: group functioning, reported as having a negative influence on the time spent in self-study -0.24. The authors conclude that enhancing the quality of problems would have a greater impact for less effort than influencing tutor performance.

Moderate influences were found between the quality of the problem, 0.43 and the way the group functioned. In addition, group function was found to have influenced the final measure of achievement of 0.42. However, group functioning was found to exert 0.57, the largest measured positive score on influencing the students' interest in the subject matter. Interest was measured by the number of students who signed up to an optional lecture on the topic after the PBL discussion. There are widespread studies linking student satisfaction and motivation with PBL methods.

Here is an attempt to quantify that PBL groups strongly influence student interest in the subject. However, in this model it also seems that group functioning has the potential to produce a negative effect on the amount of time for self study, a substantial component of likely successful achievement, in the light of the 'time on task' learning approach, (Slavin 2003) and the work of Ericsson in studying the 10,000 hours of focussed practice which he associated with exceptional levels of achievement across many fields of learning (Ericsson 2016).

2.3.14 PBL discussions and their influence on the tutorial group

Dolmans and Schmidt (Dolmans and Schmidt 1994) discuss what drives the student in PBL and identify the influence of the discussion in the tutorial group as their first theme, representing four questions out of twenty they include in their questionnaire.

1. The discussion in the tutorial group determines to a large extent what I will study
2. The tutorial group discussion is an important stimulus for my learning activities during self-study
3. The learning issues generated are the most important starting point for my learning activities during self-study
4. I study to a large extent independently from the learning issues generated

Figure 4 Dolmens and Schmidt's Theme 1: Influence of the discussion in the tutorial group

Students were then asked to profile their preference using these statements as customised Likert scales to determine how they orient themselves as self-studying. Ironically, a PBL study group is not a means of self-study by itself. It is the opposite, and each individual has to submit their individual study time and effort to some extent to the requirements and tastes of the group. The Dolmans and Schmidt study showed that the influence of the study group as opposed to other sources of stimulus for self-study rose in later years, rather than in early years, and concluded that it was because participants became better self-directed learners as they progressed within the curriculum.

2.3.15 Social reinforcement of speaking

There is research evidence to suggests that timing of verbal contributions in the ongoing interaction is of importance as well as the sheer quantity (Schroeder, Rakos et al. 1983). Bandura (Bandura 1986) and Vygostky (Harland 2003), both advocate understanding the strong impact of socialisation on the behaviour of others, both as role models, and in terms of societal norms and social rewards. These are strongly at play in the PBL discussion group, where the needs of the individual depend to a

large extent on successful collaboration within the group to produce more or better quality knowledge than could be acquired alone, through self-study. It is likely that some personality types, may be more aware of the process losses which inevitably occur when working in a group.

2.3.16 Participation in PBL discussions linked with attainment

Experiments with High School students (Schmidt 1988) demonstrated that groups which discussed problems in groups performed better in recall tests than those who analysed alone. In one study Moust et al, attempted to identify whether the amount of participation in small group discussion had an impact on achievement (Moust, Schmidt et al. 1986). They demonstrated that quantity and perhaps more surprisingly, quality of contribution were unrelated to student achievement in their study. They found no correlations between elaboration and achievement. Indeed, they found a slight negative correlation between the elaboration and achievement score on the recall and completion test.

2.3.17 Covert elaboration

However, overall, the variation between participation was larger than the variation between the achievement scores, implying that learning which was not verbalised was taking place. They accounted for this finding by describing 'covert elaboration' whereby silent students cognitively elaborated the learning, but chose not to verbalise that process in the discussion. When online, this process is often termed 'lurking', and will be discussed shortly in connection with the phenomenon termed 'social loafing'. However, in any discussion, a key component is that there are more listeners, following the contributions of others, than there are those making a contribution. One might argue, that discussion itself is founded on covert elaboration to a large extent. In other words, the quality of a discussion lies in the quality of thought for its participants not just those which are verbalised.

In addition, it is important to point out in the light of subsequent findings relating to reductions in the quantity of speech by females, particularly in mixed groups, that

the groups were not gender balanced, indeed groups were mainly female, save for a male 'leader'. This replicates conditions in which other studies have found significant reductions in the amount of talk from females, which will be discussed shortly (Duek 2000, Tannen 2001). Moust et al. related the participation finding to data which has shown that giving explanations (or help) is linked with higher achievement, but also 'receiving help', when it is the right help for that learner, is also linked with higher achievement. (Webb 1980). Receiving help in this context, sounds remarkably similar to the basic paradigm of teaching: where one person with knowledge, gives a correct explanation to another when that knowledge is necessary and timely.

This finding relating differences in verbalising thoughts, suggests that a willingness to verbalise, often linked to personality differences between extraversion and introversion, may not impact on an individual learner, however, that does mean that the group as a whole may miss out on elaboration and so one channel of learning is diminished. Schmidt et al state (Schmidt and Moust 2000) 'the collaborative effort by which students help each other in clarifying the issues, is a central element of problem based groups' (p. 24). If one of the key tenets of PBL is that the group together elaborates more knowledge than the lone learner and thereby identifies only those gaps which exist for the group as a whole, whilst benefitting from sharing the knowledge they do vocalise, then 'covert elaboration' is effectively withholding a resource from the group. Jamieson's study of PBL discussions as a community of inquiry demonstrated that some aspects of critical thinking became possible within the PBL group, and that as a rule, individual utterances demonstrated one stage of the critical thinking framework (Jamieson 2014).

If the findings are still emerging when considering exactly how individual learners construct knowledge together, perhaps it is not surprising that a simple measurement of contribution remains the most obvious measure of participation indicating learning. There are limitations to this measure, concerning, most obviously the quantity and quality of contribution to which we now turn.

2.3.18 Not participating: Social loafing, withdrawal, Social loafing or Free riders

When humans measured power in horse-power, it was observed that two horses in harness had less than two horse power, and that this process loss persisted when replicated by men pulling on a rope (Ringelmann 1913).

1 horse = 100%
2 horses = 93%
3 horses = 85%
8 horses = 49%

Figure 5 Ringelmann's horse power findings

This situation neatly depicts the 'process losses' which occur when individuals are required to work together. Thus, for an eight student PBL group, Ringelmann's experiment suggest that for the simplest of tasks there might be a process loss of around 50%.

In addition, individuals differ in the amount of effort or contribution they may wish or be able to make. In PBL groups, educators and students have had concerns about 'social loafing' or people who might take advantage of the group situation in order to benefit disproportionally for their individual effort. One simple measure which has been used, is to assess participation – the quantity of how much students say during PBL tutorials. Learners who take may fewer turns or participate less are in danger of being identified as a social loafer, or a free-rider who benefits from others' study but does not contribute their own knowledge to the group. One study, where a problem based learning approach was used for students studying interior design in Taiwan, specifically examined the 'free rider' situation (Huang, Yang et al. 2007) in correlation with Kolb learning style (LSI) and teacher interventions.

Teachers and learners were asked to assess student contributions using the same criteria.

Table 3 Assessment categories used by teachers to assess student contributions Huang et al. Grades A – D for contribution, within the four categories, speaking, tone, attitude, announcement

Appraisals	Grade A	Grade B	Grade C	Grade D
1. Speaking	Renewable speaking	Often speak	Sometimes speak	Do not speak
2. Tone	Approve and encourage others	Approve others	Coldly	Argue to win
3. Attitude	Humble and decisive	Time control and efficiency first	Critical	Arrogant
4. Announcement	Make constructive comment	Speaking with words logic	Unclear consciousness	Irrelevant

Learners with free-rider inclinations were identified as being learners with a higher self-evaluation than that observed by their peers and teacher and with having lower results. Free riders also preferred more positive interventions from the teacher figure and were less interested in discussion unless it is about the learning environment and assessment. By contrast, results suggested that KLSI, assimilator style learners were not likely to have free-rider inclinations and were associated in this study with higher grades. In general, assimilators prefer thinking and listening and are good at summarising knowledge. This style is not necessarily associated with making many verbal contributions to a discussion, but this alone was not identified with free rider behaviours.

These findings suggest a complex picture of association between the quantity and quality of speech behaviour during discussions which relates to previous findings underlying education. For example, learners have been shown to behave differently according to whether they are oriented strategically towards assessment, (Marton and Saljo 1976) and that many learners with little experience, lack accurate self-perception of their performance (Papinczak, Young et al. 2007).

2.3.19 Positive indicators for behaviour in groups

In one study, the top three dysfunctional behaviours observed within PBL discussions were the quiet student, the dominant student and the late student, which were thought to contribute to a disorganised PBL learning experience. Both students and facilitators agreed that these were also difficult to resolve successfully within the group (Hendry, Ryan et al. 2003). Lack of interaction did not really have a large impact on the effectiveness of their study within the PBL group, but students believed that tutors could influence this factor. Low motivation was not felt to happen often, but when it did, it could have a big impact and students felt tutors could influence it.

It is interesting to consider the professional context within which the medical student will ultimately communicate. Dickson et al., (Dickson, Saunders et al. 1993), discuss the 'Interpersonal Professions' described by Ellis and Whittington (Ellis and Whittington 1981) where the majority of the professional day is spent in face-to-face interaction with others, and a major component of the service they offer is achieved this way. Doctors clearly belong to this group, with some sub-groups such as surgeons and emergency medical staff acting in circumstances where this flow of communication is impaired or happens at intervals before or after treatment. Faraone and Hurtig found that one of the hallmarks of skilled practitioners was that they interacted in ways which reduced uncertainty and reacted positively to the other (Faraone and Hurtig 1985). We might expect to observe similar behaviour developing in PBL discussions. We might also expect to observe that some

personality types, particularly those noted to be more oriented to the needs of others, such as scoring highly on agreeableness, might be expected to demonstrate this to a higher level or more frequently.

2.3.20 Online discussions

Online data has the potential to produce rich self-documenting corpora which may be studied through corpus analysis and so will be discussed in that section of the literature review. However, evidence is emerging about how students participate in discussions online. Some findings suggest that in an environment where it is possible to think before contributing and not have to openly compete for listeners, that learners who are quieter in face to face environments are able to contribute more. For example, Blumer finds a significant link with higher levels of neuroticism, being associated with higher online contributions, rather than introversion or any other measure (Blumer 2012). In the online environment, perhaps skills for participation in discussions may be differently distributed and thoughtful, socially anxious learners may become more able to contribute.

In a study of 'learning sets' through their online discussions, Sackville and Sherratt identified marked differences in styles of participation between groups (Sackville and Sherratt 2006). The authors concluded that one influential factor related to the crucial importance of the initial postings. They recommend that the facilitator models the style and type of posting which they wish to encourage strongly and at an early stage. In order to establish an effective online discussion environment, facilitators need to provide clear induction activities in order to replicate the usual 'forming and norming' phase for learners. The potential benefits of learning online may also need to be reinforced in order to encourage effective group performance in the online environment in the absence of the usual social reinforcements of the face to face environment.

2.3.21 Summary of interaction and learning

So far, the section has considered the roles individuals may take in discussion and the interventions that may be made by a facilitator or a learner taking on one of many roles associated with discussion. Discussion has identified some of the interdependencies of the constituent parts of the PBL process and how the group interaction may influence learning. It is clear that there is no straightforward relationship between contribution and attainment, but that there can be a strong relationship between group interaction and interest in the topic, and motivation to study, although this may have potential for a weak negative impact on self-study. Various studies have identified effective performance in teams with forming an effective functional identity which enables team members to contribute and to avoid dysfunction and conflicts, particularly 'lateness' and 'free-rider' behaviours. Future clinicians will be required to develop skilled communication skills which will be well-used throughout their working day and increasingly in online environments as opposed to face-to-face.

2.3.22 Factors associated with success in medical school

There is mixed and patchy evidence of relevance to identifying the factors which may be associated with success at different stages in medical school. A systematic review was conducted by Ferguson et al (Ferguson, James et al. 2002) which identified thirteen studies prior to 2000 which considered academic performance, personality, gender, ethnicity, learning styles, selection interviews, personal statements and references as possible predictors of academic success in medical school. Meta-analysis was possible on the links between 'A' Level and subsequent medical school performance, which was found to account for 23% of the variance in medical school performance, but only 7% at post-graduate level. The Kolb convergent learning style, which focuses on hypothetical and deductive reasoning, was associated with the highest levels of academic attainment in the studies they reviewed where Kolb's learning style was considered. There was data suggesting that dominance, one of the facets associated with extraversion, was associated with

higher scores in oral exams. The authors noted that most studies focus on the early medical school years and larger longitudinal tests would be needed to explore longer term performance. All studies demonstrated that both very high and very low anxiety states were associated with poorer performance, which they noted was in line with well-established arousal theory (Yerkes 1908). The authors noted the difficulty of reconciling the variety of personality measures and concluded that the most fertile ground for investigation was likely to lie in the neurotic sub-trait, anxiety, and conscientiousness, particularly the facet, 'striving for achievement'.

2.3.23 Evidence comparing the attainment of UGMS students and GEMS at Nottingham

In a study of the initial cohorts at Nottingham GEMS in comparison with the UG entry medical course, Manning and Garrud (Manning and Garrud 2009) were able to show comparable levels of achievement across measures of attainment. GEMS students outperformed UGs in their percentage of completion for the course. UG students outperformed GEMS on some MCQ knowledge tests. The authors identified that this may be due to the lower knowledge base at the start of the course for scientific subjects for some GEMS students. In the Netherlands, Leivens (Lievens, Ones et al. 2009) found that in the first year of medical school the previous academic attainment level was predictive of higher knowledge test scores, but that this diminished in subsequent years.

Research is continuing to identify other factors which may be more valid in predicting later clinical performance. One review of the literature for the Institute for the study of Labour, (Brunello & Schlotte, 2011) links economic success and achievement in the complex modern economy with non-cognitive and personality measures as well as cognitive measures.

Johnson (Johnson 2015) concludes that emotional intelligence is crucial in modern medical education, basing her argument on the demands on medical practitioners to have highly developed communication skills, and the ability to read and respond to

the emotions of patients and co-workers in such a way as to deliver an effective service with the minimum of conflict.

The discussion now turns to the issue of quality, and considers what type of contributions have been observed in PBL discussions and which are associated with positive outcomes.

2.3.24 What do learners actually say in PBL groups?

Visschers-Pleijers et al (Visschers-Pleijers, Dolmans et al. 2004) analysed interactions during the student reporting phase of PBL and identified the following types of interaction:

- **Asking for, giving and receiving explanations**
- **Reporting by giving and receiving explanations**
- **Asking content-related questions**
- **Integrating and applying knowledge**
- **Structuring, relating and summarising information**
- **Applying theory and providing examples from practice**
- **Discussing unclear information or different opinions with regard to learning content**
- **Discussing a variety of literature resources and opinions**
- **Discussing disagreements and different interpretations**
- **Guiding and monitoring the tutorial group discussion**
- **Guiding and monitoring the content of the discussion**
- **Guiding and monitoring the group process**

Figure 6 Visscher's et al. interactivity categories and themes, discussed in the focus group interviews

These are all positive categories which may be applied in a positive social context and which can contribute to knowledge construction, but real PBL discussion corpora may reveal other categories, or spells of ineffective contributions within discussions.

2.3.25 Interactions within PBL discussions

There are many possible ways in which interaction and efficacy can be recorded and observed. One checklist used by Tipping et al (Tipping, Freeman et al. 1995) monitored group working processes under the following headings:

1. Interaction
2. Lines of communication
3. Who has the power – pairings, subgroups?
4. Distribution of participation
5. Impact of group size on interaction
6. Balance of task and group–building roles
7. Are people listening and building on the ideas of others?

However, as previously noted, compiling individual corpora has been privileged in this study, leaving study of chronological patterns of student interaction (1 – 3) above for future research. Item, 4 will be quantitatively measured and considered between all individuals within their group and compared to the study cohort as a whole. As already noted, as group size increases, participation decreases, but as almost each PBL group is the same size, this is not a consideration in this study.

It may be possible to consider data relating to item 6 and 7 through the corpus analysis of student talk.

2.3.26 Patterns of collaboration and participation

Bales studied interactions with a tool which categorised basic interactions, interaction process analysis (IPT) in 1950 (Bales 1950)The emerging field of social network analysis (SNA) aims to describe and measure the relationships between

individuals (O'Malley and Marsden 2008). Programmes can be used to identify the agency of individuals within groups based on data. Future study may be possible to apply simple forms of social network analysis using the corpus data, ascribed to individuals to identify further patterns which may be related to efficacy of learning. A lack of research has been identified into the effectiveness of collaboration skills in the PBL process (Hmelo-Silver 2004) and this may be a useful future research direction as it would allow us to see, who collaborates with whom.

Azer boldly claims that interaction has the highest weight in predicting group success (Azer 2009) and PBL has long been identified with opportunities to deepen learning and develop critical thinking (Hmelo-Silver 2004) but many ineffective practices may impede opportunities. In their PBL study, de Grave et al (Willem S. De Grave, Dolmans, & van der Vleuten, 2002) found that students perceive problems in interaction, motivation and difficult personality as impeding their learning in PBL. However, they also perceived that there was not much that the tutor could do about the personality issues which happened seldom and did not have a big impact on their learning.

2.3.27 Links between discussion off task and achievement

One small scale study used transcription to access one cycle of a variant of PBL, including self-study periods (Yew and Schmidt 2007) with utterances classified according to Van Boxtel (Van Boxtel, Van der Linden et al. 2000). Significant results identified that students who spoke off task the most, achieved lower scores on the test given at the end of the day. Students who asked many questions were also likely to score lower indicating that at this level, perhaps questioning basic facts indicates a lack of preparation and uncertainly about material rather than the use of higher order skills to interact more deeply with knowledge. The largest proportion of utterances, almost half, concerned basic factual recall rehearsed out loud. This is a legitimate activity in the first session of PBL when prior learning is activated, but

perhaps if it continues, particularly during the self-study periods; it indicates a superficial approach to learning.

2.3.28 Differences in question asking behaviour in educational groups

Differences have also been repeatedly found in the behaviour of question asking between males and females, (Hawkins and Power 1999). Azer (Azer 2007) suggests that action verbs could be used as a proxy for the thinking outcomes of individuals – in the way that Bloom's taxonomy has been guiding educational design for decades. However, the cognitive level of questions is not simply equivalent to the initial question word used, but relates to the way in which it is used. Azer calls for more research to understand the interactions during PBL tutorials and to measure the quality of those interactions and their impact on learning.

2.3.29 Turn taking in group discussions

There are strict social norms which organise turn taking, (Sacks, Schegloff et al. 1974) which may interact with personality type to determine who speaks when. PBL groups are comprised of apparent equals in social status, so turn taking norms may be derived from the context and be more strongly related to the individual's personal need to speak. The facilitator still has the authority of the teacher and so is usually able to take a turn whenever they wish without competition, hence the need for facilitators to monitor their contributions carefully in order not to dominate discussions and render the pattern of interaction similar to the 'normal' teacher rather than learner led classroom.

2.3.30 Valle's Tutorial assessment rating scales on group interaction and active participation

Valle's tutorial assessment scale is another attempt to classify what types of interaction are effective within the group setting and what exactly constitutes interaction or participation within this setting.

Group interaction scale

1. Accepts suggestions about his/her work
2. Accepts decisions made by the group
3. Adjusts to different group roles
4. Listens attentively to peers
5. Shows respect to peers (no interruption, attack, acknowledges other contributions)

Active participation scale

6. Helps peers to clarify ideas
7. Participates in case discussion
8. Shares knowledge with group
9. Gives feedback

Wun et al (Wun, Tse et al. 2007) used Valle's scale (Valle, Petra et al. 1999) in their study at Hong Kong University Medical School. They videoed and analysed small group discussions comparing student participation in 43 small group discussions using their traditional curriculum approach and compared it to 46 new PBL tutorials the following year. Wun's study found that students who had studied through the PBL curriculum talked significantly more. In only two (4.7%) of 43 tutorials in the non-PBL curriculum did the students talk more than the tutors; but students talked more than the tutors in 17 (37.0%) of 46 tutorials in the PBL curriculum. PBL students scored significantly higher than non-PBL students in almost all items of the tutorial assessment scales, and in the mean scores of both the group interaction scale (items 1–5) and the active participation scale (items 6–9). One exception was that the PBL groups demonstrated lower scores in showing respect to their peers.

2.3.31 What are silent PBL students thinking about?

Geerligs (Geerligs 1995) sought to identify what cognitive processes may be occuring during PBL which were not vocalised, through sounding a buzzer to prompt participants to note their thoughts at the moment the PBL process was interrupted. Through 'thought sampling' ten PBL health sciences groups over ten sessions, the results identified 74% of student thoughts were categorised as task related, with irrelevant, or miscellaneous or absence of thoughts accounting for the remaining 16% and 10% respectively. Moust's study (Moust, Schmidt et al. 1986) provides evidence that silent students are not necessarily less task oriented than talkative students. De Grave et al (De Grave, Boshuizen et al. 1996) studied cognitive and metacognitive processes through stimulated recall where students watched themselves on video and recalled their thoughts at the time. Recalled thoughts closely mapped onto those expressed to the group which suggests that there can be a close relationship between what is thought and what is vocalised.

2.3.32 A person centred model to describe PBL groups

Adopting a humanist and person-centred approach to learning in groups, Clouston and Whitcombe (Clouston and Whitcombe 2005) advance a framework for considering how the individual learner is situated in the PBL process. They propose an overlapping petals model, which seeks to describe the influences which must be in place in order to facilitate learning by the person at the centre. It is implicit that the needs of each individual would be slightly different. Learners with widely differing personalities would have differing attitudes, approaches and value systems which may mean they need to adapt to achieve successful learning within the PBL method.

2.3.33 Overlapping concerns of the person-centred model

- broader environmental factors;
- student and tutor participants' readiness to engage in the PBL process;

- congruence between the individual's attitude, approach and value systems in relation to educational philosophy;
- group dynamics; and
- communication strategies used by group members

The authors propose that the extent to which an individual is comfortable within the group setting and managing the inevitable conflicts inherent in group dynamics has an impact. Individuals also have different communication strategies. It is also crucial that an individual is comfortable with identifying what they do not know in the company of their facilitators and peers, even though this is often a source of anxiety for learners. This is one of the very few frameworks to consider the individual within the PBL group.

2.3.33.1 Gender differences in group discussions

As will shortly be discussed in connection with research in corpus analysis, gender has been a productive lens through which to consider difference in communication. Strong evidence for gender differences exists in relation to group behaviour and settings related to authority and performance, such as in the workplace, where females continue to earn less and hold less powerful positions than males (Bishu and Alkadry 2016). For much of the research in the period under consideration most facilitators will have been males, and for much of that time, males will have comprised the majority of medical students. However, there has been a substantial change in the proportion of females studying medicine in Europe and the US, and now females form the slight majority of medical students, indeed, of all students in higher education.

One detailed study, by JodyLee Duek (Duek 2000), examines issues of gender and ethnicity and behaviour of students in PBL discussions. For this PhD study the ethnicity of learners was not available, but gender is known and therefore considered. Duek considers group values in the context of Johnson's differentiation between competitive, individualistic and cooperative values within group approaches

(Johnson and Johnson 1987, Johnson and Johnson 2008). Social interdependence and the adoption of cooperative learning processes which contribute to constructive socialisation, can be seen as central to group function in PBL discussions. Duek identifies the following criteria as being fundamental:

1. Positive interdependence: each learner is concerned about the performance of all learners within the group
2. Individual accountability: each student's mastery is assessed.
3. Heterogeneity: members are diverse in terms of ability and personal characteristics
4. Dispersed leadership: all members need to perform leadership actions
5. Developing social skills: all members need to possess high degrees of communication skills, trust, conflict management and leadership, explicitly and implicitly taught.
6. Reflection: time and procedures are needed for dealing with group processing and evaluating and modifying personal and collaborative effectiveness.

2.3.34 Social constructivism and interdependence in PBL groups

Johnson's work identified that negative interdependence, i.e.: that an individual's efforts are rewarded at the expense of the group, would foster a competitive and unhelpful learning environment for group work. PBL, where the individual's knowledge and skill, is individually assessed, clearly may be in danger of fostering negative interdependence. Duek identifies that individuals need to recognise that their individual needs are better served in terms of three constructs: cognitive, affect and process. The cognitive component is the quality of the group learning they collaborate to create. The affect construct is the extent to which the emotional wellbeing of the group serves the learning and needs of the members. Bouihuijs et al (Bouhuijs 1984) tracked students who were negatively regarded in one group but found that no one student was negatively regarded in all groups, which suggests that student well-being is a more dynamic and subtle construct than poor behaviour by one individual.

Duek proposes that the process construct, how students behave, is key and identifies equity in discussion as exemplifying effective cooperative principles. Certainly, as discussed there is a belief that marked imbalance in contribution to discussion would be detrimental to the group. Much faculty development is concerned with developing skills for enabling facilitators to create equitable learning environments for their learners, especially when discussing subjects where strong views and emotions may be present, such as in discussions of medical ethics and religious beliefs. However, gender equity still proves elusive and many studies have found consistent differences in the way the genders are treated in classrooms, particularly in discussions. In overview, males take a more active part in classroom interaction, receive more attention and more positive feedback from their teachers. One recent small pilot study shows gender bias in evaluating skill in discussions is very much alive and well, and exists in how students evaluate their own University facilitators. MacNell et al (MacNell, Driscoll et al. 2015) adopted cross gender identities in facilitating online discussions, and found that the same tutor was rated significantly lower whenever thought female by the students, even if in fact male.

Duek observed small group discussion groups studying basic anatomy in a clinical context, where 9 – 12 students met twice a week for two hours to discuss a case; once with a facilitating tutor, who guided their performance and once without. She recorded discussions, allocated utterances to individuals and interviewed learners.

2.3.35 Behaviours changed when tutors present

When the tutor was present, on average females halved their contributions, measured in number of turns, compared to the tutor absent condition, when contributions were broadly equal by gender. Behaviours related to asking questions changed when the tutor was present, with most questions being directed towards the tutor rather than fellow students. The Hawthorne effect, and the wish to gain good grades suggests that student behaviour would change when the tutor is

present, encouraging some to hide potential mistakes or demands and some to increase displays of knowledge (Franke and Kaul 1978). These behaviours have been shown to be strongly associated with gender (Caspi, Chajut et al. 2008).

2.3.36 Student frequently adopted roles

Duek labelled some patterns of behaviour which she observed as roles which students tended to adopt. She identified discussion dominators, who tended to talk over other speakers' contributions, contribute much more often than others, give their assent or disagreement to other student contributions and sought to gain approval from the tutor or group members. In several groups one learner continued to understudy work allocated to other group members and then compete to contribute when it was not their turn. She termed this a hyper-contributor. This went unchallenged by the facilitators but was resented by the other group members.

Hypo-contributors were students who contributed fewer utterances than average. Sometimes this was disguised by adopting another role, such as becoming the scribe. Some students physically distanced themselves from the group, some stayed within the group but examples of withdrawal from discussion were seen in both conditions. Some students continued separate conversations in competition with the group discourse.

In overview, students and tutors believed that it was the responsibility of facilitators to manage group discussions and to keep learners on task. It was noted that positive feedback, which has been shown to greatly enhance student participation in speech was received more often by males. Duek suggests that the patterns she observed are likely to be influenced by a combination of factors including, traditional roles, student expectations, tutor expectations and tutor feedback.

2.3.37 Section Summary

This section of the literature review has discussed a number of factors relating to learning in groups through discussion. The group has been considered as the social context within which all discussion must occur, yet acknowledging that non-verbal, facilitator and interactional data are not the focus of this study. The nature of the PBL work group has been considered in the light of group size, group roles, productivity and 'social loafing', process losses, gender, turn taking, and considered quality and some factors associated with success at medical school. The following section will go on to consider individual factors, relating to personality and related learning styles, and learning through PBL discussions.

2.4 Introduction to Learning styles

This section will broadly survey the current situation with respect to learning styles theory in relation to individual learning performance and behaviour in discussions.

First, this section presents a brief summary of the main approaches and learning styles inventories which have been applied to the target or near target populations before turning to the forceful critique of the field, in the meta-analysis from Coffield (Coffield, Moseley et al. 2004). It will then describe some of the findings of the most salient evidence of distribution of KLSI types among medical students and trainees, links with medical specialty choice, preference for learning approaches and evidence for academic success. The section concludes with a short discussion of the relationship between learning styles and personality which forms the next section of the literature review.

2.4.1 Introduction to the field of learning styles

It is widely recognised that individuals differ in their approaches to learning and respond differently to factors in the learning environment (Paul, Bojanczyk et al. 1994). Learning style is a term given to describe the manner in which students receive and process information (Coffield, Moseley et al. 2004). Unlike a measurable ability, no one style is considered to be of greater value or better than any other, (Curry 1999), but in certain circumstances one style might prove an advantage or an impediment. If a learner had a pronounced learning style, i.e.; strengths in one approach and not others, then that might lead them to struggle in some circumstances. It might be hoped that understanding these factors could help educators to better meet the various needs of all students and improve the quality of learning, or at least enhance the effectiveness of the process (Felder and Brent 2005, Hawk and Shah 2007).

It is worth noting that the demands of Medical Education are somewhat unusual in comparison with some other disciplines in that it is highly demanding, over a long period of study and training in all three of Bloom's domains: cognitive, affective, and psychomotor (Bloom 1989). To succeed in a clinical training, the learner must succeed in receiving and processing information and experience cognitively through intensive teaching and study; experientially, through intensive attachments where levels of supervision and expectations vary; and emotionally, through dealing with prolonged stress and change and high professional expectations from early in their study. Ultimately, a clinician's skills are then practised in complex environments where performance impacts greatly on others, chiefly those who are vulnerable through illness and their relatives.

A learning style is not an ability, a strategy or a tactic. It is an unconsciously adopted approach which influences the 'theoretical framework' within which one exists. As such it can be seen to overlap with some personality approaches, in particular the Myers' Briggs typology, which has been widely used in industry as a framework to guide personally tailored learning and development (Myers, McCaulley et al. 1985).

However, the view that learning styles are more flexible than personality and can shift and adapt as suits the learner, seems increasingly to be held by theorists, particularly as the low test-re-test reliability for learning styles and comparability across measures reveals little coherence (Coffield, Moseley et al. 2004). Curry proposes a layered 'onion model' with a learning style as the most malleable outer skin of the self in a learning context with inner layers of more stable personality within (Curry 1999). This interpretation sits well with the constructivist approaches of Piaget, and Kolb, who sought to describe the process of learning as the aspect **by** which a person can adapt and change whilst maintaining relative personality stability (Kolb 1971).

The following table, adapted from Coffield, summarises groupings of learning style according to their perceived degree of flexibility which neatly indicates some of the range and longevity of activity within the field of learning styles (Coffield, Moseley et al. 2004).

Table 4 Summarising instruments, with date of instrument version, devised to assess learning styles according to the perceived stability of the learning style

Stable Learning Styles	**Flexible Learning Styles**
Apter's Motivational Style Profile (MSP)	Kolb's Learning Style Inventory (LSI) (1984, 2009)
Dunn and Dunn model and Instruments of Learning Styles (1992)	Felder and Silverman's Index of Learning Styles (ILS) (1988)
Gregorc's Mind Styles Model and Style Delineator (GSD) (1982)	Allison and Hayes: Cognitive Styles Index (CSI) (1988)
Myers-Briggs type Indicator (MBTI) (1998)	Honey and Mumford's Learning Styles Questionnaire (LSQ) (1992)
Riding's cognitive Style Analysis (CSA) (1991)	Herrmann's Brain Dominance Instrument (HBD) (1996)
Stemberg's thinking Style Inventory (TSI) (2001)	Jackson's Learning Style Profiler

Coffield and colleagues critically reviewed the most commonly used learning styles of the day and examined 13 out of the 71 different models. They found the research field to have a number of problems. In common with personality research prior to the breakthrough of McCrae and Costa's 'big five', FFM (Costa and McCrae 1985), there was little unity in conceptual approach and language. In addition, a conflict of

interest arises in that many of the researchers were selling their own learning style instruments whilst publishing limited psychometric data on their own tests. In addition, some of the tests used most widely have serious psychometric weaknesses, particularly in terms of their validity.

2.4.2 Kolb's learning styles inventory

Kolb is identified with the growth of interest in the learning styles field, from his earliest work in the nineteen seventies (Kolb 1971) up to the present day when his daughter, Alice Kolb, continues to market the KLSI in its latest form, (Kolb and Kolb 2005). Kolb does claim that matching the teaching style to the learning style of students will lead to improved learning and so improve academic scores (Kolb 1984). However, there is conflicting evidence for this hypothesis, and by the time of Coffield's review, in 2004, there was no significant evidence to support this claim and the initial excitement about what learning styles could offer to enhance the effectiveness of education had begun to die down.

However, Kolb's learning styles inventory (KLSI) continues to be used in exploratory studies of factors which influence how students learn. As it has been adopted as one measure for the current study, this section gives a brief description of the model, before considering the specific criticisms of the theory and the measurement instrument, summarised by Coffield.

2.4.3 Kolb's learning styles

First, Kolb establishes his working definition of learning itself:

> 'Learning is the process whereby knowledge is created through the transformation of experience. Knowledge results from the combination of grasping experience and transforming it.
>
> (Kolb, 1984, p. 41)'

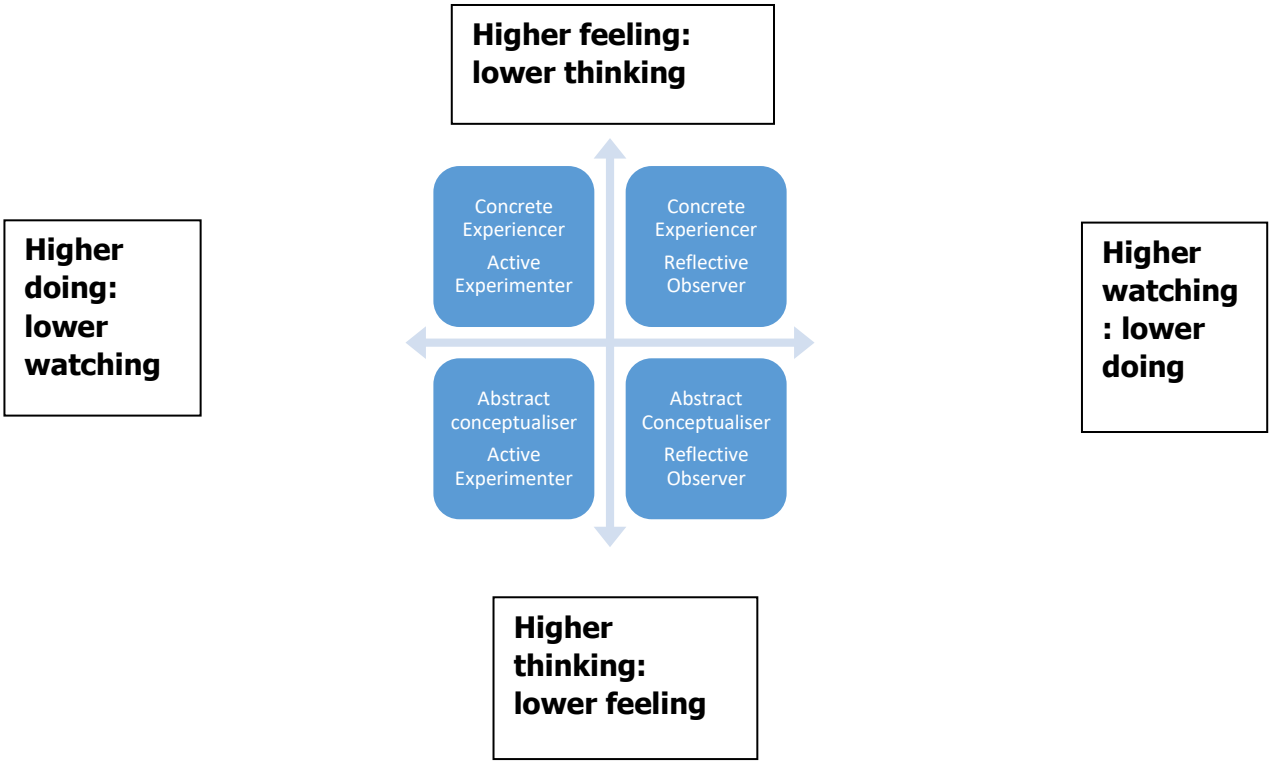
As a brief summary, learning from experience, Kolb proposes, has six main characteristics:

1. Learning is a process, not an outcome (a position in opposition to the behaviourist position, that learning can only be an outcome)
2. Learning happens constantly in response to experience
3. Learning is innately conflictual and occurs when the world outside does not conform to the world view within
4. Learning is the process whereby we change and adapt to the outside world.
5. Learning requires interaction between learner and the environment
6. Learning is a product of personal and social knowledge

Kolb's learning theory consists of four distinct learning styles, which are based on a four-stage learning cycle, describing both the learning process and how individuals relate to it within one model.

Kolb's learning model combines two continua that form a quadrant, within which the learning styles sit. One continuum relates to the extent to which an individual prefers to approach tasks through active engagement, doing; as opposed to through observation: watching. The second, perception continuum, relates to the extent to which an individual responds to a learning experience through thought or through feeling.

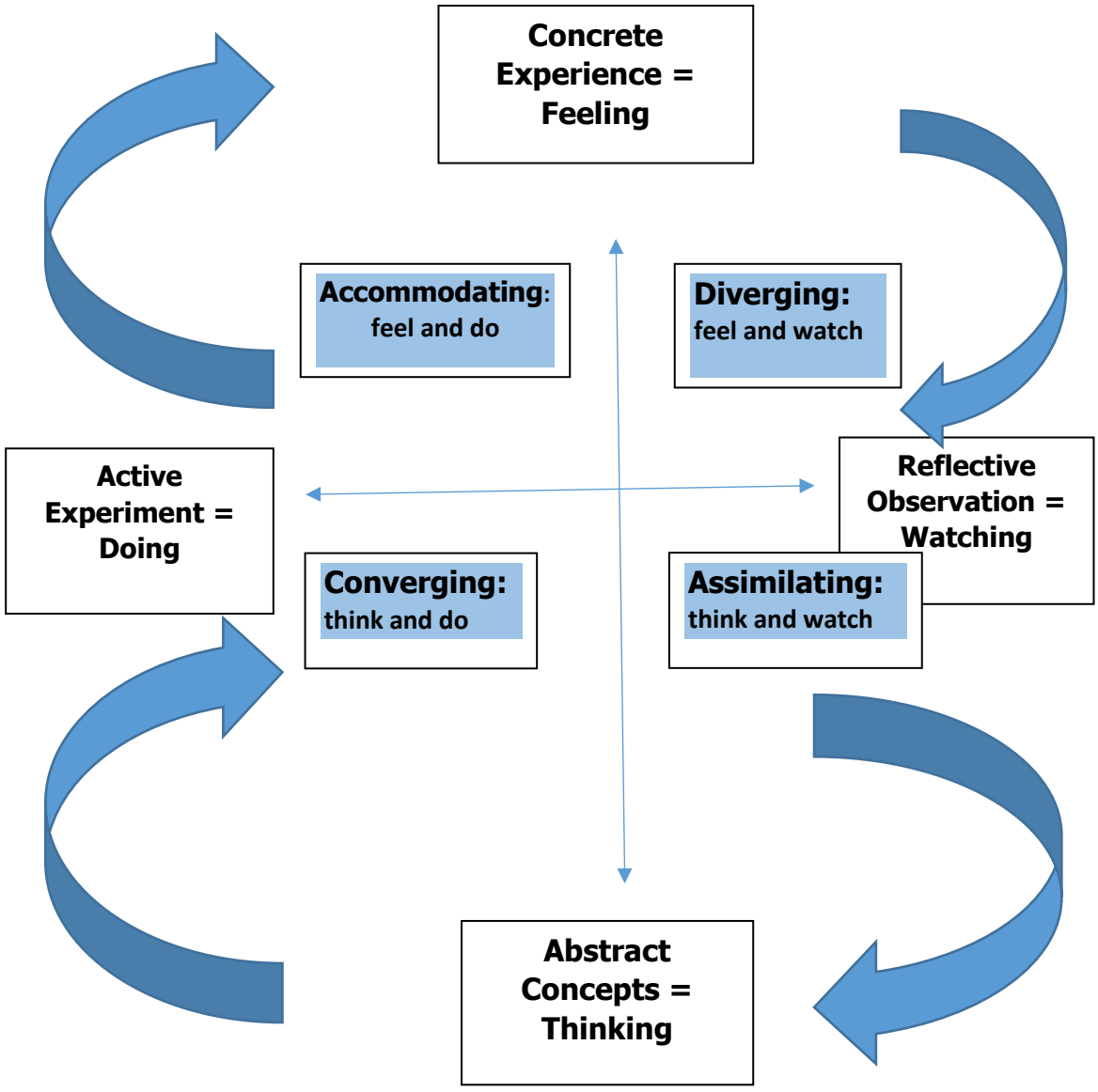
Figure 7 Kolb's Learning Styles Quadrants



The Kolb Learning Cycle

For Kolb, this Experiential Learning Cycle described the constant process of learning; each mode, being a possible entry point, but also a necessary stage in a fully realised instance of learning through experience (Kolb 1984).

Figure 8 Kolb's Learning Cycle



For example:

Learning to ride a bicycle involves:

Reflective Observation (RO) –	watching others ride and thinking about riding a bike
Abstract conceptualization (AO) –	understanding the concept of using a bike for travel
Concrete experience (CE) –	Being taught relevant practical tips and techniques by an expert biker
Active experimentation (AE) -	Riding the bike yourself

For Example, learning a software program involves:

Active experimentation –	Attempting to use the program immediately
Reflective observation -	Thinking about what was easy and what you need to develop
Abstract conceptualization -	Read the manual to get a better picture of what you did well and how to do better
Concrete experience –	Continue to use the help feature while you get more experience

2.4.4 Learning Styles KLSI

Kolb proposed that learners had relatively stable preferences for how they like to approach learning. Different KLSI types are theorised to have particular strengths in one of the four quadrants: divergent, assimilating, convergent or accommodating. The table below gives a brief summary of the strengths of each quadrant.

Table 5 Summarising the KLSI types

Diverging concrete, reflective learner
• **Emphasizes the innovative and imaginative approach** • **Views situations from multiple perspectives** • **adapts by observation rather than by action** • **Interested in people and tends to be feeling** • **Likes cooperative groups and brainstorming**
Assimilating, abstract, reflective
• **Integrates observations and thoughts** • **Reasons inductively, creating models and theories.** • **Likes to design projects and experiments.**
Converging, abstract, active
• **Emphasizes the practical and solving problems** • **Likes decision making, problem solving, and the practical application of ideas.** • **Prefers technical problems over interpersonal issues**
Accommodating (concrete, active)
• **Uses trial and error rather than thought and reflection, likes discovery learning** • **Good at adapting to changing circumstances;** • **tends to be at ease with people**

2.4.5 Nicknames for the Four Kolb Learning Styles

The quadrant labels have obviously proved somewhat unmemorable in use, and have been re-labelled several times, Here are two such examples; one by Zoghi (Zoghi, Brown et al. 2010), the next, by Honey and Mumford, whose LSI achieved widespread use in the UK.

Table 6 Nicknames for the Four Kolb Learning Styles

KLSI Types	Zoghi's re-naming
Converger	Devil's advocate learner
Diverger	Social comfort learner
Accommodator	Hands on learner
Assimilator	Fact-oriented learner

2.4.6 Honey and Mumford's Kolb variation

Honey and Mumford simplified Kolb's model (Honey 1982), returning to Kolb's idea of the intersecting continua; labelling the types as follows:

Table 7 KLSI polarities and Honey and Mumford's renaming

KLSI Polarities	Honey & Mumford's re-naming
Doing	Activist
Feeling/experiencing	Pragmatist
Thinking	Theorist
Watching	Reflector

2.4.7 Common criticisms of Kolb's Learning Cycle and Learning Style models

Kolb's model and the resulting styles are commonly criticised for not demonstrating sufficiently reliable a statistical basis, particular for over-claiming the association

between KLSI and certain professional groups and learners (Garner 2000). The theoretical basis runs close to Jung's archetypes theory (Jung 1921), which although offering useful applications, such as the Myers Briggs Type Inventory, (MBTI), discussed in the personality section of this review, is not susceptible to rigorous, scientific methodologies. Others criticise the model for not fully accounting for the role of reflection in the learning process nor for the role of interaction with the social environment (Boud, Keogh et al. 1985). However, even the sternest critics, such as Garner and Coffield, acknowledge that these criticisms do not mean that the model is wrong nor lacking useful application with learners, especially as the basis for individual self-development.

With such reservations in mind, this section will go on to review pertinent findings for similar study populations, using the KLSI, or near variants.

2.4.8 Studies of distribution and stability of Kolb learning styles among healthcare and medical students

Several studies publish the distribution of KLSI types among healthcare students and trainees, often presenting evidence for and against stability. Zoghi (Zoghi, Brown et al. 2010), in their meta-analytic study of 752 allied health professionals in Australia, found inconsistent results when compared to the variety of different small scale studies over the years. Kolb had predicted that health care professionals may predominate in the person oriented, practical dimensions, represented by accommodating and divergent learners (Kolb 1981). But in this study accommodating and divergent learners were **not** shown to be over-represented in health care students. In fact these were the lowest represented styles, with convergent types predominating in this stage of student learning.

Mammen et al (Mammen, Fischer et al. 2007) demonstrated over twelve years of collecting KLSI data for general surgery residents in Cincinnati, that convergers were in the majority (57%) and remained stable across the years of training.

Gurpinar et al studied the distribution of KLSI learning styles amongst first and second year medical students in three different Turkish medical faculties, with PBL, hybrid and integrated curriculum models (Gurpinar, Bati et al. 2011). The learning style was observed to change markedly from first to second year across students within the three faculties. The largest change in learning style from first to second year was observed in students participating in the PBL curriculum where 56.4% changed style. Most change was observed in divergers where 93% changed their learning style within the PBL and hybrid curricula. The great majority of divergers shifted to assimilating (58.6%) and converging (27.6%); and relatively few assimilators and convergers changed their learning style (5 – 6%). The authors acknowledged that the number of PBL students included in the study, (55) was a serious limitation, although comparable to the 45 first year students recruited to this study.

2.4.9 Studies of learning style and medical specialty choice

Several studies attempt to determine correlations between KLSI and choice of medical specialty, but again the results are mixed. Plovnick (Plovnick 1975) found that medical specialty career choices were strongly affected for some learning style types. Work experience in a healthcare setting was indicated as the most important factor in career choice by 100% of the accommodators and divergers in his sample. These are the types who respond most strongly, according to Kolb's model, to their concrete experiences and to people. They also selected primary care as their overall, preferred career specialty.

Engels (Engels and de Gara 2010) found that second year undergraduate medical students, general surgery residents and surgery faculty differed in their LSI results. For the undergraduates the predominating style was assimilating with a lower proportion of convergers than for the resident physician sample. The authors proposed that their data may describe a situation where assimilators, although dominant in the medical undergraduate population, may then choose to pursue

other specialties, rather than remain to study as general surgery residents and change their learning style. This was not possible to discern from their data.

2.4.10 Learning style and achievement, particularly in PBL

Clear evidence linking learning styles and academic achievement has not been found, (Coffield, Moseley et al. 2004). However, medical students do report different preferences for the same learning methods.

One study set out to identify the learning styles associated with first year medical students (Gurpinar, Alimoglu et al. 2010) and explored student satisfaction via questionnaires and results via exam scores. It claimed to be the first study linking these measures for this population. The medical students were identified by Kolb learning styles as 47.7 % diverging, 41.5% assimilating. Converging accounted for only 6.3%; Accommodating for 4.5%. PBL was found to be associated with significantly higher satisfaction scores than for traditional learning for all learning styles. Using logistical regression analysis, students with an assimilating learning style, predicted higher student satisfaction with traditional training approaches and higher scores in theoretical exams.

It might be expected that students with the assimilative learning style might have a preference for a more traditional pattern of learning and examining, as this is the type most orientated towards 'fact-based learning'. They may be able to continue to use the same study approaches which have already proved successful for them. The same students may face more challenge and be required to develop their learning styles when they are required to learn in more hands-on environments, or when the learning concerns the affective or psychomotor domain, rather than the purely factual.

The spread of students amongst the possible Kolb learning styles is clearly influenced by factors other than pure random distribution. Medical students tend to be selected from a pool of students who have studied substantially the same science

subjects, which have been examined in similar ways. This might imply that learning styles will have been strengthened which are associated with such study. It is important to note that the Derby GEMS students gain entry via the GAMSAT score in order to demonstrate sufficient scientific knowledge and reasoning abilities, without having to complete the same scientific Advanced Level subjects. This may suggest this cohort are somewhat freer from this selection bias than undergraduate medical students.

Lynch et al (Lynch, Woelfl et al. 1998) found a strong correlation between convergers and assimilators and higher grades in MCQ knowledge tests but showed no correlation at all with clinical performance measured by OSCE style tests. Their advice was to vary the assessment methodologies used to evaluate medical students in order to extend the validity of assessment methods, given the variation in individual students.

Alghasham (Alghasham 2012) explored the relationship between students' learning styles and their classroom performance in PBL. Learners were identified according to Felder's learning style inventory, so only broad statements relating to active and reflective styles may be inferred in assessing results measured by different inventories. The PBL model was a two stage process; initially generating learning objectives for self-study relating to a problem before the second meeting where they applied the understanding they had gathered. Groups were formed comprising either active (four PBL groups, 40% of total) or reflective student groups (five PBL groups, 60% of total). Performance during the PBL session was assessed using a questionnaire devised by the author relating to independent study, group interaction, reasoning/problem-solving skills and active participation, each aspect being tutor rated from 0 – 6.

Broadly, the post PBL assessment scores identified no significant differences. However, there were differences in how the students behaved during the learning task with active learners accessing a more diverse range of sources of help and

information and working cooperatively. Reflective learners focussed on fewer sources of information but used them more effectively. In the active groups, learners demonstrated more behaviours which were categorised as problem solving or reasoning. The author concludes by suggesting that PBL groups may benefit in future from balancing learners with active and reflective approaches within the same PBL groups. In addition, if learners are aware of their strengths and weaknesses and are supported in how to develop their range of approaches, that this may enhance their learning experience. However, one recent attempt at Leicester Medical School attempted to balance types within teams using the Belbin type indicator (Belbin 2000), but was not able to show any significant advantage in the PBL setting for those medical students. (Batchelder 2012)

2.4.11 Learning style and the relationship to curriculum planning

Armstrong and Parsa-Parsi, (Armstrong and Parsa-Parsi 2005) apply the ideas of Kolb's learning cycle and styles to the development of curricula in Harvard Medical School. Their proposal is to suggest that the curriculum is mapped against the quadrants of Kolb's learning cycle, to ensure balance and coherence. They term the quadrants: activate prior knowledge (Divergent), ask what is the new knowledge? (What? Assimilating), before trying it out in practice (Application, Convergent) and concluding with just doing it (If?, Accommodating).

This has close parallels with my own curriculum planning and delivery approach. I use Bloom's taxonomy, and the theories of constructivism, to plan and provide for a predictably variable range of individuals (Dennick 2012). I set the learning context by planning to activate prior learning, demonstrate relevance to the practical and theoretical domains, and give an indication of the big picture and the smaller details. I plan for each concept to be applied as soon as possible in some sort of activity, which ideally involves some reflection and discussion, before summarising and concluding.

2.4.12 Learning styles and maturity and gender

GEMS PBL learners are by definition more mature than undergraduate medical students, and in common with modern UG medical cohorts, over 50% of medical students are female. Learning styles have been shown to vary by some specific measures of dimensions which neuroscience has also identified as differing in the genders (Ruigrok, Salimi-Khorshidi et al. 2014). However, learning style inventories such as KLSI, which focus on how students handle ideas, and how they perceive their approach, tend not to identify gender differences.

One Swedish study (Heffler 2001) found no difference related to age, but did find gender-related differences using Kolb's LSI which related to handling ideas. Females demonstrated a stronger preference for the concrete dimension as opposed to the abstract dimension, indicating a greater willingness to learn from experience and others and a lower reliance on fact and authority in comparison to the male students. This also accords with gender differences in consultation and communication styles to be discussed in subsequent sections.

In early phenomenographic studies, basic demographic information was often simply left out. Few studies consider the association between age and gender and approach to learning and studying, although common sense suggests that individuals mature on the whole as they age and that an adequate level of maturity is an important predictive element among the many confounding variables at play here. To return or take to study later than one's cohort may be associated with multiple socio-economic and personal factors, each of which could have a beneficial or negative influence on achievement. Neuroscience is discovering mental processes which are still completing during a person's twenties. Formerly considered fully adult, we are now discerning that inhibition control and ethical processing may need up to 30 years to complete neurological development (Sowell, Thompson et al. 1999, Ruigrok, Salimi-Khorshidi et al. 2014).

2.4.13 The individual: Personality and Learning styles

The relevant literature pertaining to personality and learning styles straddles education, psychology, particularly in a business context with few studies having been carried out in a relevant medical school environment. Evidence from a range of different, credible personality type and learning style measures are included in the literature review. However, the researcher has adopted the position proposed by (Jackson and Lawty-Jones 1996) that learning styles can be usefully considered as a sub-set of a learner's personality type and that evidence which is able to triangulate these measures is particularly sought. There is evidence for some triangulation between personality construct measures and those targeted by the learning styles construct measures.

Previous research showed overlap between extraversion and learning styles such as activist, as opposed to introversion and the reflector style. Furnham et al. (Furnham, Jackson et al. 1999) demonstrated links in a study of telesales staff, using the Honey and Mumford questionnaire, which is closely linked to the Kolb learning styles questionnaire. He predicted that activist/pragmatist learning styles would be linked with higher appraisals by managers, and reflector/theorists would be negatively correlated. Extraverts with low neuroticism were similarly expected to be scored highly.

The results suggested that variance explained by personality and learning style together was only 8%, explaining low predictive validity within the workplace. Contrary to predictions, the activist type was not significantly related to positive appraisal, but the reflector style was consistently negatively related. For a role in telesales it is perhaps not surprising that extraversion with low emotionality is associated with higher performance evaluations.

Context is clearly vital. What is valued in the context of telesales, may be irrelevant or problematic in a field such as medicine. For example: the telesales environment did not value reflectors highly. Reflectors are often identified as being careful,

thoughtful, thorough, methodical, unassertive, risk averse and indecisive and with a tendency to hold back. The same tendencies could prove to be highly valued in performance in a medical context. Furnham identifies 'learned components of personality' as a way of describing the aspects of personality which are subject to change and may be described by the learning styles inventories.

This literature review will now continue with a consideration of personality research.

2.5 Introduction to the literature relating Personality to the PBL setting

This study seeks to examine the correlation between self-reporting FFM personality measures with a record of observed speech behaviours during PBL discussions, and indications of attainment in early GEMS. This section of the literature review therefore needs to establish what is meant by the FFM and will briefly summarise the history of personality research, and give an overview of the current situation in the field. The review will then focus on studies which have linked each measure of personality, particularly recent FFM studies, to factors which relate to learning through discussion, such as preference for groups, degrees of talkativeness, tendency to experience strong emotions, particularly anxiety, and performance and progression in careers, with particular reference to medicine when available.

2.5.1 Brief background to FFM and the history of personality Research

2.5.1.1 Pre-twentieth century personality typologies and theories

Human nature has of course always been studied as the ability to predict future behaviour confers an advantage. From the time of Hippocrates, 'humoralism' has been used to account for the physical state of the body and its disease, with Galen establishing the typology of the four humours (Benson 2012) . 'Every man in his humour', by Ben Jonson, 1598, is the best-known example of the sub-genre of humours comedy, where characters representing the four types explored behaviour

through drama. Clearly people have used typologies over the centuries to provide a relatively simple framework within which human difference could be observed to fall.

2.5.1.2 Myers Briggs Typology: a theory and a measure

The Myers Briggs Type Indicator (MBTI) has been widely used in business settings since it was first derived in the 1940s from Jungian thinking on archetypes (Jung 1921). The MBTI has four dimensions, a four factors model, as it lacks a dimension identifying neuroticism, or emotionality. Jung theorised that 'sensation is to establish that something exists, thinking tells us what it means, feeling what its value is, and intuition surmises whence it comes and whither it goes'. Myers Briggs elaborated their theory of human types to these dimensions as, introvert or extravert, (I-E), intuitive or sensing (N-S), and feeling or thinking (F-T). They added one new dimension to account for how an individual related to information, with opposing approaches of perceiving or judging (P-J). Thus, sixteen personality types were theorised. Various scales were then devised and evolved which identified how samples best fit the typologies (Myers, McCaulley et al. 1985)

Whilst according to a meta-analysis in 2002, it has been reported as having acceptable reliability and validity by many study authors, this is not considered to be the case by more modern statistical standards (Capraro and Capraro 2002). The FFM has come to be regarded as the more statistically reliable way of measuring differences between individuals. The FFM does not seek to provide a theory to account for how or why that difference exists. However, there is some comparability between the MBTI dimensions and the FFM dimensions, which will be briefly summarised once the FFM dimensions have been introduced below.

2.5.1.3 Lexical and statistically derived personality measures

In the twentieth century psychologists sought to derive a new understanding of the categories of human behaviour, combining a lexical approach, with statistics. Allport

first used a questionnaire in combination with input from people who knew the subject well to establish the feasibility of measuring personality. He pursued studies of nomothetic aspects, those which applied to all people, as well as an idiographic method which blended many approaches to produce a detailed understanding of what might be unique about one individual.

Allport and Odbert then worked on the assumption that the most relevant and important aspects of personality difference would be reflected in human language, using dictionaries to identify 4500 adjectives which they considered to be observable and stable personality traits (Allport and Odbert 1936). This corpus study seeks to link the words which people use in PBL discussions with their personality type: a reversal of this approach. Indeed, at the end of his career, Allport analysed the letters of Jenny Masterson, triangulating his findings with the views of those who knew her, and applied early content analysis by computer to this project (Masterson 1965). One recent critique of the lexical approach to personality is that it favours social aspects of personality, as language itself developed in response to social needs and lacks specificity for internal mental states which must be inferred (Trofimova 2014).

Cattell further reduced categories to 16 personality factors (Cattell 1943, Cattell 1957). Hans Eysenck, used three factors, Extraversion, Neuroticism and Psychoticism to describe personality (Eysenck 1975). Norman eradicated evaluative words, and grouped them into broad dimensions, which could be arranged along a continuum, or seen as polarities, which could then be analysed using statistical techniques, especially factor analysis (Norman 1967). Factor analysis allows us to identify the factor or abstract concept, which those measured have in common. Goldberg resolved the traits into five factors: openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism. (Goldberg 1990).

2.5.2 The Five Factor Method personality measurement

The origins of research into personality through the norms of psychological science, have converged around the five factor method, the FFM. Costa and McCrae set out this unifying framework describing the big five dimensions, the Five Factor Method, within which much human difference in disposition could be described with relative reliability and stability (Costa , Costa and McCrae 1985, McCrae and Costa Jr 1999).

Within this model, human personality is conceived of as being accounted for by five broad traits, each of which has two opposing aspects. An individual will have relatively stable traits which lead them mostly to act more often in one aspect and less often towards its polar opposite. Individuals vary in the strength of their measured tendency. For example, a moderately extravert person will be able to study quietly in the library at times, but will tend to be more outgoing and sociable in a social setting. However overall, they are likely to be more sociable, and seek out more social contacts and stimulation than an introverted person would in the same settings. Some people also tend to be more 'volatile', to experience more heightened emotional responses than do people with high emotional stability (Moskowitz 2004).

2.5.2.1 Summary of original traits and facets in the FFM

Personality tests based on the 'Big Five' include commercially trademarked tests such as Orpheus and Quintax and the Neuroticism-Extraversion-Openness personality inventory (NEO-PI), which was revised to include the conscientiousness and agreeableness domains (NEO-PIR). Each trait was sub-divided into a further six facets, which are summarised in the following table.

Table 8 A summary of the FFM traits and sub facets: NEO PI-Revised

Neuroticism: Anxiety, Hostility, Depression, Self-consciousness, Impulsiveness, Vulnerability **Extraversion:** Warmth, Gregariousness, Assertiveness, Activity, Excitement-Seeking, Positive Emotions **Openness to Experience:** Fantasy, Aesthetics, Feelings, Actions, Ideas, Values

To which were added, in 1985, (Costa and McCrae 1985).

Agreeableness: Trust, Straightforwardness, Altruism, Compliance, Modesty, Tender-mindedness **Conscientiousness:** Competence, Order, Dutifulness, Achievement Striving, Self-Discipline, Deliberation

The following table seeks to describe, in lay terms whilst avoiding value judgements, what behaviour is likely to be observed towards the extreme of each continuum, at the polar opposites. For consistency, this order will be now used for considering the FFM traits.

Table 9 A brief summary of McCrae and Costa's FFM continua

McCrae and Costa's continua	**Brief description of Traits**	**Brief description of Traits**
Extraversion v Introversion	Extravert type is: • sociable • fun-loving • affectionate	Introverted type is: • retiring • sober • reserved
Agreeableness v Antagonistic	Agreeable type is: • good natured • cooperative • lenient	Antagonistic type is: • irritable • competitive • critical
Conscientious v Undirected	Conscientious type is: • consistent • orderly • industrious	Undirected type is: • spontaneous • flexible • procrastinates
Openness to experience v in-curious	Open type is: • original • creative • independent	'Closed' type is: • conventional • uncreative • conforming
Neuroticism v Stability	Neurotic type is: • emotionally reactive • nervous • worries	Stable type is: • relaxed • at ease • calm

2.5.3 Brief comparison of the MBTI and the FFM dimensions

Furnham compared the two measurement scales of the MBTI and the FFM in 160 adults in order to replicate other studies and confirm the extent to which they may be seeking to measure the same personality constructs (Furnham 1996). One factor was of course absent; neuroticism is not gathered for MBTI. As expected,

extraversion correlated closely with the identically named scale. Of the remaining three dimensions, conscientiousness was found to correlate most closely with the perceiving-judging dimension, and agreeableness with the thinking-feeling dimension. Openness to experience was correlated across all four MBTI dimensions, but most strongly with the sensing-intuitive dimension. Furnham suggests that it is credible to consider findings established for different populations from either measure but that future studies would benefit from the statistical power, reliability and validity of the FFM, which also includes a measure of neuroticism, long established as a fundamental human trait.

2.5.4 Temperament types: a potential overview of personality derived from the FFM

Each of the five factor dimensions can be analysed at the level of facets, with six facets for each dimension. This restricts full detailed analysis to much larger populations than may be readily available within one cohort, such as within one GEMS, as was available to this study. Many of the studies report their analyses at the FFM dimension level only. Some, particularly those based on MBTI measures and related to business and the workplace, group key dimensions together to formulate four temperament types, which psychologists, such as Keirsey argue encapsulates differentiation between personalities with sufficient robustness to permit meaningful discussion.(Keirsey and Bates 1984).

Keirsey proposes that temperaments are organised in a very similar way to that of Kolb and his learning styles, (considered in Section 2.4). His temperament types are considered on the basis of just two dimensions. Firstly, whether an individual relates more strongly to abstract theories (theoretical), or to concrete experience, (grounded). Secondly, temperament types consider whether individuals are oriented chiefly towards working with other people, (personable) or by focussing on the effectiveness of results and reality (logical).

2.5.5 Quintax Temperaments

These different temperaments, as conceived by Keirsey, can be seen to relate to only three of the FFM dimensions: conscientiousness, openness to experience and agreeableness. Overall, these dimensions are associated with less cohesive evidence for accounting for human behaviour. In contrast, extraversion and neuroticism have consistently been shown to relate closely to ranges of behaviour and some consider the evidence to be more likely to be biologically determined (Caspi 2005).

Apparently slightly simpler than the FFM, this does suggest that temperament types may be less consistent and powerful in predicting human behaviour. Quintax combines both FFM and temperament types and will be reviewed in section 3.2.

Table 10 Summary table for four dominant Quintax temperaments

Temperament Name	Tagline	Description
Adaptable with Grounded (AG)	Trouble-shooters	Action-oriented, explore the here and now
Structured with Grounded (SG)	Organisers and Doers	Cautious, deadline – conscious, traditionalists.
Logical with Theoretical (LT)	Logical Strategists	Conceptual, future oriented, change strategists.
Personable with Theoretical (PT)	Passionate Idealists	Motivators, investors in people, spokesperson

2.5.6 Reports on the statistical performance of the FFM

Within the first ten years of applying the FFM, the NEO-PI had established high internal consistency of between 0.8 and 0.9 for each of the five domains, and good test-re-test reliability (0.9) (McCrae and Costa 2010). Other measures, for example Quintax, replicates this using a 'lie-scale' to highlight where an individual may seek to present themselves more favourably than would be consistent with the facts, in order to compensate for problems in self-reporting bias.

Mischel (Mischel 1968) offers a strong critique to the FFM, taking issue with the degree of stability of traits, (the so called state versus trait 'controversy') and particularly with the extent to which behaviour can be shown to vary depending on the situation (trait:situation controversy). Mischel is most well-known for the 'marshmallow' experiment which demonstrated a correlation between a child's persistence in resisting a marshmallow sweet in return for a delayed reward and their success and achievements in later life. His interpretation is not that self-control is innate, but that it can be learned. Debate is still active as part of the 'nature' nurture debate' which seeks to account for how humans are alike and how to account for the undeniable individual differences (Ridley 2003).

2.5.7 McAdams and Pals: big five unifying principles for personality

McAdams and Pals (McAdams and Pals 2006) set out a further big five unifying principles within which personality, and personality research operate. These take account of evolution and genetic expression and the implications of cultural and life stage difference. The principles relate to how we are 'disposed' to behave in regular patterns, as well as how we may characteristically adapt across situations.

- Evolution and human nature
- Life narratives and the challenge of modern identity
- The differential role of culture
- Characteristic adaptations
- The dispositional signature

Genetic variance, or evolution and human nature have been shown through various twin studies, to account for up to half of the variance in each of the Big Five traits (Loehlin, McCrae et al. 1998). The rest appears to be due to developmental factors. In other words, the way in which we interact with and respond to our environment leads to the expression of up to half our personality.

Cultural interpretations of behaviour and the norms for conduct have been shown to vary, particularly at the sub-trait level. However, at the broad level the FFM have held up to cultural scrutiny over the past thirty years (McCrae and Terracciano Mar 2005).

The focus in this PhD study is on exploring personality as measured under what McAdams and Pals term, the dispositional signature, as PBL learners adapt to learning at medical school using the PBL discussion format. McAdams and Pals sought to differentiate between these more 'fixed' determinants and those which may influence our behaviours at certain times. For example, when motivated to complete a course of study, a person may be capable of sustaining a higher level of 'conscientious' behaviours, in order to achieve that goal. To some extent, our work or professional self, is a characteristic adaptation of the personality or disposition to suit the work environment, and individuals are required to sustain different behaviours from those which they may adopt outside the work situation. However, one of the challenges of studying learning and work behaviour and personality is the sheer diversity of influences which may be strongly influencing an individual in their learning setting. According to McAdam's model, many of these crucial influences which impact on performance and behaviour, and which many have sought to measure in evaluating educational effectiveness, will inevitably fall outside the scope of the FFM and so limit the extent to which FFM can account for variance.

2.5.8 Personality stability

There is evidence that personality can change, but that this tends to occur at particular life stages across all individuals, particularly in young adulthood. Termed normative change, 'the norm' for the whole group has been shown to change and so rank order as a measure of personality traits remains the same, whilst trait behaviours, particularly at the sub-type level, may change. In overview, people tend to become slightly more agreeable, emotionally stable and conscientious as they grow older (Roberts, Walton et al. 2006) but even these changes plateau after the age of 30 and change very little after the age of 50. At the level of extravert sub-traits, as we age, social dominance may increase, but social vitality decreases.

In the UK most students enter higher education, in this case medical training at 18, about a decade before they reach what is currently considered to be full neurological maturity, at around 28. GEMS may offer an added advantage to students who have established their effective study habits through a first degree and who tend to be working in groups which have members who may bring the average age cohort closer to thirty years old. However, this suggests that the personality type will also be formed and is likely to persist as a factor in a future medical career. In overview, Caspi et al conducted a meta-analysis which concludes that the stability of FFM measures is comparable to that of IQ measurements (Caspi 2005).

This section has briefly reviewed the emergence of personality theory, the background to the FFM and current debates in the field.

It is clear that learning or working closely in small groups, might be associated with each of the five factors in complex combination and would be valuable to consider in the context of medical training. The following section will briefly consider the PBL learning environment in general, studies in non-human animals and observations about their personality and behaviour. The section will then summarise the nature of each of the FFM traits and introduce some key studies relating primarily to that dimension.

2.5.9 The Five Factors in relation to learning and discussion. An **introduction to personality, learning and animal behaviour**

Skinner and the behaviourists used rats to model learning in its observable forms as they explored learning as conditioning. In science, all rats were considered equal. Recent studies have studied learning and how individual differences, or personality, may impact on the ability of non-human animals to learn. In one study of learning and the temperament of calves (Webb, van Reenen et al. 2015) it was shown that very similar traits to those of the FFM were correlated with different learning behaviours and attainment. One dimension of calf behaviour, described in this study as proactive or reactive, appeared close to the dimensions which have proved to be the most coherently observed across all studies: extraversion, conscientiousness and neuroticism. Proactive (extravert) calves were swift to explore their environment superficially and had higher levels of activity, but overall were less persistent in their training efforts (possibly lower in conscientiousness). Calves which were reactive (introvert), explored more gradually but intensively and were able to demonstrate more persistence in their training (possibly higher in conscientiousness). Other studies have shown that more fearful animals (high neurotic score) demonstrated lower performance in learned behaviours (Hartmann, Christensen et al. 2011).

The following section will consider the definition of each of the five factor personality dimensions before considering key studies which relate to learning, especially in work and clinical performance.

2.5.10 Extraversion dimension and PBL behaviours

Extraversion is sometimes described as a cluster of traits related to 'surgency'. Surgency is the extent to which an individual displays assertiveness, gregariousness, excitement-seeking and being active, particularly in situations with people. Such behaviour is sometimes identified with seeking dominance in groups. It is also associated with experiencing positive emotions, and displaying emotional states, particularly warmth. The lexical approach inevitably uses terms which may be seen

as conferring value judgements: if an extravert is 'warm' does that mean an introvert is 'cold'? If a highly conscientious person is industrious, is the opposite 'lazy'? However, physiologically, clear difference between individuals on this dimension is consistently demonstrated.

Eysenck was able to measure physiological differences in responses to physical stimulation, such as levels of light, and puffs of air to the eyeball which demonstrate a physiological difference in response to external stimuli between extraverts and introverts (Eysenck 1977). Eysenck graphically demonstrated relatively little overlap where both extraverts and introverts could both experience comfortable levels of stimulation: extraverts only becoming comfortable towards the upper limits of what introverts could tolerate. Results have also been found in relation to performance and noise, with extraverts choosing to set headphones on average at 72 decibels, whilst introverts chose on average 55 decibels. Variations from comfortable levels were found to impair performance (Geen and O'Neal 1969). It might be inferred, that physiologically, other people and their behaviour presents a more complex and stimulating source of energy than simple physical sources of stimulation. This suggests that introverts will be likely to experience greater likelihood of discomfort, or need for isolated breaks in order to function optimally when working closely with others. In a literal sense, it has been shown that extraverts even maintain smaller physical distances from others (Williams 1971).

'Dominance' or the more neutrally termed, surgency, in the PBL setting is likely to encourage extraverts to volunteer to speak more frequently, and to speak over others, and to perhaps move around more and listen with less attention when others are speaking, or to require more energy to listen than to speak. Conversely, introverts are likely to voluntarily speak less frequently, prefer pre prepared topics but perhaps to speak for longer, to give a fuller account of their thoughts. The extraversion-introversion dimension contrasts an outgoing character with a withdrawn nature. Extraverts tend to be more physically and verbally active whereas introverts are independent, reserved, steady and like being alone.

Introverts may be described as quiet, reserved, shy and less sociable (Costa and McCrae 1992). The word introvert, however, is not a synonym for hermit, misanthrope, or a person lacking in social skills. Most introverts are perfectly friendly and are comfortable in social situations and particularly enjoy discussing subjects which interest them, to depth, with people they know (Thorne 1987). By many measures, introverts comprise at least one third of the population of the US, where traits such as friendliness and talkativeness which have been associated with extraversion are strongly culturally encouraged (Cain 2013). Introverts tend to have a small group of close friends and family in their social network, and tend to relax by withdrawing from external stimulation. Extraverts, conversely tend to create large social groupings and relax by seeking external social interaction.

In PBL groups, a small group of people with whom one becomes relatively familiar, but who provide lively discussion for 4 – 6 hours per week, might characterise an optimal compromise between these styles. Introverts may find the initial stages of group formation more demanding, whilst extraverts may particularly enjoy the stimulating nature of discussions among 8 people. Certainly, in one study of individuals in a clinical setting, albeit as patients, extraverts were shown to benefit more and contribute more than introverts in group therapy situations (Funder and Sneed 1993, Hamburger 2000). One study of conversational style found that introvert-introvert dyads held closely focussed discussions, whilst extravert-extravert pairs ranged over many more subjects. Mixed type dyads gave rise to results between these two, in that the discussions were moderated (Thorne 1987). In addition, as discussed in the section on online discussions, where a discussion allows for more thought time and does not require an introvert contributor to assert surgency to participate in the conversation in the same way, introvert contributions might be expected to increase, certainly dis-inhibition has been shown in online discussions in comparison with face to face behaviours (Suler 2004). This may be happening in the introvert – introvert dyad, as well as being inherent in online discussions. Mixed small group discussion such as occur in PBL, may be

characterised by extraverts and introverts having a moderating effect on each other's style.

Clearly, we might expect significant differences between learners who are measured as having strong scores at either end of the dimension, where a large data set was available. The study may be able to detect significant differences as permitted by the available data set of 45 participants, particularly relating to the number of turns taken to speak, the length of utterance and the breadth of topics discussed.

2.5.11 Agreeableness dimension and PBL behaviours

High scores on the Agreeableness dimension refers to the willingness of an individual to maintain positive and reciprocal relationships with others, tending to avoid open criticism and irritability. Having a higher agreeableness score may lead learners to choose not to express their criticisms and disagreements as frequently in open discussions, such as the PBL setting. In general, team behaviours have been shown to tend to the average of the individual's score (Barrick 1998) and so are not impacted powerfully by the behaviour of one individual. However, in the case of agreeableness, the lowest score on agreeableness within a team is important. One highly disagreeable member can substantially and negatively impact a team's performance. As discussed later in this section, medical students have been shown to have higher agreeableness scores relative to other undergraduate students and of those studying medicine agreeableness is associated with a preference for learning through PBL (Lievens, Coessier et al. 2002, Holen, Manandhar et al. 2015).

2.5.12 Conscientiousness dimension and PBL behaviours

Conscientiousness refers to organisational and motivational aspects of a person's behaviour. It has been identified as a moderate to strong predictor of academic success across a range of occupations, particularly sales and managerial occupations, although less so at high levels (Barrick and Mount 1991). However, some statistical issues have been identified that suggest that meta-studies may be obscuring important advantages at the level of the facets of conscientiousness,

which broadly relate to levels of order and planning and industriousness, or persistence. It has been shown that higher conscientiousness is associated with fewer tasks being completed (Driscoll 1994). In certain situations, thoroughness may be undesirable or even, in the case of certain police roles, dangerous (Barrick and Mount 1991). Indeed, high conscientiousness has been linked to poorer performance in certain types of post. High conscientiousness relating to orderliness in following rules has been identified with lower creative problem solving skills and creativity in situations where full information is not available which may be crucial aspects of effective job performance (Hogan and Hogan 1989). Recent research has identified extreme conscientiousness, where excessive ordering and planning and persistence combine to form pathological behaviours, often diagnosed as obsessive-compulsive-disorder (Samuel DB1 2011)

As early as 1984, Lipton et al were proposing that personality factors were considered along with academic grades as reliable predictors of academic performance at medical school and within the medical profession (Lipton, Huxham et al. 1984). In his study of secondary school leavers in Queensland, Lipton tracked 151 medical students, 92% of whom completed medical training. The study correlated attainment at entry in secondary school subjects with academic performance in 28 academic tests over six years of medical school against over ten separate personality questionnaires – an indication of the lack of coherence in the field of personality research prior to the emergence of the FFM in the mid-1980s. Even though secondary school grades were found to be the strongest predictor of academic success in medical school, they found high levels of unexplained variance at the subject level. The personality factors which were shown to have a positive correlation with performance were termed, 'maturity, good study habits', which might be identifying the cluster of behaviours we might now term conscientiousness. English language and maths, both of which might be acting as a proxy measure for generalised IQ were positively correlated with performance. The study did not find

significant correlations between anxiety, part of the neuroticism construct, or the extraversion construct.

A more recent study, (Lievens, Coessier et al. 2002) compared personality types across different degree subjects in Ghent. Medical students were found to have the highest proportion of extraversion and agreeableness when compared to other students. Students who were high in gregariousness and excitement seeking (facets of extraversion) and low in conscientiousness were found to be 'less likely to sit examinations successfully' as the authors tactfully put it.

Fergusson et al (Ferguson, James et al. 2003) also identified predictive power for higher scores in early medical school in association with higher conscientiousness scores; but not for the clinical years, where it was correlated with poorer performance. Doherty and Nugent (Doherty and Nugent 2011) draw the same conclusion in their meta-analysis.

Behaviour associated with PBL discussions which might be impacted by levels of conscientiousness may be conflict generated by students within the same PBL group having different approaches to deadlines, time-management, orderliness of information shared with others through the scribe role and the thoroughness of their preparation for each stage of the PBL process. However, it may be helpful for students with lower conscientiousness levels to have the added incentive against procrastination or lack of planning provided by having a close group of colleagues to account to for their behaviour. PBL also provides an environment where spontaneity is a valued part of the learning process, and where it is necessary to adopt the role of doctor or patient in a role play without having been fully prepared to fulfil that role in reality. Some highly conscientious students, as noted elsewhere, find the unplanned nature of PBL discussions uncomfortably chaotic, and have been associated with disliking PBL for this reason (Holen, Manandhar et al. 2015).

2.5.13 Durham conscientiousness index

Since 2007, data which would have been collected as part of the administrative regime at Durham Medical School relating to student behaviour, such as attendance, timing of assessment submission, reminders to submit data such as immunity status etc. has been successfully correlated with staff and student perceptions of conscientious behaviour (Finn, Carter et al. 2012). A conscientious index takes 300 such individual measures which can be either added to or are deducted from each student's score to construct an index (McLachlan, Finn et al. 2009). Longitudinal data established validity both at Durham and Cork medical schools, who conclude that it is an effective proxy measure for professionalism, an area where there currently are few such measures.

However, the index has raised controversy by calling into question the 'mechanistic' nature of such a measure and its validity when considered against the full scope of a clinician's role. As noted above, very high conscientiousness scores are associated with obsessive compulsive behaviours which may impair a clinician's ability to do their duties or work effectively with others in clinical teams, or for a student to complete their full range of tasks to a suitable but not excessive degree. Correlation with performance in more complex clinical environments has proved much less likely to be strongly correlated with conscientiousness, perhaps for this reason. In this study, it is likely that there will be a correlation between higher conscientiousness scores and achievement on knowledge tests in line with other findings in the early years of medical study.

2.5.14 Openness to experience dimension and PBL behaviours

Openness to Experience refers to how a person deals with new information at a personal and experiential level with higher scores associated with curiosity and enjoyment of new experiences and higher intelligence (Moutafi, Furnham et al. 2006). Lower scores, together with high conscientiousness have been correlated with a higher likelihood of Conservative or Republican political values! (Rentfrow,

Jost et al. 2009). The openness to experience can take a literal form, such as a study which showed than the inhabitants of a small Mediterranean island had significantly lower openness to experience scores than those of their relatives who had emigrated to the mainland (Camperio, Capiluppi et al. 2006). In spite of this dimension being included from the first NEO personality measure, it has been associated with mixed results in the literature, for example being correlated across all four of the MBTI measures, as well as with the Sensing-Thinking dimension. However, Costa et al wished to differentiate the facet of 'absorption', the propensity to become deeply attentive to subjects or experiences of interest, from the facets of extraversion which related to diverse but superficial interests. This quality is clearly of assistance in learning and sustained practice which is necessary for gaining academic qualifications and possibly for demonstrating intelligence through testing mechanisms.

PBL discussions are likely to be an appealing environment for learners who are high in openness to experience as there is likely to be more variety employed in giving explanations between eight participants, rather than a mode which focuses on one teacher. However, it may be more challenging to maintain absorption into the concerns of all other participants. Learners who are lower in this score may respond more effectively to an environment with more predictable rules and less variety of behaviour to contend with. This may be particularly true for learners who also have a high conscientiousness score which would magnify this challenge. However, the group may well benefit in their discussions from learners who are able to skilfully apply the correct steps of the PBL method and to ensure that the maximum is gained from the format. In this study, it may be possible to identify significant correlations between openness and the learner, particularly in conjunction with conscientiousness.

2.5.15 Neuroticism dimension and PBL behaviours

Emotional stability indicates the ability to deal effectively with negative emotions; where 'negative' emotions might be defined as comprising, fear (anxiety), anger and sadness, with happiness being defined as positive. Differences in emotional stability have of course been studied at the extremes, and defined as pathological psychiatric disorders. Some studies have shown high levels of anxiety and depression among medical students, particularly in the first years of medical school, (Jadoon 2010). McManus et al (McManus, Keeling et al. 2004) in their twelve year longitudinal study, found that 20% of doctors in their study would qualify as a psychiatric 'case' suitable for treatment. High neuroticism, low extraversion and low conscientiousness correlated with higher levels of stress. Thyssen et al (Tyssen, Dolatowski et al. 2007) broadly replicated these results, analysing extraversion, conscientiousness and neuroticism over six years of medical training.

A study at St George's studied the differences between their new GEMS PBL cohort and their traditional undergraduate entry cohort with respect to their anxiety on transition to clinical practice. (Hayes, Feather et al. 2004) When age and gender were controlled out of the statistical analysis, the PBL prepared graduate entry medical students were significantly less likely to express anxiety at the transition to clinical practice. In both groups, male students expressed lower anxiety levels than females. This finding is in keeping with the general findings in the literature that females consistently express more anxiety at each life stage (Lee Ellis 2008). Of course, such anxiety is usually measured by self-assessment and there are strong cultural influences which permit females to express anxiety and which deter males. Female students are particularly likely to express anxiety in connection with educational tests (Lee Ellis 2008). This PhD study will explore associations between gender and volatility as permitted within the sample.

2.5.16 Personality type and choice of career within medicine

Medicine begins by selecting the medical student onto a programme of general medical study, yet leads onto specialism in many diverse fields. In a recent study,

personality choice was correlated to speciality choice (Mehmood, Khan et al. 2013). Both males and females with low scores on agreeableness and high scores on novelty seeking, a facet of openness to experience, were drawn to surgery. Similarly empathy has been found to correlate with specialty choices with increased proportions of patient interaction (Hojat, Gonnella et al. 2002). However, as with much personality research, there remains a plethora of measurement instruments which, while relating to the 'big five', are not always transparent and interchangeable in use.

2.5.17 Personality and work performance

There is evidence that personality measures can be linked statistically with work performance (Hogan and Hogan 1989), (Day and Silverman 1989, Benson 2012). Two key measures are emotional stability and conscientiousness being linked to positive job performance across different occupational groups – when adjusted so that validity coefficients and personality constructs match. Extraversion has been shown to be positively related to performance in sales roles, and neuroticism has been shown to be a drawback in this context, being negatively correlated.

2.5.18 The distribution of FFM traits amongst medical students

A recent study of medical students, their personality type and whether they thought positively or negatively about PBL, compared 449 students in Norway, the US and Nepal (Holen, Manandhar et al. 2015). The following average distributions of personality type polarities were reported:

Table 11 FFM Dimensions: distribution for medical students n=449

FF Dimension	Mean: Scored 0 - 4	Standard Deviation
Extraversion	2.69	0.57
Agreeableness	2.75	0.54
Conscientiousness	2.87	0.53
Openness to experience	2.29	0.46
Neuroticism	1.56	0.74

Student personality types were self-reported using the NEO – FFI short form (60 item) questionnaire where strength of agreement from 0 (not true) to 2 (somewhat true) to 4 (very true) determined the position on the continuum. This distribution indicated that on average, students tended towards moderate (2 – 3) extraversion, agreeableness, and conscientiousness, milder openness to experience, and markedly lower neuroticism (1 – 2), although this measure had the widest spread according to the standard deviation reported.

2.5.19 Preferences for PBL and personality

Hierarchical linear regression analysis was used to identify correlations between personality types and their preferences towards or away from PBL. The only significant negative correlations (i.e.: personality trait for students who did not like PBL), were between neuroticism and PBL and conscientiousness and PBL. That is, the higher the neuroticism or conscientiousness score, the less they liked PBL. Students who were moderately extravert, moderately open to new experience and moderately conscientious felt positively about PBL. Some students with high scores

on the neuroticism dimension split between two positions, either liking PBL or very much not liking it. The authors suggest that belonging to a stable, close knit PBL group may have a mitigating benefit which some high neurotic students appreciated, however, some equally may conclude that they would prefer to avoid the extra stressors from PBL group work.

Some students with high scores on the conscientiousness dimension also disliked PBL. The authors suggested that the relatively unstructured and interdependent learning environment of PBL may be experienced as unpleasantly chaotic by these learners. There is some support for the view that high conscientiousness scores are associated with maladaptive behaviour, which may be diagnosed as Obsessive-compulsive Disorder, which may prove to be a challenge to manage and maintain well-being (Samuel and Widiger 2008). Students with high agreeableness scores, very much liked PBL.

The study was not designed to correlate personality type with learning attainment, but sought to explore cultural and social differences which may be associated with attitudes to PBL. However, the exploration linking personality type dimensions with positive and negative attitudes towards this learning activity does indicate useful comparators for distribution of personality types for a similar cohort. In addition, it is likely that a positive attitude to a method of study is indicative of a positive experience of learning, which tends to be associated with increasing motivation to study, leading to more time spent on the task of study (Slavin 2003), which is strongly correlated with better results.

Another recent study in the UK identified correlations between enjoyment, using the Biggs' Study Process questionnaire identifying deep surface and strategic learning styles, (SPQ) and FFM personality type among 99 medical undergraduates at St Georges. (Bigsby, McManus et al. 2013) The authors identified that less neurotic, more agreeable learners enjoy PBL more. Learners with a deep learning style (SPQ), being motivated by ideas and curiosity, rather than simply strategic desire to achieve

particular results, in combination with the above personality factors enjoyed PBL most.

2.5.20 Links between personality type and medical performance

One study of medical students at UCLA identified advantages to different personality measures at different stages of their longitudinal study of medical careers. Although initially demonstrating close predictive power for previous grade point averages, once measures of clinical practice and personal suitability were combined in assessing performance the picture became mixed and a single personality profile failed to demonstrate any predictive validity (Shen 1997).

2.5.21 Section summary

This section of the literature review has established what is meant by the FFM, summarising the history of personality research, giving an overview of the current situation in the field. The review linked each FFM measure of personality to factors which relate to learning through discussion, such as preference for groups, degrees of talkativeness, tendency to experience strong emotions, particularly anxiety, and performance and progression in careers, with particular reference to medicine when available.

In summary, there is much evidence to suggest that speech behaviour will be capable of demonstrating significant differences between learners of different personality types, through corpus analysis, and statistical analysis of comparable measures of speech quantity and turns taken, and quality, such as complexity/simplicity, proportion and types of questions.

The following section will consider corpus analysis and applied linguistics and consider their role in this study.

2.6 Applied linguistics and Corpus analysis

2.6.1 Introduction

This section of the literature review will establish how corpus analysis has been used to identify and measure aspects of learner speech in order to demonstrate a gap in the literature to which this study may contribute novel findings. First, this section will introduce applied linguistics and corpus analysis and describe current relevant definitions within the field and the distinctions between written and spoken corpora. There will be a brief discussion of the challenges of preparing a spoken transcription, and methods of measuring speech behaviour, which will be further discussed in the methods chapter. Corpus analytical studies will then be considered which concern medical subjects, or similar populations, particularly medical or life sciences students, engaged in learning through discussion. Studies which concern clinicians at work, links with performance in assessment or where they concern evidence for differences in personality will also be considered. This section will conclude with a brief summary and anticipate detailed explanation in the methods, corpus results and discussion chapters of this study.

2.6.2 Definitions

Applied linguistics, is defined within the Longman dictionary of language and applied linguistics (Richards and Schmidt 2013) as, 'the study of language and linguistics in relation to practical problems'. Corpus analysis, describes the process of analysing a corpus or several corpora, as the bodies of text which have been written or spoken and transcribed are known. Combining the two terms, corpus linguistics is a recent field, which emphasises the methodology. It has grown rapidly, as computer-based methods have emerged as more powerful means of conducting analyses and have provided a viable approach to studying language use in real world situations.

Biber et al (Biber 2006) describe corpus linguistics as having four main features:

1) it is an empirical (experiment-based) approach in which patterns of language use that are observed in real language texts (spoken and written) are analysed

2) it uses a representative sample of the target language stored as an electronic database (a corpus) as the basis for the analysis

3) it relies on computer software to count linguistics patterns

4) it depends on both quantitative and qualitative analytical techniques to interpret the findings

As Hunston (Hunston 2002) baldly states: "...a corpus by itself can do nothing at all, being nothing more than a store of used language" (p.3). There are many issues relating to the creation of corpora for study, and equally many issues relating to the use which may be made of this 'store of used language'. This study conducts content analysis, as the focus is on identifying and analysing the occurrences of particular aspects contained within the text; particularly the frequency of terms, the parts of speech and the semantic topics of the text.

2.6.3 Written corpora

Analysing word clouds has become a more commonplace way of objectively depicting the relative importance of terms within a complex text. Gill and Griffin (Gill and Griffin 2010) demonstrate this by examining documents outlining good medical practice for doctors in the UK between 1963 and 2010. Park et al (Park, Griffin et al. 2012) discuss the potential for applying such techniques within medical education to look at texts which may be familiar to us in some ways, in a powerful new light. Words can be analysed in a literal sense: the 'what', but also the 'how', which leads us into more interpretative realms or hermeneutics, the theory and practice of interpretation, in order to interpret texts on many levels.

More complex analysis is the domain of specially devised computation tools which have been in development over the past forty years. Third generation corpus

analysis tools allow us to mine the data within a corpus in order to determine ever more complex patterns within the language which give information about the underlying thoughts or preoccupations of the speakers. They have even been likened to the bringing of telescopes to bear in astronomy for their potential in revealing what previously was hidden within language (Hunston 2002).

Leech, (Leech 1992) argues that when sampling genuine language, particularly, spoken language, errors must not be corrected or eliminated but analysed as part of the rich data set. In addition, if a sample of speech is large enough, then any errors in language would be prevented from having a negative impact.

Opinions vary on the extent to which one should annotate the naturally occurring text. Traditional grammarians, such as Professor Quirk, (Quirk 1986) advise rigorous annotation to assist in linguistic analysis, but more recent eminent scholars such as Sinclair, suggest minimising this process (Sinclair 2004). Professor Quirk headed the Survey of English Usage Group at UCL, which compiled 0.5 million words from spoken sources in the pre-computer era, to form the London-Lund Corpus (LLC). This was being compiled in the same building, at the same time as I was studying English Literature at UCL. I observed evidence of the corpus as an enormous number of grey metal filing cabinets throughout the four floors of the English department. Collecting, transcribing and annotating, all were done by hand.

2.6.4 Discussion of recent varieties of corpus analyses

Corpus linguistics and computational linguistics were able to collaborate on approaches, a history recounted in Adolphs, 2006 (Adolphs 2006). Rayson (Rayson 2008) recommends a three stage process in corpus analysis: Annotation, Abstraction, Analysis. Annotation is the stage where descriptive 'tags' are added to the text which render the text machine searchable (Wallis and Nelson 2001). The next stage is to abstract the data, creating a searchable dataset. Analysis is then possible in both quantitative (frequency) and qualitative (individual phrases for quotation) approaches which might be used to address particular research

questions. For example, the use of personal pronouns in the introductory sections of lectures in science and the arts (Yeo and Ting 2014).

Rayson describes micro and macro approaches to analysing the text. Macro analysis focuses on what is similar or different across different texts or languages. Micro analysis builds on the traditional detailed approach where research focusses on the word and specific grammatical structures within the text. Rayson proposes that these two approaches may be combined, as each iteration of analysis uncovers data in relation to an initial research plan, whereupon further aspects and questions may be formulated and explored (Rayson 2008).

2.6.5 Electronic content analysis

Automated content analysis software can code text automatically, which eradicates the issue of inter-coder reliability as all codes, once established, are applied objectively. Groups of words can be grouped together into pre-determined 'themes'.

There are different approaches to completing the same type of task. Some software performs more effectively in concordancing, frequency lists or key word lists, such as Wmatrix3. Others can identify when words occur in time sequence (Wordsmith). The choice of software depends on the nature of the research questions (Kondracki, Wellman et al. 2002) and (Adolphs 2006). The ability to move swiftly between detailed statistical information and qualitative concordances may be important in mixed study approaches.

2.6.6 Written and Spoken Corpora

It has now become feasible to conduct corpus analysis on vast quantities of text which are stored digitally. There are numerous studies, such as the study of the development of character in the works of Dickens (Mahlberg 2013), or, a recent study of argumentation in Prime Minister's Questions, (Sealey and Bates 2016), which make use of texts which have already been made available digitally, but not specifically for the purpose of corpus analysis. However, spoken word corpora are

rare. As Pastizzo et al assert, even the most common analysis, that of spoken word frequency analysis, "is conspicuously absent in the literature" (Pastizzo & Carbone, 2007 (p. 1025)).

2.6.7 Speech corpora

The Cambridge and Nottingham Corpus of Discourse in English (CANCODE) is a corpus of five million words of spoken English, recorded in naturalistic settings. Participants were aware of being recorded, but were sampled from a range of relatively informal settings such as at the hairdresser, during a meeting. This permits data to demonstrate features of English in use which certainly differ from how English language is taught to non-English speakers. One example, noted by Channel is the frequency of what is termed, 'vague language' where the speaker 'softens' the effect of reporting facts through adding terms such as 'hereabouts', 'about twenty people', 'we'll meet at sixish' (Channel 1994). The Director of the project, notes that there are inherent limitations, such as an unrealistically positive focus of talk, as conflictual subjects may well have been avoided in the presence of the recording (Carter 1998).

2.6.8 Educational speech corpora

There are several precious and large scale academic corpora, including some speech, which have been established. The T2K-SWAL corpus consists of 2.7 million words captured at four US universities. Unusually for an academic corpus, almost 1.7 million of those words are from recorded spoken sources, while only 1 million originated in written material (Biber 2006). The Michigan Corpus of Academic Spoken English, (MiCase) is a searchable resource of over 1.8 million words where one can select according to discipline, level and interactivity of the setting. Below is an extract from a transcription which represents one of the closest models; an interactive discussion among Junior Graduate students in Biology.

2.6.9 Educational speech corpora

The extract follows the standard notation, showing a conversation between two speakers, where each change of speaker is shown by beginning a new line, and S for speaker, together with the identifying number. In this discussion, two students in a study group are discussing their progress. S1: denotes speaker one, and S2: speaker 2 (Micase 2007).

S1: sorry we could only get the old book.

S2: oh that's okay.

S1: okay. <*PAUSE:09*> so, we have questions on electron transport or what?

S2: um, you kn- that oxyl acetate, you know that, there's like a, a weird cycle. it's, it sorta seems separate

S1: citric acid cycle?

S2: no, it's something, sort of related to it but it's not, it. [S1: okay] like, i was reading in the book. it's like

S1: oh, [S2: the] the the t- okay i know what you're talking about

S2: okay, i just...

S1: for the fatty acid breakdown right?

S2: uh yeah, i think so

S1: okay do you know what chapter that is? it's (way) past this

Title: Biochemistry Study Group
Academic Division: Biological and Health Sciences
Publisher: Michigan Corpus of Academic Spoken English, English Language Institute, University of Michigan
Interactivity Rating: Highly interactive
Number of Participants: Students: 4 Speakers: 5
Recording Date: November 6, 2000
Recording Duration: 109 minutes
Word Count: 16192

Figure 9 A typical sample from an educational speech corpus

This extract serves to illustrate many of the features of spoken corpora. The phrases are short, colloquial, often incomplete and sometimes overlapping with

other people's speech, although this extract only shows this occurring once, when S2 interjects into S1's utterance, at line 6.

In order to determine key words in context and identify the salient features of a new spoken corpus, it will be necessary to select a suitable spoken corpus as a comparator for the analysis. The British National Corpus, (BNC), contains a total of 100 million words, of which 10% are transcribed from spoken, naturalistic sources, from the late twentieth century. The BNC Sampler, is a subset of one million written, and a one million spoken word corpus (BNC 2007). The sub-set includes educational discussions of science and assessment and is offered in the Wmatrix3 programme as a comparator corpus. The British Academic Spoken English (BASE) corpus contains 20 seminars of which 7 have been transcribed (Thompson and Nesi 2001). It was developed at the Universities of Warwick and Reading and contains a high proportion of native English Language speakers.

2.6.10 Studies observing PBL processes

A small number of studies have conducted PBL observations and analysed the contributions of students. Rivard and Straw considered the role of talk as opposed to writing and compared both individual conditions with a combination of talking and writing in scientific conceptual development (Rivard and Straw 2000). Students of all levels of ability who had the opportunity to discuss their ideas, outperformed students who did not, both on immediate recall and subsequent tests. Students who had the opportunity to talk and subsequently formulate their ideas through writing performed the best. The authors suggested that elaboration through talk is an important stage in elaborating understanding but that the writing stage provides structured repetition which enhances learning further. Grave et al demonstrated similar effects with a PBL discussion prior to a comprehension test on a related text, leading to enhanced recall, comprehension and suggested that the discussion motivated further self-study through being relevant (Grave, Dolmans et al. 2002).

2.6.11 Studies classifying verbal interactions during learning discussions

There have been a select number of studies over the past fifteen years which have sought to describe, codify, and build understanding of clinical reasoning processes by studying transcripts of student speech. However, very few have specifically sought to identify speech behaviours of individual students over a full cycle of PBL.

Van Boxtel et al (Van Boxtel, Van der Linden et al. 2000) evolved an episodic coding system which categorised utterances from verbatim transcripts taped from discussion sessions. Inter-rater agreement for the coding system applied by two judges was high, demonstrating 89% agreement. The categories they identified were specific behaviours such as questioning, conflict resolution and reasoning, with or without elaboration. The learners were in a school physics class. The researchers noted that the process was very time consuming.

Several authors have sought to categorise how learners employ critical thinking, and clinical reasoning during PBL discussion using verbatim transcripts and coding systems. Jamieson, applied a theoretical model and an adapted coding system to examine how different learners elaborated their understanding (Jamieson 2013). Tokode (Tokode 2016) applied a theoretically derived coding system to identify group constructions of learning in PBL discussions, at Derby GEMS.

2.6.12 Transcription studies linking PBL discussion with attainment

Yew and Schmidt aimed to provide a more detailed analysis of student speech during PBL discussions, including self-directed learning episodes with no facilitator present (Yew & Schmidt, 2007). The authors extended Van Boxtel's coding tool to provide a closer fit with categories and thereby enhance objectivity. This coding system was trialled on transcripts from a modified PBL learning event in Singapore held over one day with one problem. Notably, the periods of self-study which are usually unobserved elements within PBL were also videotaped and translated into

text scripts. There were again two judges with an inter rater agreement of code application reaching 80%. Again, conflicts and resolution, critical questioning and elaborations were coded. Individual student speech and non-verbal behaviour was linked to performance in a post case test. Different patterns in the use of questioning were identified when comparing high scorers on the knowledge test with those gaining low scores. Students asking questions related to verifying data or making requests were associated with lower scores. Students asking more complex questions were significantly correlated with higher scores.

The authors call for further research with larger groups of students, their study involved just five students, with verbal interactions amounting to 7.5 hours of recording with a total of 1075 utterances, coded as 349 episodes. The authors suggest that the statistical significance of their results suggests promising territory for future research which could generate bigger effect size through studying a larger number of subjects. They also note that implications for PBL facilitators may include modelling under-represented approaches, such as guiding learners to deal more effectively with conflicts in knowledge and ideas and posing critical questions and seeking clearer explanations from learners. The current project may be able to replicate significant associations between questioning and attainment over a larger number of subjects.

2.6.13 PBL transcription studies linked with speech behaviour

Visschers-Pleijers et al (Visschers-Pleijers, Dolmans et al. 2004) reduced the categories in Van-Boxtel's tool to five categories and applied it to one session of PBL, achieving lower inter-rater agreement of 71%. The length of time required to undertake the analysis was again cited as a drawback which would severely limit future study.

Wmatrix is a linguistics programme which was developed at Lancaster and has been used in a number of studies in applied linguistics. The coding of language use is

done automatically by the tagging programme, especially for parts of speech and semantic categories. Da Silva and Dennick demonstrated viable use of Wmatrix2 to analyse one cycle of three PBL sessions for one first year group of Derby GEMS (Da Silva and Dennick 2010). This exploratory study described how specific relevant technical vocabulary is employed in relation to the PBL cycle within one group. Sessions became longer as the cycle progressed and comprised between 1000 and 3000 specialist, technical terms.

By analysing the frequency of occurrence of technical nouns, they were able to show that the first session focussed on initial considerations related to, the case (alcoholic liver damage); the second session dwelt on the underlying physiological mechanisms, and the final session deepened the potential medical implications of liver function changes.

By analysing the semantic categories which were used most frequently they were able to confirm that initial discussions dwelt on probability and disease. The analysis showed a statistically significant increase in discussion of anatomical topics and cause and effect in the middle session, and a move towards discussion of treatment in the final session. The authors were also able to use parts of speech analysis to explore the pattern of use of terms associated with clinical reasoning, such as subordinating conjunctions phrases (that, if, so, because) and question words (why, what, how). The initial session over-used 'so' and 'then' terms associated with describing sequence, and the question words 'how' and 'what'. The middle session demonstrated a statistically significant increase in the use of 'why', 'where' and 'because' terms associated with understanding physiology. The final session showed a marked decrease in questions overall and subordinating conjunctions. This paper concluded by calling for twelve future lines of research enquiry to be continued and posed the following questions relating to individuals during PBL discussions (Da Silva and Dennick 2010).

Contribution by individual learners to PBL discussions

1. What proportion of the total discussion is contributed by individual students?
2. Is the quality of a student's contribution, such as questions generated or explanations given, associated with his or her learning style, academic background, personality or other variables?
3. Is the quantity or quality of an individual's contribution to the collaborative learning process a valid predictor of academic success in knowledge and clinical reasoning assessment?
4. Could corpus analysis be used to formally assess a student's contribution towards the PBL process?
5. How can information obtained from analysing transcripts of PBL sessions be fed back to individual students, groups of students and the facilitator to improve their PBL performance?
6. How do students use their language skills (e.g. imagery, metaphor, etc.) to construct explanations for concepts?
7. How do the technical vocabulary and reasoning skills of individual students vary and develop over time?
8. How do the technical vocabulary and reasoning skills of PBL groups develop and change over time as they move through cycles of PBL scenarios?
9. Do student discussions cover all the learning objectives and factual and conceptual content provided in the guides and other resource material provided for each scenario?

2.6.14 Response to da Silva and Dennick's future research questions

This PhD study may contribute evidence towards the first five questions relating to the individual and their speech behaviour during PBL discussions. In addition, the methodology and the spoken word corpus itself, contributes to future approaches for

questions 7 to 9, with the findings and corpus analyses of this spoken corpus contributing one measure, taken at the earliest feasible moment in medical school, against which future development may be measured, although the exact individuals have now themselves (mostly) graduated and become doctors.

In relation to question number 6, this corpus was analysed to identify patterns in student explanations, seeking examples of similes and metaphors. The almost complete lack of examples of such use of figurative language in the entire corpus was itself significant. The findings were presented at the medical education conference in Prague (Haines 2013). Although this meant that this avenue of exploration was at an end, it did shed light on the nature of medical language acquisition in first year medical students. They are simply too concerned with mastering the basic terminology to be able to use figurative speech, which is perhaps, along which higher order question use such as 'if, and when' indicates that the basics have been mastered. For these first year students when discussing a demanding PBL scenario part way through their first year at medical school, this was simply not observed. It remains to be investigated whether students themselves benefit from the explanations which facilitators and teachers routinely give using similes and metaphors in order to explain what is unknown in terms of the known. Perhaps learners are not yet ready to benefit from this approach whilst mastering the basic medical terminology. This is in keeping with findings in language acquisition.

2.6.15 Corpus analysis of the role of language in educational discussions across differing disciplines.

One recent PhD thesis is based on a study of differentiating language features between different disciplinary seminars, uses the BASE corpus (Tanguay 2015). However, most publications listed on the BASE site relate to analysis of spoken language in the lecture environment, rather than the more interactive environment of seminars. PBL discussions would qualify as the most interactive of all small group

discussions, as close teamwork and discussion is essential to the format and participation is demanded to some degree from each participant.

From the pre-computerised era of corpus analysis, one PhD study from 1989, (Lukar 1989) analysed student discussions from across a range of disciplines at the University of Nottingham, and compared the proportion of speech contributed by students as opposed to facilitators. The proportion of speech was found to vary markedly between different disciplines.

One of the highest proportions of student discussion was found among the undergraduate life sciences students, the highest of the scientific participation rates. Lukar quantifies the proportion of small group teaching experienced by students at this era in Physiology to represent 25% of their first and second year curriculum. This is comparable to the proportion of PBL discussion which is at the centre of the Derby GEMS curriculum, set within a balanced series of lectures, clinical skill and other experiences. This average proportion of study through discussion was observed across the Faculty of Medicine and equalled the measure for social sciences at that time. However, the overall average of student talk across the series of twenty classes which were analysed was 33.25%, whereas the total lecturer talk was 60.55%. When considered by subject, even in small groups discussing cases in physiology, the total teacher talk time remained high at 60%. The postgraduate courses did record higher scores for student participation in talk, for example, education theory and method the highest scores were recorded at 64% and 68%. Not all postgraduate courses automatically scored high on student talk. Education method, scored only 8% for student talk.

Lukar also counted the number of teacher questions, but not student questions, but included teacher responses. When compared to other subjects the physiology class teachers favoured questions which were categorised as analytical or evaluative only 10% of the time. The category of problem solving, was found to be under represented in many disciplines and relatively low in physiology. The probe question

was found to be represented strongly, particularly in physiology. However, the author states that these sequences often followed an unsuccessful initial question which was either poorly phrased or pitched at a level where the students were unable to respond. The physiology small group session was also unique among the 32 groups studied in that the intention by the teacher was for the student understanding to be central and so has much in common with a PBL small group discussion. The lecturer stated the aims as follows:

> *'by going some way towards solving problems based on medical case histories the students will a: gain perspective on the use of drugs (learned about theoretically) in real situations, b: specify particular areas of difficulty in treating simultaneously with more than one drug, c: establish criteria which can be used to decide when treatment is essential, inessential and undesirable.' p. 237*
>
> *(Lukar 1989)*

When the lecturer talk was examined in these physiology sessions it was found to have the lowest proportion by far of the teacher lecturing compared to all other disciplines. For the two physiology classes, out of 60% or 58% of the word count, the teacher was giving a 'mini lecture' for 19% or 18% of that time. The teacher was asking questions for 13% or 14% of the time and responding to student questions for the remaining 28% and 26% of the time.

2.6.16 Individual student talk analysis in one group

Lukar did study one small group in detail to identify the amount of talk, measured in time, per individual student. One group of seven Post Graduate education students contributed 64% of the session talk overall. Individual contributions ranged from 2.90 to 12.60% of the discussion between three talkative students, two reticent students and two moderately contributing students. Lukar notes that much of the interaction is between students responding to each other's comments or silence, rather than to teacher questions. The class was based on student presentations by a smaller number of students which were then discussed with their student audience. The length of utterance is not recorded, so the role and teaching method

are chiefly determining the amount of student talk likely from each individual student on this occasion. Where a larger sample of talk from each group might be taken, these differences due to role might become less pronounced and be more indicative of a student's own preference for speaking out in groups.

Lukar summarises her findings identifying age and experience as being associated with higher amounts of student talk, with student preparation being associated with higher participation, and where the expectation of the students' role determined what might be expected in terms of spoken contribution on any one occasion.

2.6.17 Studies which examine informal writing, teenagers seeking health advice

One example of using corpus analysis in the healthcare setting is that published by Harvey et al, (Harvey, Brown et al. 2007) describing a study of a 400,000 word corpus created from questions submitted by teenagers to a confidential health concern advice web site. The paper identifies this corpus as providing very rare access to self-documenting, highly-specific information about teenagers' health concerns which would be very unlikely to be expressed in the traditional healthcare setting with adult healthcare providers. The analysis was able to identify pervasive themes of preoccupation with normality and worry, and meticulous description where face to face sources have previously identified use of vague language. The study used illustrative quotations to support their findings.

2.6.18 Performance features in clinical skills assessment

The GMC commissioned a study of the linguistic and cultural factors in the membership of the Royal College of General Practitioners examination (Roberts 2014). The study aimed to:

1. 'understand the performance features of the Clinical Skills Assessment (CSA), part of the licensing exam for all GPs in the UK, and, in particular, to

investigate the extent to which linguistic/cultural factors contribute to poor performance.

2. Develop an analytic framework and design associated materials to raise awareness among examiners, GP trainers and candidates of the linguistic and cultural demands of the exam.'

There were video recordings made of 198 candidates at the SCA in March and May 2011. A detailed analysis was prepared of 40 cases (shorter than ten minutes per case) creating a spoken word corpus of 85,000 words from which clear patterns can be identified and statistically evaluated against larger datasets. The authors state that:

> *'the generalisability of this method comes from the fact that within a specialist dataset such as this, 85,000 words begin to exhibit linguistic patterning that is sufficiently repetitive to make statistical claims about the way talk is characterised in this setting and to find clear distinctions from general language use.'*
>
> *'The video data set is relatively small for corpus linguistics …. By contrast 40 cases is a large data base for micro-analysis and only certain features, identified in previous intercultural research, could analyse in depth."*
>
> *'..allows for applicability of concepts, patterns and phenomena to other contexts, provided these contexts meet the same conditions as those in the original data.'(Roberts 2014) p. xvii)*

Broad analyses revealed little difference in the structure of consultations for poorer candidates as opposed to high scoring candidates. All groups used similar phrases. However, poorer candidates continued for longer in gathering data, and were more likely to be interrupted by the simulated patient seeking clarification when giving explanations. Cases were initially annotated blind, that is, the annotators did not know if the candidate had received high or low marks from the assessors, in order to mitigate against confirmation bias. Broad phases were identified when candidates were concerned through the simulated consultation with: data gathering, physical examination, explanation, discussion. The researchers identified episodes of alignment, misalignment, misunderstanding, display of 'exam techniques', and

handling ethically challenging situations. The corpus analysis software was able to identify the frequency of certain key phrases and provide a quantitative basis for comparing unsuccessful and successful candidates.

When micro-level analysis was performed using Wordsmith, similar formulaic phrases were used by all candidates, however, successful candidates used them in ways which seemed slightly more customised and empathetic in the circumstances. For example: successful candidates used the stock phrase 'tell me a bit more about' early in the consultation and did not appear to be buying thinking time through getting the simulated patient to speak. Overall, candidates spoke for 68% of the time and this did not differ markedly for candidates with different results. Successful candidates were quicker to identify when a misunderstanding had occurred and repaired this more swiftly and were consequently interrupted by the simulated patient on fewer occasions. When cases required discussion of ethical issues, more time was allocated to this by the candidate by limiting the data gathering and treatment management elements of the usual consultation structure.

The authors conclude that there are small but cumulative elements in the linguistic and cultural domain which impact on the performance of the candidates in the Clinical Skills Assessment (CSA) and that targeted training and support could be usefully directed towards candidates, their medical educators and examiners to assist in developing assessments of such communication heavy domains.

It is anticipated that the current study may identify statistically significant patterns of communication which may be found to correlate with higher or lower performance particularly in the domain of questioning and focus on appropriate semantic topics.

2.6.19 Studies which examine personality through transcribed or self-documenting speech behaviour

The focus of the current study is on demonstrating how students of different types behave differently through their speech. One rapidly developing area applies some

of the tools of corpus analysis to self-documenting corpora, as discussed above in considering the teenage health online questionnaire, but which seeks to identify evidence for personality traits through naturally occurring self-documenting corpora, such as blogs. This side-steps the intensely laborious stage of transcribing and verifying speech yet still provides access to very large naturally occurring language corpora, which may capture some of the informal characteristics of speech corpora.

For example, Iacobelli et al, used blogs to identify 3000 bloggers and stratified their corpora according to high or low scores on FFM personality measures (Iacobelli, Gill et al. 2011). The authors discuss the difficulty of identifying suitable comparator corpora within the BNC as they were capturing language which had not been susceptible to existing corpora.

2.6.20 Personality and language use

Iacobelli present a table which summarise their findings linking personality measures with typically observed written language behaviour, as demonstrated in the context of essays (Pennebaker and King 1999); for blogs (Nowson 2006): and for emails: (Oberlander and Gill 2006). Creating a spoken corpus of PBL discussion individual speakers with a known personality style, would allow these findings to be further examined in the novel context of the PBL discussion.

Table 12 *Personality and language use adapted from (*Iacobelli, Gill et al. 2011*)*

Trait	Type	High Scorer	Low Scorer
Extravert **Quintax uses same terminology**	Behaviour	Warm, assertive, action oriented; thrill seeking	Introverts: low key; deliberate; easily stimulated
	Speech behaviour	Use social words, refers to self and other, positive emotion words, certainty, less tentative, more present tense verbs, reference to communication	Negations and negative emotional expressions, causation words, greater tentativeness, (email) achievements,
Agreeableness (Personable as opposed to logical in Quintax)	Behaviour	Compassionate, cooperative, considerate, friendly	Suspicious, unfriendly wary, antagonistic, uncooperative
	Speech Behaviour	Uses first person singular and positive emotion words (essays)	Negative emotion words, talk about body states (blogs)
Conscientious-ness (structured)	Behaviour	Disciplined, dutiful, persistent, compulsive, perfectionist	Spontaneous, impulsive, achievement less important
	Speech behaviour	Use of first person singular and positive emotions words on essays	Negative emotion on essays, talk about body states (blogs)
Openness **(grounded v theoretical, in Quintax)**	Behaviour	Appreciates art and ideas, aware of feelings, imaginative	Straightforward interests, resists change
	Speech behaviour	Uses longer words, expresses positive feelings, (essays) uses inclusive terms (blogs)	First person singular, present tense more causation, negations (blogs)
Neurotic	Behaviour	Emotional instability, anxious, hostile prone to depression.	Emotional stability; calm, less easily upset.
	Speech behaviour	Uses first person singular (I) and negatively associated terms, discrepancies, multiple punctuation on emails.	Refers to others on blogs, uses more nouns and adverbs on email

2.6.21 Further studies using corpus analysis to link language with personality

In Yarkoni's study – boldly titled, 'Personality in 100,000 words', 694 individuals were recruited to contribute random samples of their blogs to be studied in connection with their personality type (Yarkoni 2010). An email invited them to visit a web page where 83% completed a short or long version personality questionnaire (NEO PFI questionnaire). The population was strongly skewed towards female respondents, who accounted for 524 of the subjects. Participants were then invited to compose two emails to a close friend, describing activities the week before and in the week ahead. The authors argue that such communication has been considered as being comparable to everyday speech.

The data set was analysed for the top 5000 occurring words. The authors demonstrated consistent evidence for linking both individual words and semantic categories with personality traits both at the FFM level, and with many of the subsidiary facets. Particular findings included a strong positive correlation between length of words and degree of openness on the FFM. Participants with high agreeableness scores also used fewer swear words. Participants with high extraversion used more socially oriented words.

Oberlander and Gill (Oberlander and Gill 2006) built a 65,000 word corpus of student emails using an opportunistic sample of University students whose personality types were already known. A total of 105 students were invited to write two short emails, as if describing events of the past week or the coming week to a good friend. The corpus was further stratified according to high or low scores on the personality dimensions but that did result in very small sub-corpora, for example 6,000 words for high extraverts. When analysed using key words in context (KWIC), to calculate the log likelihood using wmatrix, (Rayson 2003) consistent findings in line with the previous literature were found. For example, extraverts used the informal greeting

form 'Hi' and many more exclamation marks, than did introverts, those with low extraversion scores. Low extraversion scorers (introverts) used 'hello' as a greeting, and more quantifying terms, such as 'how much' how many'. Their verb use differed also. High extraverts favoured future verb forms such as 'I'll be' and 'would be' and avoided negatives. Introverts preferred statements such as 'I think', 'I've done', 'I am' expressing autonomy of action. They also used more negatives, such as 'I don't', 'I didn't' and 'I should be'. The work reinforces the evidence for the feasibility of demonstrating personality difference through corpus analysis of text, building on earlier work, particularly that of Pennebaker (Pennebaker and King 1999). The authors conclude by recommending that the role of gender be investigated in future studies.

Wmatrix would allow all of these findings to be replicated or otherwise, using a student PBL corpus which identified individual contributions and gender and is discussed further in the methods chapter.

2.6.22 Linguistic studies of gender use, particularly through corpus analysis

There are widespread stereotypes about the differences between male and female styles of communication. However, there is a growing body of scholarly literature which demonstrates persistent, strong evidence for differences in the range of communication strategies used by speakers of different gender, which is still discernible in mixed discussions (Tannen 2001, Tannen 2002, Schiffrin, Tannen et al. 2008). Developments in corpus analysis permit us to use increasingly large data sets which leads to stronger more generalisable evidence than from many earlier studies.

In overview, women have often been found to have more social focus in their language, using speech to build and support social relationships. Men have often been found to have a more instrumental focus, particularly on information giving. This has been broadly summarised as a female rapport versus a male report style

(Newman, Groom et al. 2008). These authors propose that one key marker of this difference is in the difference in the use of 'content' language, (nouns, verbs and adjectives) as opposed to 'function' words. In English there are only around 200 commonly used function words, such as pronouns, articles, prepositions, conjunctions and auxiliary verbs. These words are generally used to combine content words together. As they occur more frequently, they may provide rich opportunities to generate a large data set and show up differences in use. Argamon et al (Argamon, Koppel et al. 2003) used a prediction equation to classify writing by gender, held in the BNC, with 80% accuracy. Function words were the strongest indicating category. Female language has been found to contain more references to positive emotions, whereas male language contains more references to anger (Mehl and Pennebaker 2003). Biber (Biber and Reppen 1998) used parts of speech to create an index to demonstrate that female language was more 'involved' than male language. The involvement index identified an involved style as being rich in the present tense with more pronoun use, whereas, the opposite, more nouns and longer words, was consistent with male language use. Analysis of parts of speech, such as that offered within Wmatrix3 may prove to be a useful tool for this current study in order to demonstrate and explore gender differences in a PBL speech corpus.

2.6.23 Gender and language use in clinical consultations

Roter et al conducted a meta-analysis of studies with the aim of quantifying the effect of physician gender on communication during medical visits, published during the years 1967 to 2001 (Roter, Hall et al. 2002). Of the 23 observational studies, 29 publications and 3 large physician self-reported data sets which qualified for inclusion, the main sources were from the primary care setting. The broad conclusion was that female physicians engage in more patient-centred communication than their male counterparts, taking an average of 2 minutes longer per consultation, representing 10% longer. The exception were male physicians in obstetrics and gynaecology, whose consultations were closer when analysed to the

pattern observed in female physicians. It is of course true that most patients in the O&G consultations would be women, and that one aspect of skilled communication is to have the skill for the speaker to adopt attributes in common with the listeners in order to communicate more successfully. The factors which were measured were pooled and compared and the overall average consultation time was calculated.

2.6.24 Conceptual Groupings of Physician Communication Categories

The authors present their findings according to thirteen dimensions, within three categories, as follows:

Information Giving

- Biomedical
- Psychosocial
- Directive
- Nondirective
- Quality

Question Asking

- General
- Biomedical
- Psychosocial
- Closed-ended
- Open-ended

Partnership Behaviours

- Active
- Passive

Socioemotional Behaviour

- Social Conversation
- Positive Talk

- Negative Talk
- Emotionally Focused Talk
- Positive Nonverbal

Length of Visit

The following table is adapted from the paper to show the three categories of talk used to group the thirteen dimensions, as described above. The following figure shows eight of the 13 dimensions, labelled in red, to present a simplified summary. The error bar position indicates the effect size; most marked at +0.2 for female use of positive talk; and -0.1 for male use of directive talk.

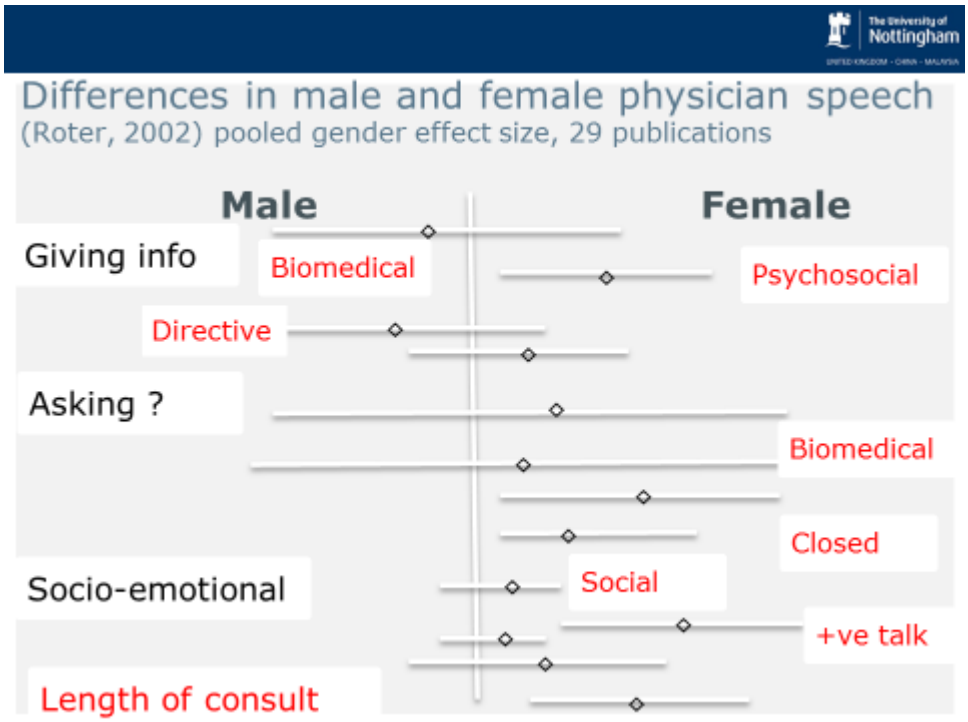


Figure 10 Presenting a summary of meta analyses of gender differences in health care consultations

For every category which was evaluated, men performed just two dimensions more often: giving directive information, and giving bio-medical explanations. Both these dimensions belong in the giving information category. Female physicians were measured to have used all other measured aspects more frequently.

In their extensive, meta-analysis, Roter et al, compile a comprehensive picture of communication which occurs between doctor and patient across four decades. Significant and consistent gender differences were identified between the modes of communication from genuine consultations.

Differences in gendered communication may already exist early in medical school and may be perpetuated during small group teaching such as PBL settings, as they rely so heavily on mixed-gender small group discussion. It may be helpful for learners and facilitators to become aware of differences, or rather restrictions in their range of available strategies for communication. It is logical to argue that medical students require the opportunity to develop a range of effective communication skills for their future work with patients of both genders in the clinical setting. This PhD study may be able to identify and describe gender differences in communication by gender which support and inform such development.

2.6.25 Gender differences in questioning style

Differences in questioning style between the genders were also detected in a study by Hawkins and Power (Hawkins and Power 1999). The authors transcribed 18 hours of small group discussion held in mixed groups comprising 4 – 7 members. Discussion centred on a decision-making task. The researchers found that significantly more probing questions were asked by females when compared to males, which they proposed was consistent with findings that female language is often directed towards supporting communication. The probing question, invites an elaboration of the speaker's initial utterance and can be both supportive and lead to more data being available to the whole group. In overview, 15.6% of the sentences were questions, of which, 89.2% are probing, using Bales' original 1957 classification. Learners and facilitators benefit from developing their questioning skills. As discussed, in the section on educational concepts in general, question use is a significant marker for internal cognitive processes.

2.6.26 Summary of the section

This section of the literature review has established how corpus analysis has been used to identify and measure aspects of writing and speech in relevant populations. Where possible, this has related to learning through discussion in higher education, PBL or seminars in closely allied disciplines. The section has also reviewed studies where spoken or similar corpora have been used in relation to personality, learning style and gender. It has identified clear gaps in the current knowledge base and specific calls to pursue research relating to the aims and objectives of this study.

2.7 Literature review conclusions

These four key sections have been reviewed through the relevant literature.

- Educational concepts in the PBL setting
- Key concepts and definitions relating to individual difference, learning, especially in groups, and educational attainment
- Psychological concepts in the PBL setting, personality and learning styles
- Corpus analysis, the creation of spoken word corpora and application of linguistics in educational and medical education settings

The results of the literature review have now been presented by considering each of these sections in turn, establishing key definitions and debates. Key studies have been synthesised, analysed and evaluated with respect to relevance to this study, according to the stated literature review methodology.

This literature review therefore provides an important starting point for establishing what different types of individual learners say during one PBL cycle, and what that might suggest about the speech behaviours of the different genders, personality types, learning styles, higher and lower scoring learners, through PBL discussion in the early years of medical study. Through providing and analysing evidence of genuine, naturalistically gathered speech, it may be possible to demonstrate

significant differences between individuals as they learn through PBL. This may give further insight as to factors which are associated with effectiveness in learning. This understanding of the individual characteristics of learners may well inform PBL facilitators and PBL learners in developing their understanding and application of the processes of PBL learning, thereby meeting the overall research aim to better describe and understand the differences in individual speech when learning through PBL discussion.

This literature review has identified no single overarching theory within which the study of learners through discussion may be considered. In each of the four key areas, there are multiple theories, and a number of different theoretical frameworks which underpin the research within the field. This study may contribute towards testing certain areas of theory and the discussion will return to briefly consider where evidence supports or refutes certain theories.

Finally, this literature review demonstrates that there is a gap in the literature in this area, and that this study has potential use for learners and facilitators of PBL and other discussion based learning. The research aims and objectives will now be stated, leading into chapter three, detailing the methodology for this PhD study.

2.8 Research questions

It is essential that the research methodology is fitted to the research question – and not the reverse (Silverman 2007). The primary research aim of this study is to establish and evaluate evidence of differences in speech behaviours by individual learners during one cycle of PBL discussion. Through linking speech to the individual, it becomes possible to explore differences through quantitative and interpretive corpus analysis and illustration and to link behaviours with results on entry and exit to the first year.

The process of investigating individual behaviour and preference around learning does not lend itself easily to hypothesis creation. It is grounded in a particular time

and place which is confined to studying one particular population acting in a particular way. By definition, this is an observational study, which may have relevance elsewhere with similar populations in similar contexts but is necessarily limited in its generalisability. In their guidelines for Strengthening The Reporting of Observational Studies in Epidemiology (STROBE) (von Elm 2008) the authors criticise the lack of systematic reporting for observational studies even in formal biomedical research. The STROBE guidelines have been used in evaluating the strength of reporting the outcomes of medical education studies as well as informing the reporting of this research.

Such a study would normally preclude quantitative methodology, however, through creating a speech corpus, the corpus itself becomes evidence which can be interrogated and can provide objective quantitative evidence where otherwise self-reporting, observation or interviewing may be the only means. The speech behaviours can be measured quantitatively and then statistically examined to identify any significant correlations with pertinent variables, such as exam and personality test results. The following table summarises the key research questions which will be addressed through the study. The research questions are stated as null hypotheses where the statistical analysis has the potential to demonstrate them. Where corpus analysis has the potential to demonstrate rich and diverse difference, this will be stated as a general prediction.

Table 13 Research Questions posed

Research Questions related to statistically significant differences
Gender differences
Is there a difference in the amount of speech contributed by males or females?
Is there a difference in the number of questions asked by males or females?
Can corpus analysis demonstrate significant differences between the male and female corpora in line with previous studies of gender difference in speech?
Performance and group differences
Is there a statistically significant relationship identified between different PBL groups and PPD or MCQ measures for this cohort?
Is there a connection between interview grade and PPD score at the end of the first year?
Is there a relationship between the interview grade and performance in GAMSAT pre-entry?
Is there a connection between speaking more words overall and the score students received on interview, in the GAMSAT; in their MCQ exam; in their PPD?
Is there a connection between the Flesch simplicity index measure when applied to speech, use of questions and performance in knowledge tests?
Is there a difference in the frequency or type of questions asked or terminology used by higher and lower scoring learners?
Is there a connection between personality type and performance in knowledge tests in year 1?
Personality type, learning styles and speech behaviour
Is there a correlation between extraversion and the number of words, or number of utterances (turns taken) which a person makes during a group session?
Can corpus analysis demonstrate significant differences in student speech between personality and learning style corpora created according to the FFM, temperament type and Kolb learning style inventory?

The following chapter on research methods will discuss how the research questions will be addressed and the reasoning behind the choices.

2.8.1 Statement of the research aim and objectives

This study will therefore attempt to address the following research aim:

An objective study exploring the differences in individual speech when learning through discussion in PBL

2.8.2 The main objectives of this study are to:

- Create a PBL speech corpus of at least 100,000 words, which identifies each individual student's contribution to a PBL discussion of the same case at the same time within six, gender-balanced PBL groups, (45 first year students).
- Use this data to determine to what extent speech behaviours can be shown to be closely associated with factors which may influence learning.
- Conduct multiple linear regression analysis in order to identify statistically significant connection (or lack of connection) between an individual's comparable speech behaviours (the proportion of total student speech, proportion of total questions, proportion of total turns taken, proportion of an individual's sentences which are questions, a measure of language complexity), and gender, individual personality and learning style types and entry and first year knowledge, interview and professional development scores.
- Create sub-corpora, identified by these factors associated with individual learners: gender, personality type, learning style, high and low scorers in knowledge and Personal and Professional Development (PPD) tests.
- Analyse the PBL speech corpus and sub-corpora to identify significant differences in word and technical term frequency, parts of speech, particularly use of questioning, and semantic categories relevant to anatomy, disease medicine and cause and effect.
- Use this data to inform PBL facilitators and learners about the implications of individual differences to learning through PBL discussions.

- Propose some possible steps facilitators and learners can take to better support individual learning in the PBL group situation.
- Consider the wider implications for all teaching through discussion, in the medical curriculum such as in ethics, professionalism and medical humanities and also in the arts and social sciences curriculum where discussion is the pre-eminent teaching methodology.

3 Methods

3.1 Introduction

This chapter will introduce the research paradigm adopted for this study by considering the background of the researcher, giving an overview of the theoretical frameworks within which the research is conducted; positivism and post-positivism and the current qualitative and quantitative approaches and the issues which define the quality and generalisability of possible methods which may be employed and interpreted within such frameworks. There will be a brief consideration of the research strategies which pertain to some degree, ethnographic, phenomenological, case study and general grounded theory.

The research study will then be located as a mixed-methods observational study, combining quantitative methods to collect data about learners and learner speech behaviour and triangulate such findings through corpus analysis and to interpret corpus data within a qualitative approach within a grounded theoretical framework. Both quantitative analysis of word frequencies, parts of speech and semantic categories, will be employed, but individual cases will also be used qualitatively to explore and illustrate typical instances in order to address how individuals learn through discussion.

Each variable will be presented to establish how and why it was selected, how it will be measured, and interpreted and presented in the statistical results section and known limitations will be reviewed.

There will be a discussion of issues relating to the creation of this spoken word corpora, and the basis for analysis and illustration which will be presented in the corpus results chapter and known limitations will be reviewed.

3.1.1 Introduction to the research paradigm for this study

3.1.1.1 My own background in the arts and computer sciences

My first and higher degrees, in English Literature and Creative Writing, both concern themselves with the description and depiction of individuals through their behaviour, primarily through reporting their thoughts and conversations. When a modern author writes from a character's viewpoint, they seek to represent the world as experienced by that person, sharing what they alone would notice and ignoring what would not concern them. In my own creative writing, I consciously choose how to relate a story through weaving different characters' viewpoints, internal thoughts, dialogues and conversations and description together. This choice determines the view of events which is shared directly with the reader and what is left unsaid, for a reader to discern meaning in the way we are likely to do in real life.

In addition, throughout my thirty-year career (so far) as a teacher, mentor, mediator and manager, I have been aware of being strongly reliant on interpreting an individual's speech behaviour to guide my interventions relating to learning and performance in tutorial, small group teaching, management, appraisal and assessment settings. Which of the many possible interventions I could choose, would be guided by my interpretation of what might be the most helpful for that person at that moment, in the light of Vygotsky's notion of the Zone of Proximal Development (Vygotskiĭ 1978).

Early in my career, I worked as a programmer and systems analyst and used very precise forms of language to instruct inanimate machines, through the reduction of complex business systems down to computer coding. I have become acutely aware of the limitations for generalisability of much of social science, particularly for so complex a phenomenon as learning, through lecturing on learning theory to a variety of scientific academics. I am also keenly aware that much of value cannot be easily understood through quantitative study alone.

In summary, I was keen to combine each of these long-term interests into one project; to collect data and create a suitable body of evidence of speech behaviour from a cohort of students: the first year Derby GEM students, 2011. The study would use systematic, quantitative methods to gather data, then employ statistics to explore relationships between data points, some of which were based on exploring models suggested by previous research examined in the literature review, but also informed by the research process itself. The main focus of the study will be to use corpus analysis to examine the speech corpus of individual medical students engaged in collaborative problem based learning in relation to individual factors. It would be necessary to identify the contribution of each individual. It would be helpful to minimise variability through capturing data from learners in one cohort, at the same time, engaged on the same learning task.

3.1.2 Locating the study within the research paradigm

3.1.2.1 Positivism and post-positivism

The enlightenment and the establishment of the scientific method demonstrated magnificent results in the fields of industry and engineering in the eighteenth and early nineteenth centuries. However, the second half of the nineteenth century gave rise to debates about how a new approach of systematic, positivistic enquiry, other than deducing laws, may be required, particularly for understanding human behaviour (Creswell 2009). As industrialisation brought people together en masse, and as political theory grappled with understanding human behaviour in order to harness it for industry, anti-positivists attacked "science's mechanistic and reductionist view of nature, which, by definition, defines life in measurable terms rather than inner experience, and excludes notions of choice, freedom, individuality, and moral responsibility" (Cohen 2008) p. 17. Establishing individual facts could also hide the complexity in systems. Meaning and understanding may also incorporate multiple viewpoints, personal experiences, or multiple truths as opposed to one simple law, labelled as 'the truth' (Myers 2000).

Medicine may sometimes be classified as an 'interpretive' science, such as geology, as it has some laws or certainties, but accepts that at the heart of the human experience of health or illness, birth and death, there are many forces at work which are best understood, given our current knowledge by being 'evidence-based' rather than totally evidenced. Increasingly, health science research extends a scientifically rigorous approach to incorporate qualitative, interpretive, narrative evidence relating to the individual, in order to achieve its ends (Pope, Mays et al. 2006). Mixed methods research, combining both approaches, is increasingly common, as one is used to inform the other and a wider scope of data results in better informing of health policy (O'Cathain and Thomas 2006).

3.1.3 Qualitative research

"The most basic way of characterizing qualitative studies is to describe their aims as seeking answers to questions about the 'what' 'how' or 'why' of a phenomenon, rather than questions about 'how many' or 'how much'." (Green and Thorogood 2009) *p. 5.*

The research question needs to guide the choice of methods. As stated in the quotation above, qualitative research does not often involve numerical data, and does not usually begin with a hypothesis. Results and conclusions are based around theories and frameworks, with an inductive approach employed, which allows ideas and issues to emerge from the subjects examined. Breadth of study is often sacrificed for detail (Creswell 2009). Qualitative research is not just non-quantitative research: it has its own methods and disciplines. Increasingly mixed methodologies are emerging as offering a blend of precision and rigour whilst achieving depth and quality of understanding of the human experience which numbers seldom provide alone.

However, the subjectivity at the heart of qualitative research is a crucial issue. Research which relies fully on the perspective of the researcher alone, is limited by that one viewpoint. Inevitably what the researcher elects to notice or record says just as much about the researcher as about the researched. Even quantitative

methods are subject to such bias, as will be discussed in the section discussing the production of the speech corpus through transcription.

3.1.4 Generalisability of qualitative research

Transferring qualitative findings from one context to another requires closer scrutiny. This is of particular concern where the focus is on a group of individuals, such as PBL learners in Derby, UK. Just because this group behaves in one way, or has a set of perceptions about themselves, doesn't mean that any another group will feel the same way. However, if the study is not conceived as one which will 'prove' a hypothesis, rather one which will shed light on one and possibly similar or comparable situations, then this is an addition to the knowledge base. Medical education, and education in general has cause to remember that research often has more in common with a 'case study' approach where generalisation from a case study should be done with caution. Reproducibility, such a crucial test of quantitative scientific endeavour, proves more elusive in examining complex human behaviours, such as learning. However, it might be considered possible if the research is described in depth. In addition, studies of specialised groups of individuals may fruitfully be compared to similar specialised groups, rather than from a set of individuals to a population (Stake 1980). The nature of statistics itself, apparently quantitative as it is using numerical values, relies completely on assumptions about a general population, although with certain checks and balances, which will be discussed later in this methodological chapter. The generalisability of the findings of this study will need to be discussed and interpreted with caution.

3.1.5 Ensuring quality in qualitative research

Qualitative research should not be equated to a lack of rigour and should be subjected to quality assessment like any other research – it demands "theoretical sophistication and methodological rigour." (Silverman 2007) p. 209. For quantitative research there is one gold standard of reproducibility, "single tangible reality that an investigation is intended to unearth and display" (Lincoln & Guba,1985,p.294). For

qualitative research, issues of credibility, transferability, dependability and confirmability are key to validating the quality of findings.

Eight different validation strategies are described by Creswell of which he states, at least two should be applied to ensure quality of research (Creswell 2007). These are:

- Prolonged engagement and participation in the field
- Triangulation of different sources of evidence
- Peer review or debriefing, to externally check the process
- Negative case analysis, in which hypotheses are revised until all cases fit
- Clarification of researcher bias and prior assumptions
- Member checking, in which findings are fed back to the population under study
- Rich and thick description
- An external audit by an independent reviewer

Each of these eight points will now be considered in relation to how each factor contributes to ensure the quality of this research.

For this research study, the researcher has been engaged in medical education over a decade, and in higher education for over two decades.

The study rests on triangulating observational data, such as measuring the proportion of speech or questions made by an individual, with self-reported data, such as from a personality and learning style questionnaire. In addition, the corpus itself is a rich source of evidence which may triangulate the quantitative evidence. For example, the corpus itself may show evidence that volatile (neurotic) students use fewer technical terms related to medicine in their discussion, which may help to account for their lower exam performance, as identified through statistical analysis.

This research process has been peer reviewed through taking advice from highly experienced academics in the fields of medical education and corpus analysis, from three medical statisticians, two former PhD students who have published in closely allied fields and through presenting at five conferences. The transcriptions for the PBL discussions have been shared and compared between two PhD students, who have cross referenced transcripts to ensure accuracy.

There were few firm hypotheses, so negative case analysis was not appropriate, which Creswell describes as a process in which hypotheses are revised until all known cases fit the hypotheses (Creswell 2007).

The researcher has endeavoured to clarify how the study is located within their prior assumptions and previous experience, but of course, as a solo project, with interpretive elements, there will always be bias and assumptions which are integral to the work.

The pilot study, which will be described shortly, was undertaken with the full participation of one PBL group and their facilitator, and the resulting transcriptions, personality and learning styles data were all provided to the participants. A debriefing session was held where implications for their future study and PBL discussions were reviewed, particularly gender differences in communication styles.

The corpus analysis technique provides rich data which will be described and interpreted with reference to the context, identity and individuality of the speakers. This is a powerful demonstration of its potential for providing rich and thick accounts of phenomena, in keeping with the original definition of the term 'thick description' as commonly employed in qualitative research (Ponterotto 2006).

The PhD examination process completes this external audit by two independent reviewing examiners.

3.1.6 The research strategy

3.1.6.1 Ethnography, phenomenology, case studies and grounded theory

Within the qualitative domain, there are several prominent strategies of inquiry (Creswell 2009) which can be used according to the research question, resources available, and researchers intentions and experience. In brief, ethnography, phenomenology, case study and grounded theory were considered for this study as the research questions were developed in depth.

Ethnography is an approach used by anthropologists, who by living amongst the culture they are trying to study, observe the patterns of behaviour and try and understand what is happening. However, it is increasingly appearing as a methodology chosen by social scientists and educationalists. It appeals from the perspective of validity – although the researcher must take care that their presence does not influence the behaviour of their subjects. Ethnography has been used extensively within the medical context – for example Fox used ethnography to study the behaviour of surgeons (Fox 1992). One of the earliest considerations of medicine's hidden curriculum is described in Becker's classic ethnography "Boys in White" (Becker, Geer et al. 1961). Although a fascinating way of conducting research, this was not deemed an appropriate strategy for this study, due to the position of the researcher in medical education in the Medical School in Nottingham and being an assessor of clinical educators and not a PBL facilitator. It was also established during the literature review and in preparatory discussion with PBL facilitators at Derby, GEMS, that this approach had already been successfully and recently employed in the Derby cohort in the PhD thesis of one of the PBL facilitators, entitled, 'PBL but not as we know it: an ethnography of the practice and facilitation of 'problem-based learning' within a hybrid graduate-entry medical programme in England' (Jennings 2013).

Phenomenological studies use a prolonged period of study of a particular phenomenon in order to understand and interpret it, in a similar fashion to ethnography. It uses a more perceptive approach and is less naturalistic, but is

often discussed as assisting in the removal of "taken for granted" assumptions by the researcher (Green and Thorogood 2009). The phenomenon usually takes one aspect of human experience, often an emotion such as grief or anxiety as its focus (Moustakas 1994). Speech behaviour during PBL is therefore far too broad a topic to be labelled such a phenomenon. It was hoped that some of the findings might be indicative of learning behaviour during other purposive learning discussions, not simply PBL at Derby and so adopting a phenomenological approach would not provide a broad enough answer. In addition, the researcher in this study would struggle to remove assumptions at the level required due to many years of experience facilitating and teaching through small group discussion methods.

A case study looks at a particular incidence of something, or at a group performing a particular function – and from this can then go on to discuss inferences and possibly generalise about a situation. Case studies can be defined as a qualitative methodology per se (Yin 1994), but the word is confusingly also used simply to define a specific area under study (Stake 1980). The researcher may be participating in what they are studying, or just observing. One fellow PhD student, has based a study on selected cases from the same data set as da Silva and one from this study in order to explore how students and facilitators collaborate in PBL discussions to construct knowledge or new understanding (Tokode 2016).

This study is a case study of sorts, as it focussed on one institution's practice: transcribing PBL discussions for one half of a cohort during the same cycle of PBL. This 'case', does provide opportunities to examine the data in depth, but it is one of the limitations of the study that it was not feasible, practically or financially to record more cases, i.e. more full cycles of PBL by whole cohorts across the first and second year. For example, da Silva (Da Silva and Dennick 2010, Da Silva 2014) videoed and transcribed first and second year PBL discussions on different topics. It was mooted in the early stages that this study could collect data for the same students in several PBL groupings over two years in order to identify the degree of stability for individual behaviour in differing group settings. This proved impractical and financially

unfeasible and would have yielded a corpus dataset approaching half a million words: a feat which has only been achieved in major, funded, specialist research settings. As it would require an infinite number of cases to 'prove' human behaviour in any generalisable sense, this study compares the speech behaviour of individuals across the same PBL case, in six separate PBL groups (cases) over a number of sessions, with speech behaviour samples totalling over 100,000 words.

The grounded theory technique, first described by Glaser and Straus (Glaser and Strauss 1967) is a way of inducing social phenomena. The theory is developed from the ground up, that is, the researcher should have no preconception or hypothesis to prove or disprove. They merely have an area of interest they wish to theorise about, and develop this theory in parallel with the data they collect. This means that the data collected informs the next questions asked, as the theory continually reshapes the next stage of the research. This is called an iterative process – and the theoretical sampling means that generalisations from the data are more justifiable.

Grounded Theory as a methodology can use any number of data collection techniques – and indeed use other methodologies within it, such as case studies. It therefore has a broad and flexible appeal – any form of data can be used to create a grounded theory. It is a useful way of informing policy and practice (Charmaz 2009).

It was felt that developing a grounded theory would be an appropriate research strategy for emerging research questions particularly in considering the interpretation of evidence from the corpus. It would not be without difficulties, but overall it seemed to offer an in-depth perspective which would then be convertible to conclusions about the nature of discussion within PBL groups in a representative cohort which opportunity presented for sampling.

The analysis of the speech corpus would also be carried out qualitatively in that patterns once identified, could be interpreted. Some research questions, suggested by prior research have been posed but the methods also maintain openness to

emerging hypotheses as the analysis progressed. The analysis would be approached as a single case study, without need for generalisation.

3.1.7 Correlational, cross-sectional and longitudinal studies

The study was clearly non-experimental, as the variables were outside the control of the researcher. As an observational study, it combined elements of each of the main three types: correlational studies, cross-sectional studies and longitudinal studies (Tavakol and Sandars 2014) quoted below.

"In correlational studies, researchers make a claim about the relationship between variables in theories or models. For example, what is the relationship between student ability and their score on the UK Clinical Aptitude Test (UKCAT)? By calculating the correlation between students' marks and UKCAT scores, researchers can address the association between student ability and UKCAT."

However, this study also employs cross-sectional approaches as it identifies independent and dependent variables in a population, such as the proportion of student speech and the gender of the speaker and gathers data during a short period of time, (Levin 2006) and then describes the association between them.

Finally, there are also elements of a longitudinal approach, (Goldstein 1968) as it was planned for data to be gathered retrospectively for measures of performance on entry, for behaviour to be measured close to the end of the first academic year, then to gather performance measures from the close of year one.

3.1.8 Corpus Analysis Methodology

Corpus analysis is now regularly used to examine complex patterns of speech, often within literary works (Adolphs 2006). Corpus analysis has been used on self-documenting corpora such as online discussions and advice boards to discern patterns in vernacular speech (Harvey, Brown et al. 2007). This study differed in that it sought to create a corpus of spoken language being generated in a very specific way at a specific time by a particular group of people and then to compare individual difference within this group.

3.1.9 Building the Corpus

3.1.9.1 Pilot Study 1: feasibility of attribution of speech to individuals from existing video and audio recording

Previous PhD research had established a method for creating a transcript which could then be used to create a spoken corpus, which could then be analysed. An approach had been pioneered with a previous intake of students at Derby GEMS and data for one PBL group, completing one PBL cycle had been analysed and published in Medical Education (Da Silva and Dennick 2010). The researcher re-used the da Silva-Dennick corpus for the first pilot study, to determine the feasibility of allocating speech to individuals from a group discussion. The transcript for one cycle of PBL involving seven participating students and one facilitator, was analysed to identify individual speaker's contributions by watching the video and listening to the audio recording and marking a printed transcript with each speaker's individual contributions. Using both video and audio sources, this proved entirely feasible, with a very low proportion of utterances which were not possible to ascribe to an individual speaker, usually through issues related to over talking and inaudibility, which are discussed below. The transcript had been produced by an experienced professional transcription service with experience in medical terminology. This transcript was 'cleaned' by adding more accurate versions of the medical terminology where possible. The accuracy of the transcript was high, with well over 95% of the speech rendered accurately. However, certain issues emerged which were particularly relevant to fast-paced discussions between up to eight or nine people.

3.1.9.2 Issues of cross talk and inaudibility in corpus production

Where sections proved inaudible, due to over-talking by several people at once for example, these were clearly marked in the transcript using the recommended notation [inaudible] or [over-talking]. This tended to occur from time to time in all groups, and meant that the transcriber was not able to fully describe all the speech behaviour. However, overall, inaudible or cross talk sections accounted for less than

5% of the total corpus, within an acceptable level of error for any measurement process.

In discussing the methods used to create the British National Corpus (BNC), Crowdy describes a similar level of pragmatic transcription and error checking (Crowdy 1994). Transcribers were instructed not to spend much time discerning inaudible speech, but to time the gap and mark [inaudible] in square brackets. Where possible over talking was to be shown by bracketing the words said by each person in the speaker's separate line. These issues were specifically checked and amended by a different person, often with specialist subject or dialect knowledge.

In this case, the transcription was independently checked and compared by two researchers, one with a medical training, and medical terminology was clarified. This suggests that the current methods were effective in reaching acceptable levels of accuracy in transcriptions for corpus analysis, as they mirrored, and in some senses exceeded that used by the largest comparator corpus.

Clearly, there is still margin for considerable error in transcribing speech, which does not exist when creating written corpora. There are clearly issues relating to which speaker is heard most clearly in a group. Being heard may relate to literal audibility – where someone chooses to sit away from the recording device, but also to the power of the voice itself, its pitch and volume, and also to the perceived power, confidence and assertiveness of the speaker. These issues were discussed in the literature review.

This pilot process established that it was feasible to identify the individual contributions of a speaker using video and audio to a reasonable degree of accuracy. The results of this initial pilot were presented in a poster at the Ottawa conference in Miami in 2010 (Haines 2010). This poster is included as Appendix G. In this PBL group, the student who spoke the fewest words appeared in no way to be deemed to be a 'social loafer', judging by the responses of the group viewed on the video recording. Whenever the quietest person did make a contribution to discussion, the rest of the group were immediately silent and listened intently, signifying, that the

contribution was greatly respected, when it did occur and often prompted deeper engagement with the subject matter by the group as a whole.

3.1.10 Pilot study 2 – Feasibility of ascribing speech to individuals using only audio recording, personality and learning styles questionnaires

Data collection was planned using one further trial group of PBL students. Students volunteered, together with their facilitator, in order to establish the feasibility of the methodology and to give a first-hand account to other facilitators and students who would be working in the larger scale data collection project. The consent process established that this pilot group was only willing to be audio recorded, and were not willing for their performance to be video recorded. It was not possible to discern the reason for this reluctance or to allay fears and so the pilot proceeded on the basis of needing to establish the feasibility of identifying an individual's contribution to discussion based on an audio transcript alone.

The technical aspects of the data collection worked well and the same transcription service which had been used for da Silva and Dennick's research was used to provide a comparable level of accuracy in the unattributed corpus. The next stage, where the researcher listened to the recording with the printed transcript a sufficient number of times in order to identify which person had contributed to the discussion, was extremely difficult, and therefore valuable to planning the scaled up data-collection phase. It simply proved much more difficult for the researcher to identify the different contributions by participants without video evidence to back up identification. Several of the voices were of a similar tone, and had the same gender and regional characteristics. Even though participants had provided a very brief introduction by stating their names at the start of the recording, these very short clips of their usual individual speaking voices were insufficient for identifying their later contributions during rapid discussions. It became clear that to scale up the data collection it would be necessary to record much fuller samples of an individual speaking in a range of circumstances. This might more closely and fully replicate

the type of speech of a lively PBL session where a group of up to eight people and a facilitator may speak swiftly about complex technical information. This type of transcription is also at the limit of what even an experienced transcriber may be able to produce before inaccuracy or unreportable speech begins to climb to unacceptable levels as a proportion of the transcription, particularly due to issues of inaudibility and cross talk as discussed above.

After extensive listening, it proved possible to identify, with reasonable confidence, over 90% of the utterances made and ascribe them to the individuals and to validate, feedback and discuss the findings with this voluntary group, in accordance with Creswell's recommended eight features of high quality qualitative research discussed above. The results for participation by each individual and the facilitator, as a proportion of the speech, were reported at this lower degree of accuracy. The group and facilitator reported valuing the feedback they received. Discussion centred on the findings that the males in the group spoke across the females, and contributed a larger proportion of discussion overall. There was a wide variation of contribution by individuals, varying from 743 to 7,000 words with the facilitator contributing 4,000 words. The findings were reported in a presentation at the 13th Ottawa conference (Haines 2012).

As a result of the discussion of the results with the PBL group, the facilitator chose to change his practice and to recommend that other facilitators did also. Hitherto, he had often taken on the role of the patient in any role-play where the students role-played the history taking doctor. The reasoning had been that this gave another opportunity to facilitate the learner experience such that the students gained full value from the activity. This group agreed that it would be a better stratagem for a relatively quiet student to take on that role in order to build confidence in spontaneous participation. The group felt this change did not preclude facilitation, where necessary, but was likely to enhance the structure supporting their participation in discussion. This experience also had an impact on the design of the larger scale study.

3.1.11 Building the corpus: main study, recruitment of participants and sample size

In the induction week, September 2011, the researcher and Professor Dennick, the Professor of Medical Education, were invited to address the whole cohort (94 students) as part of their induction week. Students were told about the role of medical educational research in the medical curriculum and heard about the publication in Medical Education of a previous study which videoed and audio recorded PBL discussions (Da Silva and Dennick 2010). The process of consent was explained and a question and answer session ensured that the students had a detailed understanding of the implications of their consent, and the ways in which the data could be used in the future, in its anonymised form. Consent forms were given out to each student, and were collected up on that day with a small number being collected within the same week. A total of 89 forms were completed and collected from a total of 94 students. A sample consent form is included in Appendix C, along with the ethics committee permission for the study (Appendix A, Appendix B).

One student returned their consent form and clearly indicated that they definitely withheld consent for any group in which they were involved, to take part in the study. Even though it had been explained that the focus of the main study was solely on student speech, not facilitator speech, several facilitators were also unwilling to take part. After discussions during facilitators' meetings, two facilitators withheld their consent, ruling out 16 students from their two PBL groups. Three new facilitators had recently been appointed and the date of data collection was set to take account of their need to establish working relationships with their groups and familiarity with the PBL curriculum as a whole, in order for them to willingly participate in typical sessions. One facilitator agreed on the condition that their words were redacted, but agreed that the number of words could be counted and the proportion of questions could also be counted.

On the day of recording, further discussions were held privately within some groups which led to further PBL groups withdrawing from the study. As each group can be ruled out on the basis of one person withholding consent, on the first day of data collection, six groups, from the original total of 12, remained within the study, having consented only to audio recording.

3.1.12 Data Collection recording

The researcher had provided a brief training session about using the handheld digital recording device in order to support facilitators who had been asked to use them to record their pbl sessions. Spare batteries, spare devices and two researchers were available outside the PBL discussion rooms throughout the three-day cycle. The researcher produced a handout which reminded participants of the nature of the project and provided a log for facilitators to record any comments about the sessions, such as who was present or absent; whether the session was 'typical' for that group, and any extra notes. There were instructions and suggested text for each participant to record to give full examples of their normal speaking voices which could then be used to help identify their contribution within the transcription.

The learning styles questionnaires were also handed out in a pack for the facilitator and group to complete during the first session. The personality test was briefly explained again, and the students received an online invitation by email, to complete this during the course of the PBL cycle. A blank example of the Kolb Learning Styles questionnaire (Appendix H) and a completed summary of the Quintax online survey (Appendix D) are included.

Groups were responsible for recording their own sessions on the Tuesday, Friday, Tuesday sessions which completed the selected PBL cycle. The researcher was available throughout to assist with technical support issues and to collect the KLSI questionnaires.

3.1.13 Timing of the PBL cycle for data collection

A number of factors were taken into consideration before selecting the cycle of PBL which was to be recorded. The pilot groups had taken place at different points in the year. Pilot one, had used video recordings and transcriptions prepared by da Silva in 2008/9. The researcher had conducted the second pilot with audio-only recording and verification of results with participants, in May 2011. The research project had been introduced at Induction in September 2011.

For the main data collection, where more groups were participating, the PBL cycle was chosen in order to ensure that, as far as possible, the pbl discussions recorded would represent the most 'typical' sample of PBL discussion for first year students. In discussion with facilitators, and based on experience of teaching over many years, it was determined that the selected cycle must fulfil the following conditions:

- The cycle recording should only take place when first year students have had sufficient experience working in different PBL groups to be developing a 'typical' style of learning, using the PBL scenario. Students are randomly allocated to PBL groups twice during their first year, and so this should be during their second allocation.
- The cycle should not precede or follow any substantial phase of assessment to avoid the chance of 'contamination' of learning attitude becoming over-focussed on assessment rather than learning.
- The cycle should only take place when a PBL group and facilitator have had the chance to already complete at least 2 cycles, and thus have established a working style together, so that the 'sample' is representative of the way of working for that group
- The cycle should only take place when a new facilitator has had the chance to facilitate at least ten cycles in order to establish their confidence with the way PBL might be expected to function within the curriculum for those students at that stage (first year); again to ensure a representative sample of group behaviours

- The cycle should concern a 'typical' PBL scenario which represents those undertaken for such students at such a stage.
- The cycle should be when the researcher was available to give technical support and to brief facilitators and students sufficiently, to remind them of the issues and to discuss their participation and consent at any time during the data collection.

Relatively few PBL cycles could fulfil these conditions within the first year. The cycle concerning cardio-vascular and heart failure at the close of January 2012 was selected. This is usually described as a challenging PBL scenario by participants and facilitators, yet was most likely to fulfil the criteria listed above.

3.1.14 The limitations of the data collection methods used for building the corpus: **Sample size**

Clearly the study would have been enhanced by being able to recruit every PBL group from the year cohort. Initially, on the basis of active opting out by only one student, this looked to be within reach. However, as might be expected, in the event, other students proved to be less willing to take part when the data collection became a reality for their group on that day. The nature of the consent for this study was always conceived of as one which allowed participants to opt out at any time prior to the data being completely collected, without giving any reason.

The interdependence inevitably affected the sample size achieved. In other words, students and facilitators were aware that their decision to opt out of the study meant that the entire group would be unable to participate. Perhaps this worked in favour of the study where individuals had minor concerns, as they might be willing to take part in order that their fellow PBL group members could have the experience. This might be a positive factor when a PBL group was having a positive or relatively successful experience of working together. However, where a group was experiencing difficulties, perhaps managing personality conflicts within the groups, or for any reason felt less willing for their speech to be transcribed and held

up to scrutiny, they may be more likely to opt out. This might mean that the participating sample were atypically effective in their behaviours, and that this study will not be able to represent or analyse more dysfunctional speech behaviour. It might be possible to see if the personality types and learning styles represented are representative of a population as a whole, or by comparing to similar norms, but this can never be known for this study.

3.1.15 Transcription accuracy: comparable studies

The level of accuracy achieved using similar methodology for a group as a whole and using video and audio transcription had been established as being capable of leading to publishable work as a whole (Da Silva and Dennick 2010). The researcher's first pilot had demonstrated that it was possible to ascribe utterances to individuals using the same video and audio material. However, the second pilot study had demonstrated that ascribing individual contribution based on audio recordings alone is a more ambitious task, and one for which it is not easy to reach the very highest levels of accuracy. As discussed in the literature review, each creator of spoken corpora, particularly for group discussions such as PBL, mention the challenge and time consuming nature of this task.

For the fully scaled up study, the usual experienced transcriber was not willing to transcribe the discussions recorded by these groups for this data collection as the task was beyond her capabilities and willingness. It seems that one factor was that there was a problem with the basic audibility of some of the recordings. There was a range of different data recorders used, and as they were managed within the groups themselves, there was some variation in how well the devices were positioned with respect to the seating arrangements of the groups. In effect, some individuals sat at some distance from the data recorder, on the margins of the group and this reduced the actual detail of the audio recording for them. Some groups proved to have very complex, fast moving discussions which meant that the original transcriber felt that an unacceptably high proportion of conversation was not available for her to transcribe and so would label [inaudible] at very many points in

the transcription. The researcher discussed this at some length and in detail with the usual transcriber, but she withdrew from working on the transcriptions.

The researcher submitted the recordings to an educational technologist to be audio-enhanced using Audacity (Mazzoni 2016), a free, open source software package which allows manipulation of audio files. In this case, the overall volume level was able to be boosted in the recordings so that even the quietest contributor could be brought up to a level of audibility. A new transcriber was sought, who had experience of more complex, sustained group discussions. Ubiqus were approached as they had experience in transcribing events for clients such as the World Bank, the World Health Organisation (WHO) and the Royal Courts of Justice (Ubiqus 2017). A trial transcription was completed to a high level of accuracy. It was even possible for this transcribing company to allocate each utterance to an individual speaker in most cases. A team of transcribers was used, then transcriptions were checked and edited by an experienced lead transcriber. This led to a higher level of consistency and accuracy for producing a corpus of individually allocated speech than had previously been possible.

However, in the case of one group, the team of transcribers still experienced difficulty in differentiating between two individuals, and submitted a transcription where almost 50% of the utterances were unallocated. However, the researcher re-worked this transcription and achieved an acceptable level of accuracy by spending more time repeating the audio than would be possible on a commercial basis. All transcriptions have been checked for accuracy against the original recording by the researcher, and have achieved an acceptable level of accuracy which is comparable to that achieved by da Silva and Dennick, and in comparison with the working methods for the BNC, as discussed during the literature review. The proportion of unattributed, or inaudible speech remained well below the 5% error margin which may begin to cause concern for missing data. Indeed, (Dong and Peng 2013) confidently assert that missing data in psychological and educational research is the rule rather than the exception, quoting a variety of meta analyses arguing that it is

common for 15 – 20% of quantitative data to be missing even in published studies. They offer advice on principled approaches to ensure that such studies may fairly represent the data they do gather.

3.1.16 Cleaning transcriptions: issues and decisions

Many transcription corpus studies apply a policy of 'cleaning' the corpus. Da Silva applied a policy of removing any discussion which was deemed 'irrelevant' to that study by the researcher. This might include talk at the start or close of a PBL session which related to social activity or course deadlines, anything not germane to the discussion of that particular PBL scenario. In this study, the researcher has taken the decision that all discussion between participants which has been successfully recorded and transcribed is of relevance to determining the discussion and speech behaviour of individuals. Some groups did 'self-edit' for this. On occasion, the transcription included such comments as 'have we finished with the PBL now, shall we switch off the recorder?' for example. Other groups take a couple of minutes to settle to their task, or break off part way through an intense and factual discussion to share comments about other experiences. This represented a marginal proportion of activity in each group, but the researcher believes that such discussion contributes to an understanding of the learning style and personality type of an individual, and reflects the way people behave when they are working in a group. Consequently, the researcher has not cleaned such utterances from the transcriptions in this study.

However, in monitoring the transcription standard of the professional transcribers, the researcher has noticed two marked tendencies which influence the results. The transcribers themselves have made less stringent effort to transcribe discussion which is not closely related to factual discussion, particularly when there has been the added difficulty of cross-talk or potential inaudibility. In other words, in sections of transcription where people were engaged more in process or socially motivated talk, this may be less fully represented in the transcription, and therefore in the

corpus. The impact of this tendency was mitigated by the researcher's efforts in checking all recordings and transcriptions.

The second tendency is one which I will discuss in more detail in the discussion of results. In many cases, where more than one person is speaking at once, the strongest voice is the one which gets transcribed. In very many cases, that tends to be the male voice, both because it may be deeper and stronger and therefore more audible, but also, the statement tends to be framed in a more definite manner. This is well supported by findings in the literature, that there are marked gender differences in the willingness of an individual to make a strong claim in agreement or disagreement of something. Overall, women will tend to make less definite statements, particularly in mixed gender discussions, and will tend to frame their contributions as suggestions or questions, rather than as statements (Tannen 2001). Thus, the corpus is skewed away from giving a full representation of quantity of contributions by female participants when they were 'competing' with male speakers. This makes it particularly important to examine gender differences overall, and to look closely at examples of these patterns when they occur.

As the transcription was restricted to audio, there was no possibility of analysing non-verbal behaviour of PBL participants. The transcript was restricted to identifying the speaker, by adopting the convention of inserting a code for each new speaker in brackets and beginning a new line for each new speaker. Speakers were anonymised and allocated a code which identified their PBL group, and their position in a master list. Thus, <S8.2> was a speaker from PBL group 8, who was positioned as the second member in the master list.

3.1.17 Sample size of PBL Speech

The main challenge to the corpus creation was the differing sizes in samples of speech associated with PBL groups, as opposed to individuals. In other words, for 50% of the sample, 3 PBL groups, there was a full sized sample of their full 3 PBL sessions. For two further PBL groups, the sample covered two PBL sessions, while one group only contributed one session. This occurred for a variety of reasons. In

two cases the facilitator was absent through illness and so one facilitator effectively ran two PBL groups, and chose not to activate the recording in either PBL grouping for that session. On another occasion there was a technical problem which prevented the digital device from working during the session, although on inspection it appeared to be functioning adequately and a spare device would have been available outside the room. In spite of a reasonable degree of planning to adapt to predictable incidents, the data sample size was reduced.

One medical statistician suggested introducing a variable to identify each PBL group which could be used to monitor whether that factor was found to be statistically significant when it came to analysis. Another medical statistician had no issue with any missing data, but was concerned that the number of subjects available to the data set was less than one thousand. In response to advice and further reading, the researcher decided that in terms of the main focus of the study, this was technically a sampling issue. In other words, all speech was being sampled, and larger sample sizes were available for some PBL groups than others.

The situation could be resolved to a satisfactory extent by identifying and using measures which allowed comparison across differently sized samples. In any event, spoken corpora were to be created to sample speech classified according to certain other factors which may be expected to occur frequently, for example, male or female or introvert or extravert. So long as there were sufficient sample sizes for these sub-divisions there would be triangulating data available through corpus analysis. When the PBL corpus was to be split into two, male/female, extravert/introvert for example, each may be expected to contain around 50,000 words. In some cases, for example, KLSI four sub divisions were possible. In some cases, the high and lowest scoring quartiles according to certain tests were created. For both these categories it might be expected that the sample should contain around 25,000 words and that analysis would be limited and noted as such if the available data varied from this markedly.

The PBL corpus as a whole is considered to be small to medium, for a spoken word corpus. Although it would have been excellent to have created a full corpus for every PBL group, for 3 sessions, this was still a substantial sample. It represents eleven separate PBL sessions conducted by eight to nine people in each case. The overall length of discussion sampled when measured in time is approximately twenty hours. Da Silva and Dennick based their article on one PBL group's three session PBL cycle. One recent EdD study, presented at the AMEE Conference, 2013 and submitted as a thesis by Dr Jamieson rested on the transcription of two hours of PBL discussion for six groups, studied in detail (Jamieson 2013), (Jamieson 2014). The study of transcripts of RCGP simulated consultation OSCE examinations (Roberts 2014), rested on 40 case transcripts, each lasting under ten minutes.

3.1.18 Corpus Analysis tool: Wmatrix3

Wmatrix3, licenced for use from the University of Lancaster, The School of Computing and Communications, was produced by Dr Paul Rayson, Reader in Natural Language Processing and Director of University Centre for Computer Corpus Research on Language (UCREL). This research centre claims 177 publications in this area, particularly in 'noisy environments' such as speech. Wmatrix3 is a powerful web and cloud based tool which has been used at Nottingham to conduct corpus analysis of PBL speech. Da Silva used Wmatrix2, the previous version, to conduct analysis of PBL transcripts, leading to publication in Medical Education and a PhD thesis on clinical reasoning. The tool (Rayson 2003) allows for powerful analysis of frequency, parts of speech and semantic use, showing keyness, through key words in context, calculating the log likelihood of words occurring by chance (Rayson 2008).

3.1.19 Method of corpus analysis: Preparation of the PBL corpus and sub corpora

Each corpus is created as a simple text file. In order to differentiate between different speakers, a speaker identifier can be added, and enclosed in <> to ensure that the corpus analysis software ignores the content of the brackets. First the PBL

corpus was created which contained each of the 45 subjects' speech. Next, this corpus was subdivided to create the corpora listed below containing only speech contributed by speakers of that subgroup.

All transcriptions were combined into one PBL corpus, in text format, to permit analysis of the responses of all students using the corpus analysis tool, Wmatrix3. Sub-corpora were also created, to permit comparison according to factors which had been identified as being of interest and would permit triangulation with other quantitative measures. Thus, the individual contributions of each speaker were separated, then recombined to contribute to the relevant sub corpus. The individual's contributions were in chronological order across the PBL sessions to which they contributed, but the interaction between individual students could not be preserved.

3.1.20 Relative corpus sizes

The sizes are quoted in order to demonstrate that the sample size is large enough and representative enough to warrant corpus analysis. The size of the sample size for top and lower quartile varied between 10 or 11 subjects. Also, the size of the sub-corpora varied, in part, due to the sample size contributed by each participant within the qualifying group. For example, there could have been a low number of high GAMSAT scorers who contributed a full three pbl session sample size. This means that we cannot draw conclusions directly from the size of the corpora, or the number of question marks or number of words or sentences. As a consequence, statistical analysis focussed on the percentage contribution by each learner.

3.1.21 Corpus Analyses

For each corpus, a basic analysis was carried out in order to establish the overall frequency of word occurrence (often popularly shown as relatively sized word clouds); by parts of speech (the grammatical category or purpose of speech); by semantic category (the topic or subject to which words belonged). The results were reported only for the strongest levels of significance, measured by log likelihood of

occurrence and effect size, measured by Bayes. These terms are discussed at the start of the corpus analysis results chapter.

3.1.22 Comparison between corpora

The crucial choice when conducting corpus analysis is against which other corpus to conduct a meaningful comparison. For a speech corpus conducted on medical technical terminology in a post-graduate educational setting there are very few suitable comparators. The US Michigan Corpus of Academic Spoken English (MICASE) consists of 152 transcripts, totalling 1,8 million words. However, when closely inspected, there was only one highly interactive study group discussion in biology and biochemistry, at a lower level of study, that contributed 40,000 words which might have made it a suitable comparator. In the event, the entire PBL learner corpus was first analysed and described, and then compared to the British National Corpus, an educational spoken corpus to establish a baseline.

However, the true purpose of the study is to explore differences between different learners within the same situation: PBL post graduate medicine first year students. As a consequence of this, the main comparator for this study is against each other, i.e.: the other sub-corpora. This does limit the overall size of the log likelihood measures. In comparison with a general spoken corpus, biological terms would appear extremely significantly over-included in the PBL corpus. However, when compared to each other, statistically significant measures were still frequently reported, even for the core medical topics of the PBL scenario. It is this rich territory which is reported in the corpus results chapter.

For each of the core analyses, one corpus is compared to another, for each of the three key processes and differences are reported in order of the significance, measured by log likelihood. It is possible to exclude certain instances from the report where the frequency is very minor to permit the exclusion of unhelpful differences. For example, some terms may occur very infrequently throughout the corpus. However, where there was a large difference in frequency between sub-corpora, that might be of note for the study. For example, one term might occur

only twice in one corpus yet 45 times in the other. This magnitude of difference, whilst not reaching large overall totals, is strong evidence for difference in speech behaviour and has been included in the analysis. Similarly, where there are very high numbers of occurrences of common terms such as 'the'; across all corpora only the most significantly differing cases are reported.

3.1.23 Further analyses

In many cases, question words were explored separately, often under the CS or subordinating conjunction category, as they had been found to differ across PBL discussions and might be associated with clinical reasoning (Da Silva and Dennick 2010). RRQ, the parts of speech category for 'general adverbs beginning with wh' category was also used to compare questioning approaches. Where there was significant difference between use of the key semantic medical categories, B1 Anatomy and physiology, B2, Health and disease and B3 medicines and medical treatment, these were compared.

3.1.24 Format of the results: key words in context, keyness

For each analysis Wmatrix3 reports a table of results. The results chapter will present selected evidence in the form of the tabulated results together with illustrative concordances. As there were very many significant results for each analysis, only the strongest and most relevant to the research questions will be reported in detail.

After the initial 'descriptive analysis' of the overall corpus, the sub-corpora were systematically compared. Wmatrix3 reports on the 'keyness' the frequency with which a term or part of speech or semantic category appears in one corpus in comparison to the other. For example, when comparing the sub-corpora created by males or females, it is necessary to select one to be the first corpus, against which the other is compared. The results then describe which words, or parts of speech or semantic categories are over-used or under-used in the first corpus compared to the other. In each report the results can be looked at from the viewpoint of each corpus, and be positive (over-occurring) or negative (under-occurring). In other

words, the results will first report what over-occurred in the first corpus, then state how this means that it occurs significantly less frequently for the other corpus. Next the results will consider what under-occurred (was less frequent) in the first corpus, before reporting how this feature would be more frequent in the other corpus. By this means, the results will provide a succinct picture of the differences between the features of both corpora through two tables and permit more detailed illustration through concordances of the most salient evidence.

3.1.25 Concordances

Wmatrix3 also allows all instances of a category which is subjected to analysis, to be identified and pulled out through the concordance feature. Again, this is possible in separate lists according to over or under use. This feature allows the example to be seen in context with the other utterances by the same learner and by other learners with the same characteristic of interest.

3.1.26 Discussion of corpus Results

Some possible implications are discussed as trends emerge through the data, before being fully discussed, together with the statistical analysis in the final discussion chapter.

3.1.27 Quantitative Data Points

The data points will now be presented.

There were 44 measurements made in all, for each of the 45 students who participated in the study. These were stored on an anonymised master Excel spreadsheet, of which a summary version is included in the appendix. These data points included demographic details, related to gender and previous degree. Next came measurements related to personality type, including each of the big five dichotomous measures, and KLSI, for each of the possible four sub-categories. In addition, scores on entry for each candidate, ratings in the entry interview, (A, B or C) and the GAMSAT score on entry were included. Next, the test scores from the close of the first year for the MCQ knowledge test, and for the PPD combined

portfolio were recorded. For each of the PBL sessions available, measurements of speech for each individual, within their own PBL group, were collected. Averages for each measure for each PBL group were calculated for comparison. The key comparator measures were:

- the proportion of total student speech per person for the overall sample (based on the number of words, expressed as a percentage of the total for their PBL group, for student speech)
- the proportion of turns taken to speak, per person for the overall sample (based on the number of turns taken to speak, expressed as a percentage of the total number of turns taken to speak for their PBL group for student speech)
- the average length of turn, per person for the overall sample (based on the number of sentences per person, divided by the number of turns to give an average length of utterance)
- the proportion of sentences which were questions, per person for the overall sample (based on the number of question marks allocated by the team of transcribers in response to hearing the vocal tone rising to denote a question through grammatical interpretation during transcription.)
- the total averaged readability score as supplied by the software within word, based on the Flesch readability score, which divides the average number of syllables per word and words per sentence according to a formula, given below and expresses the result on a 100 point scale, where the higher the score, the easier the reading level, or lower the complexity of utterance. Microsoft Word recommend aiming for between 60 and 70 for a general document, in order to ensure ease of readability.

  The formula for the Flesch Reading Ease score is:

  206.835 – (1.015 x ASL) – (84.6 x ASW)

  where:

ASL = average sentence length (the number of words divided by the number of sentences)

ASW = average number of syllables per word (the number of syllables divided by the number of words)

3.1.28 Summary of data points collected

There were fourteen comparable data points which were considered for inferential analysis using the enter method of multiple logistical regression in minimal model combinations so that they might identify potential statistical correlation within the acceptable bounds of analysis for independent and dependent variables, which will be discussed in the statistical results chapter. An overview of the key variables in SPSS is provided in Appendix F.

3.2 Quintax Questionnaires: personality typing

It might be argued that the behaviour of students with respect to completing an online personality typing questionnaire provides further evidence of their personality type! Fortunately, for this study, the completion rate for the participants has been of a high order with only one individual failing to complete both the personality and KLSI questionnaire, although some responded only after several reminders. This type of compliance with administrative request has recently been discussed by the General Medical Council as an aspect of professionalism and evidence for conscientiousness, one of the FFM measures! See the discussion and the literature review for further discussion of these issues, (McLachlan, Finn et al. 2009) (McLachlan 2010).

An acceptable typological framework for describing individual differences was sought. The most obvious approach would be to employ a widely validated personality style measure which identified key differences between learners, especially those which have been associated with differences in speaking, and particularly speaking in groups. For example, surgency, (whether one was extravert or introvert) or the level of neuroticism have been associated with differing amounts of speech, as one of their key defining behaviour differences. It was envisaged that a learning style measure, collected in addition to an individual personality type measure for each participant, would give opportunities to consider correlation. Learning styles are claimed as a sub-type of personality, but of interest to specific considerations of learning and personality. (Jackson and Lawty-Jones 1996).

3.2.1 Selection of personality typing questionnaire

There are a number of widely used personality type measures, as discussed in the literature review. However, the measure selected for this study needed to fulfil the following criteria.

Necessary requirements

- Demonstrating acceptable levels of validity and reliability
- Demonstrably based on research literature

Desirable features

- Modestly priced per individual, as up to 100 students were targeted
- Easily administered, preferably online
- Willing to work with the researcher to provide rigorous information suitable for publication
- Preferably with relevant norm groups for comparison with the sample population

The market-leader for personality testing within the business environment, Myers-Briggs, as discussed in the literature review, fails to meet several of the above criteria (Barbuto 1997). In particular, it is costly, not particularly keen to cooperate with the academic community, and criticised for its lack of adoption of modern statistical methods. It is closely based on the typological basis which is not normed against a standard working population.

For some years, the Medical School had a working relationship with providers of an individual personality measure called Quintax. For at least 5 years, the creators of this type measure had been contributing a day towards a module on the University's MMedSci which the researcher convenes. The module on Continuing Personal and Professional Development (CPPD), considers the topic of Psychometrics; i.e.: the measurement of mind in a selection and development context. The Quintax type measure was included in a comparison against the identified criteria above, and found to be the most suitable option for this purpose.

3.2.2 Discussion of the features and limitations of the Quintax method for personality typing

The general limitations of the self-administered self-report questionnaire as a basis to report personality and learning style have been discussed in the literature review.

This section will present how the Quintax questionnaire was administered in this research before summarising the salient features of the performance of the questionnaire which justified its suitability for this research in terms of the reported scores for reliability and validity.

Each of the participants in each of the 6 PBL groups was sent a personal email invitation to complete the Quintax 72 item questionnaire online and there is a summary chart which they received in feedback in the appendix for illustration. The completion time was less than 20 minutes. Essentially this questionnaire was used under conditions of self-administration. This might not be appropriate when being used in a selection environment, but for conditions of research among an existing cohort of intelligent individuals who have been briefed on the method, it was decided it was suitable for self-administration. The researcher is a trained administrator of the Quintax questionnaire, and has attended a two-day course which covers the detailed questions concerning the design, scoring and feeding back of the Quintax instrument. It is also possible to administer the questionnaire in a paper format, with hand scoring, but this option was not used.

For this research study, upon completion of their online questionnaire, a participant received an immediate, automated report detailing their personality type, giving examples of where this style might prove a strength and highlighting where equally they may benefit from development. No individual discussion of feedback was attempted. No individual validation of results or cross-referring with other sources was done. Participants received an immediate report, for which the basic cost was £25 per individual, where it might cost them more than a £100 to receive a comparable report from a private company source. This was a particularly valuable feature, as those who participated received immediate feedback, which may have encouraged those who had delayed, through the mechanism of peer group pressure. It also fulfilled one of the obligations of the researcher to offer useful development advice to participants in return for their support. As there were multiple technical

problems with producing the transcriptions, this was especially valuable, as it was not possible to feedback any information on group speech behaviour in a timely fashion, as had originally been hoped. Three students out of the 45, never completed a personality questionnaire, in spite of several reminders, and their missing data for Quintax is shown in the summarising spreadsheet in the appendix.

3.2.3 Quintax methodology

Quintax is a personality questionnaire which allows for a profile interpretation of a respondent's traits, along with the potential to place the individual in one of a number of normatively generated types. The name derives from Quin – Latin for five, to show it is closely based on the FFM analysis which has gained widespread acceptance in the literature of personality theory, deriving from the work of Costa and McCrae (Costa and McCrae 1992). 'Tax' is designed to show a link with a 'taxonomy, or a system of types to which people may be assigned, depending on their responses to a questionnaire. It has been in development since 1995 and in formal use since January 1998. The British Psychological Society (BPS) reviewed it in 2003, published in 2008, and awarded a rating of reasonable to overall reliability and particularly good construct validity, and good for all other categories, such as quality of documentation, materials, and the use of normative reference groups, according to the European Federation of Psychologists' Associations (EFPA) Test Review version 3.42.

3.2.3.1 ***Brief Description of the five scales used in Quintax***

Table 14 *Table describing Quintax scales, in preferred order for Quintax*

Quintax scales	Brief description of Type	Brief description of Type
Extraversion I or i E or e	Introverted type is: retiring sober reserved	Extravert type is: sociable fun-loving affectionate
Criticality: Personable v Logical P or L p or l	Personable type is: Motivated by harmony courteous lenient	Logical type is: Motivated by facts willing to disagree critical
Organisation: Adaptable v structured A or S a or s	Adapted type is: flexible responsive motivated by novelty	Structured type is: impatient with unforeseen planner motivated by completion
Intellectual Focus Grounded v Theoretical G or T g or t	Grounded type is: here and now focus conventional practical, factual	Theoretical type is: original creative independent
Emotional Involvement Calm v Volatile -C or -V -c or –v	Calm type is: not easily upset by events or strong emotions maintains control over feelings usually resilient to everyday stresses	Volatile type is: shows effects of pressure emotional involvement let's others know of their feelings

The Quintax Questionnaire reframes the scales derived from McCrae and Costa's well-established work, keeping the same scales, yet expresses them in a less value-laden way. In addition, the scales are used so that they can show strength of preference; i.e.: a score on each scale will be determined in relation to the normed population as residing at towards one pole of a scale or another. This allows the scale to take account of the strength of preference rather than taking an all or nothing approach. When reporting an individual's position on the continuum, Quintax has a simple notation system; a capital letter indicates a strong preference, a lower case letter, a more moderate preference. In addition, the neuroticism, stability scale is used to underlie the entire type description. In other words, each type is either –v, or -V, towards the volatile end of the emotionally reactive spectrum, or –c or -C, towards the Calm end of the same spectrum. This measure is also considered, by Quintax, as more indicative of an individual's permanent state, rather than a trait which may demonstrate more variation according to context and be subject to more control by an individual.

3.2.4 Potential outcomes of the Quintax measures

Mathematically it can be deduced that there are 32 possible types, each of which can be held strongly or moderate to weakly, as denoted by the use of lower case letters. This leads to a potentially large number of types being potentially recorded for the 45 students of this study, if the full potential for Quintax reporting was to be used.

To limit complexity, it was decided not to use the strength of measure difference, so both i and I, were both considered as 'introvert' for the purposes of this study. Obviously, this choice also reduces the richness of data which may have been a benefit, but was simply not practicable for this study. Another option exists for considering the broader groupings of Quintax, which enables it to be used as a development tool in workshops and with the usual smaller number of employees

who may be considered within the workplace; the setting for which Quintax was designed

3.2.5 How the Quintax scales are measured

A 72 item questionnaire builds a profile of scores, with twelve questions allocated for each of the five dimensions, and one further set of twelve questions building a score on a social desirability scale, discussed below.

3.2.5.1 Reliability and Validity

One of the key factors in selecting this scale as opposed to MBTI, concerned the reliability and validity of the scales, this inevitably involves some technical discussion including an account of the development of the instrument in order to reach the form which was used in this study.

3.2.5.2 Self-Reporting issues and social desirability scoring

Questionnaires which seek to establish a measure of an individual's behaviour, as discussed in the literature review, must always be limited by the condition that they are self-reporting. In other words, an individual responds in the light of their insight into their own behaviour and their desire to be truthful about it to those putting the question. In this study, as was painstakingly explained to participants, there was no advantage or disadvantage that would accrue to them as a result of this personality measure being used. No one type is deemed any better than any other, and it definitely had no bearing on their progress within their medical studies. It was hoped that this would encourage participants to be willing to self-report, free from concern about manipulating their responses.

The questionnaire does make use of a 'Social Desirability' measure, derived from Eysenck's subtly named lie scale (Eysenck and Eysenck 1963). This is an independently reported measure which is not fed back to a candidate, but where a high score is generated, it might signal heightened self-deception, or a desire to present oneself well in the light of others in a consistent way (Nederhof 1985).

There are twelve questions from the battery of 72 which are concerned with this behaviour.

3.2.5.3 Reliability: item response by factor

Tables of reliability have been calculated to demonstrate the item response reliability for each item of the twelve questions for each factor against the normative group: 681 employed adults between the ages of 18 and 60, approximately equally distributed between the genders and representing professional and other types of occupation, who have completed the Quintax test in the past. Cronbach's alpha has been calculated for each factor and ranges from 0.66 to 0.86 have been reported. In other words, using a similar methodology to that employed in the assessment of Multiple Choice Questionnaires (MCQs) for medical students, each question within a set is analysed with respect to the scores in the overall set in order to establish how reliable it is in determining strong scores in that set. This process should identify questions which are not performing effectively. Cronbach's alpha score indicates an acceptable level of internal consistency for each of the scales. The designers note that their intention was to have broad and relevant measures rather than to seek scale 'purity' above other things.

3.2.5.4 Test-retest Reliabilities

It was possible to test and then retest a subset of people against the scales for 55 retail managers at an interval of between 7 and 10 weeks. This produced results of between 0.66 and 0.86 using Cronbach's alpha which compared very closely to the scores reported for each scale based on the norm group taking the test only once. This suggests that the test is suitably stable in its ability to produce the same results from the same people over a time interval.

3.2.5.5 Validity

The scales may well be proved reliable, but it is also necessary to consider their validity; broadly how well the scales are measuring what they claim to be measuring, and how relevant that is to the purposes of the questionnaire.

There are three broad fields of validity to consider: content, criterion-related (predictive) and construct validity.

3.2.5.6 Content Validity

This relates to the appropriateness for the inclusion of an item for testing. Content validity would be high when important content is asked about with appropriate frequency and depth, reflecting its importance in the 'real world'. Quintax allocates 12 questions to each of the scales it purports to measure, and has calculated Cronbach's alpha, which sometimes indicates redundancy, (i.e.: content is subject to repeated questions leading to little useful data) where a score above 0.8 is reported. Thus, the design of the questionnaire is designed to deliver content validity.

3.2.5.7 Construct Validity

Construct validity endeavours to measure the extent to which an item is accurately or fully measured. Where a construct is particularly 'abstract' or debatable, one of the best methods is to compare scales of measurement with each other to determine the degree of agreement there is in measuring what they consider to be the same construct. For personality questionnaires, this would be agreement between measures of the same traits. Statistical techniques, such as factor analysis, can also be used to determine how the underlying fields affect the way named constructs are being measured. Quintax is derived very closely from the FFM measures and has been compared with the other more established measures of personality and has been shown to have reasonable to good reliability and validity, and to perform well in all comparisons.

It may be that, given the challenges of measuring personality, that this is as good as it currently gets and is certainly suitable for verifying the choice of personality measure for the purpose of this study, where issues of cost and accessibility were also crucial.

3.2.5.8 Predictive validity

Overall, we might use a personality questionnaire and be hopeful that it will predict future behaviour, thus a personality questionnaire which seeks to measure an individual's typical behaviour in the workplace, might suggest future behaviour in a workplace too. This predictive validity would be the equivalent of a 'holy grail' for employers hoping to use such a method for selection purposes. However, personality tests, are not commonly believed to be able to demonstrate a strong predictive validity. This might be termed the criterion related validity and is a hotly contested field in the literature, see for example,(Schmit, Ryan et al. 1995). However, broadly measured traits cannot reasonably be thought to correlate with very specific outcomes, or indeed, each other. Quintax advises users to use their scales to help them identify what areas of performance are important criteria in successfully performing the particular jobs which they are designing or seeking to fill. The Quintax scales can then be used to identify who might be qualified within a desired performance profile, and equally who might be ruled out as unsuitable for a certain type of role. Quintax has carried out some of this analysis, identifying correlations between their scales and factors reported using a 360 degree rating within one company for example (Wilkie, Robertson et al. 2006).

3.2.5.9 Specific Construct validation studies

Where Quintax has reported results which have impressed the BPS in their review, is in their data which compares results achieved using their measure as against those produced by other comparable scales. In other words, in the world of self-reporting questionnaires, this set of scales shows good validity in comparison to other well-established measures. A satisfactory comparable relation has been demonstrated between Quintax results and those from a range of comparable personality typing scales, such as ORPHEUS, Ravens, NEO Five Factor Inventory, 16PF5, and ASE's business Personality Indicator. It is worth noting that MBTI (Myers–Briggs) also demonstrates a significant correlation with Quintax, in spite of the division between types being determined for MBTI without using normative or statistical data. Also, although there is overlap between the most dimensions of MBTI with the big five,

there is only a weak overlap between the MBTI dimension of thinking/feeling with the scale for criticality as measured by Quintax.

In summary, the Quintax online questionnaire was adopted for the study, but due to restrictions in the sample size, students would be measured according to the overall broad dichotomous values, such as Introvert or Extravert, rather than make use of the potential subtlety of strength of score.

3.3 Kolb Learning Styles Inventory (LSI)

3.3.1 Discussion of the features and limitations of the Kolb method for Learning Styles

Each of the participants in each of the 6 PBL groups was given a paper copy of the Kolb (version 3.1) and invited to complete the 12 item questionnaire during their PBL session, summarise their score and then to return it to the researcher. A blank copy is presented in Appendix H. The completion time was less than 20 minutes. Essentially this questionnaire was used under conditions of self-administration. For conditions of research among an existing cohort of intelligent individuals who have been briefed on the method, it was decided it was suitable for self-administration.

For this research study, upon completion of the questionnaire, a participant received a summary description detailing their learning style dominant type, giving examples of where this style might prove a strength and highlighting where equally they may benefit from development. No individual discussion of feedback was attempted. It also fulfilled one of the obligations of the researcher to offer useful development advice to participants in return for their support. As there were multiple technical problems with producing the transcriptions, this was especially valuable, as it was not possible to feedback any information on group behaviour in a timely fashion, as had originally been hoped. Eight students out of the 45 never completed a learning styles questionnaire, in spite of several reminders.

3.3.2 Kolb methodology

Kolb is a learning styles questionnaire which allows for a profile interpretation of a respondent's traits, along with the potential to place the individual in one of a number of normatively generated types. It has been in development and use since the early 1970s (Kolb 1971).

Mathematically it can be deduced that, once normed, there are 4 possible types, each of which can be held strongly or moderate to weakly. For the purposes of analysis, the students have been allocated into their dominant type; hence one of four categories.

3.3.3 How the scales are measured

A 12 item questionnaire builds a profile of scores, 3 questions allocated for each of the four dimensions. This questionnaire has been in use for over forty years and redundancy is claimed to be reduced to a minimum. There are however serious concerns which relate to all learning styles and Kolb is no exception. However, Kolb has been used in an evolving form over forty years, and thus has been used for many of the studies, particularly in the US which relate to comparable populations to the one under investigation in this study. The questionnaire was also provided free by the Hay group, with the permission of Alice Kolb, on condition that the research results were reported. The version used for this research is the Kolb, 3.1 (Kolb and Kolb 2005). A blank copy is provided in Appendix H.

3.3.4 Reliability and Validity

One of the key factors in selecting this scale, as opposed to Honey and Mumford's Learning style questionnaire, perhaps better known in the UK, and of slightly more recent origin (Honey 1982) concerned the reliability and validity of the scales, this inevitably involves some technical discussion including an account of the development of the instrument in order to reach the form which was used in this study.

3.3.5 Reliability, validity and KLSI

The Kolb's publish a full handbook to accompany the inventory which has been selected for use, including norms based on an updated sample of 6977 'more diverse and representative sample' of users and reporting internal and external reliability scores, (Kolb and Kolb 2005). The authors re-state the advantage for the KLSI, that, in their view, it is strongly based on the theoretical frameworks of the major educational scholars of the twentieth century. They also claim that they hold an

important position in being adopted widely throughout education, particularly higher education, as the LSI of choice. The authors take issue with the Honey and Mumford interpretation of the KLSI adapting it linguistically for the UK business market, as they claim that the clarity of the dimensions has been confused or lost and present cluster analysis to demonstrate this (Honey 1982). The authors acknowledge that the test-retest data for learning styles can be weak. Studies, as summarised in Coffield, (Coffield, Moseley et al. 2004) suggest an inconclusive picture, regularly reporting that half of students will change their learning styles from one measurement to the next. Cronbach's alpha does consistently report appropriate values, varying from 0.67 to 0.82 for this questionnaire. The authors account for this by suggesting that the subjects change in response to the different experience they have in between the tests. They also claim that as the key purpose of the KLSI is to profile the individual, rather than for comparative purposes with others, that this is best served by the responsive test which charts the preferences of the learner against the framework and themselves.

3.3.6 Assessment of Performance

As the test population are students at a GEMS they are subject to rigorous testing of their learning achievements, both as a condition of entry to the course, and also at the close of the first year: the year in which the PBL cycle was measured. This section discusses the nature of the performance measures which are available for these students. The following section closely follows the official guidance and description for the GAMSAT test and how it relates to the GEMS students in this study (GAMSAT 2017).

3.3.7 GAMSAT

The programmes provide opportunities for medical training to graduates who have made a mature decision to commit a lifetime to a demanding profession. By extending the selection process to graduates of the humanities and social sciences as well as the physical and biological sciences, students with a diverse range of

interests, skills and talents are represented. There is no preference or advanced standing offered to applicants with particular first degrees.

The programmes are open to graduates of any field who are able to meet the selection criteria.

3.3.7.1 Structure and Content

The purpose of GAMSAT is to assess an applicant's ability to understand and analyse material, to think critically about issues and, in the case of the Written Communication section, to organise and express thoughts in a logical and effective way

The test is divided into three sections designed to assess performance in the areas of:

I Reasoning in Humanities and Social Sciences

II Written Communication

III Reasoning in Biological and Physical Sciences – in the following proportion:

Chemistry 40%

Biology 40%

Physics 20%

3.3.7.2 The GAMSAT Scores

These sections are combined to give one score on a scale of 0 - 100, where section III has been weighted twice as heavily as I and II.

3.3.7.3 Minimum GAMSAT Scores

To be considered for interview for admission to the graduate-entry programme at Nottingham, candidates must achieve a minimum score of 55 in Section II, and 55 in either Section I or Section III with a minimum of 50 in the other section

3.3.8 Nottingham School of Graduate Entry Medicine & Health - GEMS

3.3.8.1 A brief summary of the Course

Graduate Entry Medicine (GEM) is a four-year, fast-track programme designed to broaden access to medical careers and attract graduates from a wide range of professional and educational backgrounds. Successful GEM students will graduate with Bachelor of Medicine, Bachelor of Surgery (BM BS) from the University of Nottingham.

There are two phases to the course:
• An 18-month foundation programme structured around problem-based learning (PBL).

The foundation programme is based at the purpose built Medical School on the Royal Derby Hospital site.

• 30 months of clinical training in healthcare facilities across the mid-Trent region joining students from the undergraduate course for this phase.

3.3.8.2 Problem Based Learning (PBL)

The PBL Course is divided into a number of clinical modules reflecting the core medical specialities and systems of the body: Structure, function and defence; Respiratory; Cardiovascular; Limbs and back; Alimentary; Endocrine; Urogenital; Neuroscience, and Integrative. Modules are then divided into 4-7 patient case studies introduced in PBL on a weekly basis. There are also two Personal & Professional Development (PPD) modules – one in year 1, one in year 2.

3.3.8.3 Progress Assessment

Formative assessments occur throughout the PBL Course and are taken as a progress indicator. In addition, throughout the foundation course, students are required to produce a Portfolio of Personal and Professional Development (PPD)

including drug reviews, patient case-studies from GP visits, ethics coursework and reflective assignments. Summative assessments including written papers, on-line multiple choice and clinical reasoning exams, and clinical skills assessment occur at the end of year 1 and at the end of the 18-month foundation programme.

3.3.8.4 Entry Requirements

The GEM course is open to graduates of any discipline who have obtained, or are predicted to obtain, a minimum of UK Bachelors (Honours) degree classified 2:2 or better. A higher postgraduate degree (MSc, MA, MPhil or PhD) will be accepted in lieu of a first degree if awarded by 1st August of the year of study. Degrees from overseas institutions are accepted if NARIC agree that it is comparable to a UK Bachelors (Honours) Degree at 1st or 2nd class Honours level.

1. Paid or voluntary work experience in a healthcare related setting is essential. Work experience needs to be sufficient to demonstrate two things

    a) that candidates have found out at first-hand what a doctor's work involves and are making a genuinely informed and realistic choice about a future career in medicine;

    b) commitment to the care of sick, disabled or disadvantaged people. The importance of this should not be underestimated – each year a number of applicants are rejected because of inadequate work experience.

2. Performance above cut-off in the GAMSAT aptitude test.
3. Interview attendance.

3.3.8.5 Interview Grades

The second key metric on entry is the grade allocated during multiple mini-interviews. Learners will be graded on responses to questions based around the following themes:

• Are you realistic and fully aware about what it means to be a doctor?

• Your interest in and commitment to health care.
• Whether your personal attributes are suited to the study and practice of medicine?
• Considered the emphasis on PBL and whether this suits
your approach to learning?"

3.3.8.6A note on the reliability of Interview Testing

It is not the concern of this study to calculate the reliability of entry scores which have been set by the University. However, it is worth noting that the reliability data for Multiple Mini Interviews is currently being studied and shows signs that it certainly is a more reliable form of predictive information that the traditional single interview, even when a panel is involved in grading applicants. In this case, there are at least seven interview raters who work independently of each other using a pre-determined categorised description. There are five grades, A, B, C, D, E. Only students achieving C or above continue in the selection process.

3.3.9 Assessments at the close of year 1

Students were comprehensively tested on their achievements at the close of the first year. These assessments were made in compliance with the University Quality Manual for procedures and so were subject to a professional level of rigour, specifically to assure the quality of their validity – chiefly, that they accurately measure relevant traits for developing medical students, and they do so reliability. For reliability, we would expect the same students to achieve a similar score were they to take the test again in a similar time period (test, re-test reliability), and that they have achieved a similar score to candidates of similar ability (item performance analysis).

Two assessments are reported for this study:

- Personal and Professional Development (PPD)
- Multiple Choice Questions (MCQ)

3.3.10 Personal and Professional Development Score

As discussed in the literature review, there are many valid aspects of the clinical role which relate to the psychomotor and affective domain; in other words, being an effective doctor is not just a matter of technical knowledge. The PPD assessment seeks to measure progress of candidates in these non-cognitive domains, and might be seen as a continuation of the interview grade. There is an increasing focus in society and therefore in the Medical Profession, on aspects of performance which go beyond merely certifying what a medical doctor must know in order to be competent in their job. Professionalism, or the ability to behave in a way which upholds the standards of the medical profession is increasingly being explicitly identified and even assessed from the earliest stages of the medical profession. The GMC sets out specific values and behaviours which they expect their members to display, and this forms part of the career-long revalidation process, incorporating reflective reports of incidents in practice. In addition, foundation level doctors are now assessed on their views of how they should behave in clinical situations through situational judgement tests which hinge on issues of judgement about how something needs to be done, as well as what, in a clinical context, should be done. The PPD assessment is an attempt to assess these aspects of a future clinician's performance (GMC 2016).

The PPD score given reflects an average of three assessed components:

- Written reflection in semester one where the candidate describes their personal reaction to a significant event in their course experience.
- Evidence-based practical project, which is to be related to the concept and use of Best Evidence Medical Practice
- Portfolio of evidence, including, a written reflection in response to information gathered through a 360 degrees appraisal, from two peers on the course and two members of staff. This closely parallels current practice in clinical professional appraisal, where such feedback is a component of a doctor's career-long revalidation.

3.3.11 Multiple Choice Questions

This is the most traditional form of assessment for knowledge during the initial stages of early medical training and focuses on clinical and related knowledge. Students were required to give 200 responses to multiple choice questions, many of which were in the single best answer format, where they are required to choose one from five. Extended matching questions account for a large proportion of the remainder, where students demonstrate their understanding by pairing items. Questions are designed to mitigate the effects of guessing without penalty points and are monitored for their ability to discriminate fairly between candidates. The results are adjusted and reported as a percentage score. This is the end of year measure which is most closely related to the GAMSAT entry score, as it rests on clinical knowledge and reasoning.

3.4 Statistical Methods

3.4.1 Data points

Data points relating to each student in the study, 45 in all, were compiled in a master spreadsheet. The data for 45 student cases was input into SPSS 22. As advised by the statistician, the data was explored according to model groups, which were also scatter plotted to explore where correlational relationships could be said to contribute. The results of these analyses are presented in the statistical results chapter.

3.4.2 Consideration of missing data

Data which related to student entry and first year progression were provided by the GEMS Office and were complete. Data which was collected especially for this study were not complete for each of the 45 students. Three students failed to provide the results of their Quintax personality survey. Eight students failed to provide the results of their Kolb Learning styles inventory. One student failed to provide either measure. Students were excluded from the corpus or statistical analysis for which the result was not known. Although this was disappointing, overall, this represented an acceptable level of missing data for Quintax measures, being around 5%, but weakened the KLSI, being around 14%.

Substituting data using standard bootstrapping options was explored, but after statistical advice from two statisticians, it was decided not to impute (estimate) any data, and to continue to treat the Kolb Learning styles inventory data, which had the most gaps, with caution, and as a triangulating adjunct to the personality type data. It was perfectly possible to continue the corpus analysis as planned, omitting the spoken contribution of those for whom the personality or learning style measure was unknown from the relevant corpus. This led to an acceptable reduction of sample size in analysis of the personality type corpora, a greater reduction in the sample size related to KLSI, but allowed speech to be included in gender and high or low scoring corpora, where these characteristics were known.

3.4.3 Special limitations related to multiple linear regression as a statistical method

This study was not straightforward in statistical terms, as it is essentially a corpus linguistically based study, which seeks triangulation of observations about speech behaviour with other factors which may be related to student success. The design of the study could never justify sufficient power to justify generalisability due to the limited number of participants, even had the whole cohort taken part. The population is already a specially selected one – GEMS students, from which a further sample has been derived. The study design also required the cooperation of the whole group of students and the facilitator over three sessions, which also reduced the likelihood of complete data being achievable.

The main method advised by the statistician to be used for exploration of multiple factors which may be associated with effectiveness, was multiple linear regression analysis. A correlational study is limited in its reliability and can never infer causality, however, it could allow the identification of measures of correlation and can suggest the relative strength which different data points may contribute to any significant differences. It would also be feasible to explore multiple linear regression analysis to identify any significant correlations between key data points to guide the selection of models for simple and multiple logistical regression, in view of the following restrictions.

3.4.4 Advice on restricted number of variables in linear regression

Peduzzi advises a conservative guideline of ensuring ten cases for every one variable which will be included in linear regression (Peduzzi, Concato et al. 1996). This condition is met for each variable, except for the KLSI diverger (7) and accommodating (6) types, for which data was available. More recent authors have suggested that this be relaxed to advise that the number of data points examined in linear regression is limited to no more than one per five subjects, which is exceeded in every case (Vittinghoff and McCulloch 2007). It is also recommended that the researcher considers the frequency of the least frequent item. In other words, a

variable with a limited frequency, such as the dichotomous variables, for gender and each of the FFM, or trichotomous, such as for the interview score, or four way, such as for the KLSI, would remain good candidates for linear analysis, with fewer cases than might be required for other types of data. For this study, that suggests that even with the inclusion of a variable such as the KLSI with low case incidence, robust analysis could be undertaken.

These data points were first examined using standard descriptive statistical methods to explore their frequency and interval data, to establish the degree to which they were normally distributed. These results are reported in Chapter 5.

Table 15 Dependent Variables: capable of being used in analyses

1. Interview score (1, 2 or 3)
2. GAMSAT %
3. PPD %
4. MCQ%
5. % group total student speech per person
6. % sentences which were questions
7. % total number of turns taken per student
8. Simplicity index

Table 16 Independent variables capable of being considered in analysis

1. Gender: male or female
2. Orientation: Introvert or Extravert
3. Volatility: Calm or volatile
4. Criticality: Personable or Logical
5. Ordering: Adaptable or structured
6. Grounded: Grounded or theoretical
7. Quintax Temperament type: Logical/Theoretical or Personable/Theoretical or Adaptable/Grounded or Structured/Grounded
8. Kolb LSI Assimilator/Converger/Accommodator/Diverger

3.4.5 Summary of the Methods Chapter

This chapter has introduced the research paradigm and located it within appropriate theoretical frameworks. It has considered current qualitative and quantitative approaches and the issues which define the quality and generalisability of possible methods which may be employed and interpreted within such frameworks.

The research study has been located as a mixed-methods observational study which will create a speech corpus and combine quantitative and qualitative methods in order to address individual differences in learning through PBL discussion. Corpus analysis will provide objective, quantitative analysis but individual cases will also be explored qualitatively to interpret and illustrate typical instances. Data will be interpreted and illustrated within a grounded theoretical framework (Savin-Baden

and Major 2013) but also within the nexus of educational and psychological theory which seeks to account for differences and similarities in human behaviour.

Each data point has been presented, with a justification for how and why it was selected, how it will be measured and interpreted, together with a consideration of some of the limitations of the methodology. Statistical results will be presented in Chapter 5.

The chapter has given an account of the creation of this spoken word corpus and the basis for the corpus analysis and illustration which will now be presented in Chapter 4.

4 Corpus Analysis Results

4.1 Introduction to the chapter

This chapter presents the results of the analyses of the spoken word corpora. First the chapter gives an overview of the rich quantity of results which were potentially available for reporting, and the organising principles for presentation and illustration. This section will introduce the rationale for reporting significance, taking both the significance of presence and absence into account, using the log likelihood and Bayes effect size measures.

The full spoken PBL corpus will be described in terms of frequency, question and conjunctive sentence use, parts of speech, and semantic categories before being compared to the closest comparator corpus to establish key word in context, (KWIC) comparative analysis. The sub-corpora will be introduced and the major significant analyses will be reported in three sections: for gender, then personality and learning types, and finally by academic performance. There will be additional illustration of key results through quoting sample concordances, related clusters of quotations by individuals. There will be a brief summary for each of the key corpora. There will be some indicative discussion of results in this setting, for ease of explanation and the main results will also be formally discussed in the discussion chapter.

4.1.1 Overview of the corpus analysis results

In almost every case when comparing sub corpora, there was an embarrassment of riches in terms of statistically significant results. This data may be termed both rich and thick, (Fusch 2015). In all, over twenty sub-corpora were created, and for each, an analysis was made comparing them to their relevant comparator in terms of word frequency, parts of speech and semantic category. In addition, conjunctive speech and question words and other salient features were analysed to inform discussion.

For this reason, a selection of analyses has been chosen for fuller illustration of the most compelling data.

4.1.2 Reminder of background issues in corpus analysis

This study was conceived in order to create a spoken word corpus to permit exploration of speech behaviour from a quantitative and qualitative perspective using corpus analysis techniques. This exploration also serves to triangulate the statistical findings observed. Whilst 45 subjects, although 50% of the available cohort, is a statistically limiting sample for inferential statistics; the spoken word corpus is categorised as small to medium and comprises over twenty hours of PBL discussion. For comparison, the educational section of the British Academic Spoken Corpus (BASE) currently contains transcriptions of 22 hours of seminars. This section is of a comparable size, but is likely to contain samples from fewer speakers overall, and to illustrate less naturalistic patterns of discussion, given the more formal nature of seminars (Adolphs and Knight 2010). As the sample of words is sufficient, it is possible to identify repeating patterns within the speech sample. As it was considered sufficient for the purposes of the British National Corpus, (BNC) spoken word section, to cover the entire demographic variety of the UK by sampling the speech of 124 speakers, a spoken word corpus of 45 speakers is a significant achievement (Burnard 2007). It is in line with current practice to suggest that patterns from 112,000 words of this corpus may reach similar levels of 'saturation' required for qualitative representation tests. In qualitative research, it is considered sufficient to stop sampling once no new examples of themes appear, where the data is said to have reached saturation (Fusch 2015).

The spoken word corpus transcribed 112,334 words of which student discussion accounted for 109,761 words, the difference being accounted for by the speech identifier tags for each utterance, for example <S2.1>. Each utterance was ascribed

between each of the 45 student speakers, within each of their six PBL groups over one cycle of PBL. This permitted the corpus as a whole to be examined, but also permits sub-corpora to be created to explore differences in speech behaviour which might be of significance in relation to categories such as gender, personality type and performance on entry and at the close of the first year of PBL study.

The corpus was prepared in accordance with guidelines from the compilers of the BNC, spoken corpus (Crowdy 1994) and guidance from the Wmatrix3 software, University of Lancaster, Linguistics (Rayson 2003). The files were prepared in a text file, *.txt. Each individual speaker was tagged with a unique identifying code; i.e. <S2.1> was a member of PBL group 2, and was designated speaker 1, and held in a master list which allowed for anonymising participants' data. The speaker identifier is enclosed within <> brackets to exclude them from the attentions of the corpus analysis tool.

4.1.3 Sample of an individual's corpus contribution

The following text is an extract giving a sample of spoken corpus text giving the contribution by one speaker, in chronological order for this speaker, but not showing the discussion of all participants and facilitator. This text illustrates many of the varied utterances observed within the corpus. This clearly demonstrates the variation in length of utterance, from one word, for example, 'yes', to the 64 word contribution read out from the PBL scenario where the patient is described: 'Agnes Mason is…..'. The extract also includes statements where the contents were inaudible, yet were made by that speaker, shown enclosed in square brackets []. There are examples of statements which were clearly identified as a question, signified by a question mark added by the transcriber, such as 'Turn what off?' There are also examples of statements which occur frequently in spoken discussions where sentences might have several elements but not be completed: 'Did we say anything else, or is…' There are also examples where we may infer that a student is also undertaking another role, here as scribe for the group, 'I am running out of room', 'you will not have to watch me' as they describe their actions. There are also

some comments which relate to the process: 'We put it as a question we must ask her', and which contribute humour to the working of the group: 'That's a fun age to give up'… The following, figure 10, gives an extract which illustrates the range of utterance which may be contributed by one speaker, although, of course, it does not show the interaction chronologically with the other speakers.

S11.7>	Which is why she is dehydrated.
<S11.7>	No, is that not good because that puts it back into her blood supply, which takes it away from the oedema and it makes her rehydrated.
<S11.7>	Turn what off?
<S11.7>	Do not think of the heart just as a pump.
<S11.7>	Yes.
<S11.7>	She worked in the mines.
<S11.7>	Did you ask how many pillows she sleeps on?
<S11.7>	You will not have to watch me when I write everything.
<S11.7>	Yes, peripheral vascular disease.
<S11.7>	Did we say anything else – or osteoporosis induced...
<S11.7>	That is a fun age to give it up, about 40.
<S11.7>	I am running out of room.
<S11.7>	Agnes Mason is an 85 year old lady who arrives in accident and emergency by ambulance. Her husband, aged 90, called for help as she had collapsed suddenly while doing the housework. By the time she arrived in A&E Mrs Mason is alert and oriented and not complaining of any chest discomfort. She is severely dyspnoeic and has bilateral pitting oedema to mid-calf.

Figure 10 continues…

<S11.7> Could there be muscle weakness, dizziness –

<S11.7> The heart is becoming less efficient.

<S11.7> She could be orthostatic.

<S11.7> We put it as a question we must ask her.

<S11.7> It might be because she is standing still doing some sort of housework. Perhaps she is doing the dishes; that causes pooling in her legs.

<S11.7> Is that not basically heart failure?

<S11.7> Could stenosis do it?

<S11.7> ...the aorta might be stenosed as well, so it might not be going out effectively.

<S11.7> I am ready when you are.

<S11.7> Ageing. In America there are 35 million people who are over 65 and this is going to double by 2030, so ageing is a problem. Age is a major risk factor for atherosclerosis, hypertension, heart attack, stroke, heart failure. It is more of a risk factor than diabetes, sedentary lifestyle and genetics, which is a bit scary.

Figure 11 Sample of corpus text: Speaker 7 from PBL group 11, a selection of contributions

4.1.4 From the master corpus, sub corpora were created of the following types and sizes:

The following table presents the corpora and sub-corpora which were created and their relative sizes.

Table 17 Summary of the Corpora and Sub corpora and their relative sizes

Corpus	Word Size of individual sub-corpora	Notes relating to excluded subjects
PBL corpus	109,761	Data known for all
Gender		
Male	68,165	Data known for all
Female	41,584	
FFM Personality Measure Corpora		Total word counts (approx. 100,000 words) shown for subjects where personality measure is known. Excluded where not known for 3 subjects
Introvert	48,154	
Extravert	54,332	
Logical	51,334	
Personable	46,018	
Adaptable	85,541	
Structured	19,119	
Grounded	26,297	
Theoretical	74,370	
Calm	58,154	
Volatile	43,072	

Personality Measure, by temperament type		Total word count shown for known subjects (3 unknown temperament types)
PT	30,536	
LT	38,943	
AG	23,321	
SG*	3,964	Very small sample, to be treated with caution
Kolb Learning Styles		Total word count (approximately 91,000) shown for known subjects (7 unknown learning style subjects)
Assimilating	31,589	
Convergent	23,334	
Accommodating	17,681	
Divergent	18,539	
Results-related Corpora	Highest/lowest quartile	Word count for total high and low quartiles, (approximately 50,000 words)
High GAMSAT	20,184	
Low GAMSAT	28,506	48,690
High post PBL: MCQ	23,406	
Low Post PBL: MCQ	30,187	53,593
High post PBL: PPD	27,291	
Low post PBL: PPD	21,597	48,888

*The SG temperament type (where this measure was known, for 42 of the 45 subjects) stands out as an exception.

The sub-corpora appear to include well-balanced sample sizes, even though the basis for selection was not to provide this, as it may be with many other corpus constructions. The high and low quartiles of performance, exclude the speech of the mid quartiles and so would be expected to approach 50% of the PBL corpus in total; which is in fact the case.

4.1.5 Basic Corpus Analysis

The corpus was examined as a whole to establish descriptive statistics before being compared to the most appropriate comparator spoken corpus, to identify likely areas of difference.

The following analyses were carried out:

- Frequency of basic nouns

  This allows us to examine the most frequently occurring nouns, and to what degree these are technical, as might be expected in such a corpus.

- Frequency of CS, subordinating conjunctions, (if, because etc.)

  This allows us to examine the proportion of speech which may be engaged in causative reasoning, as proposed by da Silva and Dennick (Da Silva and Dennick 2010).

- Analysis by parts of speech (POS)

  This allows us to triangulate findings in the other categories and identifies evidence of other speech behaviours, for example interjections, used in interruptions and pronoun use.

- Analysed by semantic (meaning) categories.

The sub corpora were then subjected to the same fundamental analyses in comparison with the appropriate matching corpus in order to identify significant differences in speech between individual students of that particular type.

4.1.6 The frequency results

Frequency of word occurrence has become popular through word clouds and can be a powerful tool. These depictions are based on relative word frequencies and are commonly calculated using simpler versions of corpus analysis software. For example, the entire 600,000 word content of the British Journal of General Practice was produced as a word cloud to serve as the cover of the journal. An article then discussed how the word cloud depicted differences between the reality of what had been published and the stated editorial policy (Atenstaedt 2012).

Word clouds depict precise statistical operations in an easily understood pictorial format, just as graphs can show precise and complex data in a readily understood form. It is possible to calculate the log likelihood for each occurrence of a word or phrase within a given text. This means that a cross-comparable index can be calculated to allow us to objectively demonstrate any statistically significant differences in occurrence. In the work of Silva and Dennick (Da Silva and Dennick 2010) the three phases of the PBL session were compared in terms of the frequency of technical nouns in order to explore the underlying PBL process using Wmatrix 2. Technical nouns were found to vary significantly between the different meetings. This study uses Wmatrix3 and replicates this approach by comparing the spoken corpora produced by different participants in the PBL process across their discussions to identify any significant differences.

4.1.7 Concordance and Collocation

These two features allow identification of how a word is situated within a text, when it occurs, or with what phrases it tends to be collocated, allowing a more qualitative analysis of texts. For example, da Silva and Dennick were able to identify indicative patterns of speech which may be signifiers of clinical reasoning, such as 'what if', 'because', etc. and were able to examine where they tended to occur in the process of PBL discussions.

The following table lists the nouns which were most frequently observed in the PBL corpus, together with their actual frequency, and the relative frequency, which is the proportion that that term occurs as normalised percentages. In other words, the relative frequency provides an index measure which allows us to compare occurrence between corpora of different sizes.

Table 18 Frequency of nouns in PBL corpus

Noun	Frequency	Relative Frequency
heart_failure	369	0.38
heart	338	0.35
blood	294	0.3
oedema	293	0.3
pressure	235	0.24
things	179	0.18
stenosis	171	0.18
water	156	0.16
thing	151	0.16
people	118	0.12
valve	112	0.12
regurgitation	108	0.11
volume	99	0.1
lungs	87	0.09
symptoms	87	0.09
blood_pressure	86	0.09
fluid	82	0.08
hypertension	82	0.08

It is evident that virtually all of these most frequently occurring nouns could relate to the specific matter of this PBL cycle, the operation of the heart and heart failure.

This suggests that the discussions recorded by the corpus remain close to the subject matter of this PBL cycle.

Table 19 Analysis of parts of speech (POS) in the PBL corpus

POS term	POS code	Frequency	Relative Freq
Noun (s)	NN1	12878	13.28
Adjective	JJ	6515	6.72
Article, the	AT	4885	5.04
General preposition	II	4132	4.26
Is	VBZ	3754	3.87
Coordinating conjunction, and, or	CC	2806	2.89

The analysis of POS for the PBL corpus is reported for the top five categories of parts of speech and for coordinating conjunctions. Nouns, singular and plural feature strongly, as do 'is' and 'the'. There are frequent uses of adjectives and general propositions. There is also high occurrence of coordinating conjunctions (CC), 'and' and 'or' which might suggest more complex statements requiring conjunctions, and also comparison between alternatives.

In addition, two categories are of great interest as they are most concerned with clinical reasoning; conjunction statements (CS) and question word occurrence and so are now presented for the PBL corpus.

Table 20 Frequency of conjunction use, CS Parts of Speech, (POS) category

Conjunction	Frequency	Relative Frequency
and	2241	2.31
that	866	0.89
but	665	0.69
if	623	0.64
because	549	0.57
or	492	0.51
so	372	0.38
when	169	0.17
as	134	0.14
cos	84	0.09
if	71	0.07

This table shows the frequency of use of representative words used as subordinating conjunctions (CS), many of which are associated with reasoning. High frequency is also to be expected with such a corpus, but which might vary between different types of learners, depending on their engagement with reasoning and their willingness to share this mental process out loud in a group. In association with the conjunction words, the next table gives the occurrence of specific question words within the corpus.

Table 21 Question word Use

Question	Frequency	Relative Frequency
what	729	0.75
which	466	0.48
how	191	0.20
why	176	0.18
when	121	0.12
where	109	0.11
who	97	0.1
how*	68	0.07
whatever	19	0.02
Total	**1976**	

A Microsoft Word search through the PBL corpus identified more questions than the total (1976) identified through question words: a total of 2,490 as identified with a question mark. Question marks had been added during transcription in response to a raised voice after any utterance. For example, an examination of the concordances for the occurrence of 'does' reveals that it occurs 340 times in connection with asking a question, for example, 'Does this mean it's heart failure?' or seeking confirmation, in the term 'doesn't it?' Other terms, such as, 'isn't it?' also occur in connection with question marks, and account for the increased number of question marks when compared to the use of the question words in the table above, provided by the Wmatrix3 categorisation of question words.

4.1.8 Semantic analysis

Semantic analysis requires comparison of each word with a dictionary which links the meaning of the words with a category. There are several key semantic categories, B1, B2 and B3, relating to aspects of body and disease which we would expect to feature highly in semantic analysis of the PBL corpus, as shown by the medical

content of the most frequent nouns shown above. There are other categories of interest such as categories related to aspects of individual personality, such as emotion, failure, agreement and semantic categories related to cause and effect etc.

Unfortunately, it is not always possible to match every instance with the correct semantic category. The Wmatrix3 system claims a 92% level of accuracy for this aspect of processing using the USAS tagging system (Rayson 2003).

Table 22 Overall semantic categorisation of the PBL spoken Corpus

Semantic Category	**Semantic Code**	**Fre**	**Rel Fre**
Grammatical	Z5	25217	26
Pronouns	Z8	15091	15.56
Existing	A3+	4801	4.95
Anatomy	**B1**	**3669**	**3.78**
Discourse	Z4	2313	2.39
Disease	**B2-**	**2246**	**2.32**
Unmatched	Z99	2158	2.23
Likely	**A7+**	**2029**	**2.09**
Medicines	**B3**	**563**	**0.58**

This table summarises the semantic categories and the frequency and relative frequency with which they occur within the PBL corpus. Anatomy (B1) and disease (B2) figure prominently, as do categories which may be associated with whether there are signs and symptoms, and what they mean, such as likelihood (A7+). Grammatical words, (Z5), such as 'a', 'the', etc. feature frequently, and pronouns (Z8) also features prominently. One further category, B3, relating to treatment (B3) occurs relatively fewer times in the overall corpus: 563. Below is a sample of the instances of words classified as B3, as shown through the concordance function.

because a while ago the **doctor**	told my grandma she had it
uld look at a differential **diagnosis**	of it I am Dr D o
ecstasy and those kinds of **drugs**	it is like a two pronged at
ay one of the differential **diagnoses**	between aortic and mitral p
think a lot of the time as **doctors**	we are too afraid to be dir
ship is that you know your **doctor**	is going to tell you the ba

Figure 12 Concordance showing examples of B3

The following table shows the relative occurrence of representative terms within this category, not in context, but in order of frequency.

Table 23 B3: Most frequent terms within B3 category

Word	Semantic tag	Frequency	Relative Frequency
clinical	B3	46	0.05
treatment	B3	40	0.04
drugs	B3	37	0.04

4.2 Study of the characteristics of the PBL spoken corpus in comparison with BNC Spoken (Educational)

However, the next stage of analysis requires us to compare the speech corpus with a comparator in order to evaluate whether any feature is significantly different from what might be expected. To achieve this, the British National Corpus, spoken education section, was selected as the comparator corpus. This is of an equivalent size, at 80,463 for the spoken educational section, slightly smaller in fact, but of all the existing spoken corpora available, this was the most suitable, being based on educational situations, including seminars, and lectures (BNC 2007).

The figure below sets out the range of possible choices from the BNC, with the spoken corpus to the left.

SPOKEN (990,704 words)		WRITTEN (1,010,690 words)	
Context-Governed (496,852)	Demographic (493,852)[by socio-economic class]	Imaginative (231,663)	Informative(779,027)
Leisure (136,606)	I (AB) (164,933)	Drama (23,786)	Pure science (32,974)
Educational (80,463)	II (C1) (98,700)	Poetry (30,144)	Applied science (117,685)
Business (134,275)	III (C2) (137,686)	Prose Fiction (177,733)	Social science (29,868)
Public/Institutional (145,508)	IV (DE) (92,533)		World affairs (277,128)
			Commerce & finance (92,057) Arts (51,645) Belief & thought (43,626) Leisure (134,044)

Figure 13- BNC Sampler Corpus

4.2.1 Whole Corpus compared to BNC spoken corpus, Educational

The following tables list the significant differences between the PBL corpus and the sample comparator corpus, in this case the BNC spoken, Educational category. The corpora are compared in terms of word frequency, parts of speech and finally by semantic category. Where appropriate, Wmatrix3 has been used to focus on the highest number of occurring words, but also to identify when words have very low occurrence in one corpus relative to another.

When making comparisons, in addition to relative frequency, we have two further measures which have been selected to differentiate between levels of statistical significance in difference. Rayson strongly recommends the use of log likelihood,

which is the formula which allows us to make comparisons in conjunction with an appropriate effect size measures, such as Bayes' (Rayson 2008) (Wilson 2013).

4.2.2 Significant differences and log likelihood (LL)

This study is restricting reported results to show those 'keywords' where there are a significant number of occurrences (keyword frequency) but also an unusual frequency when compared to an equivalent corpus. It also seeks to demonstrate the 'keyness' of these keywords; i.e.: the importance as a content descriptor. In each analysis there are two objective measures of significance, log likelihood, and Bayes' effect size.

4.2.3 Significance measures

There are overlapping definitions of what constitutes a measure of significant difference when comparing corpora. This is one definition from Scott:

"*Key words are those whose frequency is unusually high in comparison with some norm.*" (Scott 1997).

However, a word may be a key word in context because it does not appear, or it may over or under appear in comparison to other comparable texts. This may be significant. Scott again provides a definition.

"A key word may be defined as a word which occurs with unusual frequency in a given text. This does not mean high frequency but unusual frequency, by comparison with a reference corpus of some kind" ((Scott 1997) *p. 236).*

Finally, through statistical means it is possible to objectively compare significance. In Biber's words:

"The keyness of a keyword represents the value of log likelihood or Chi-square statistics; in other words, it provides an indicator of a keyword's importance as a content descriptor for

the appeal. The significance, (p-value) represents the probability that this keyness is accidental." ((Biber 2006) p. 138.)

4.2.4 Log Likelihood

The log likelihood figure demonstrates the significance of any difference between corpora. A negative figure indicates that the first corpus under represents the item when compared with the second corpus; a positive figure shows that the item is over represented. The measures of significance for log likelihood critical values are quoted when they reach the 99th percentile, i.e.: $p < 0.01$, which equates to a log likelihood level of 6.63. The significances given by Rayson, are given below.

Log-likelihood cut-off values

For 1 degree of freedom (1 d.f. is where comparing 2 corpora) the log-likelihood critical values are shown in the following figure:

- 95th percentile; 5% level; $p < 0.05$; critical value = 3.84
- 99th percentile; 1% level; $p < 0.01$; critical value = 6.63
- 99.9th percentile; 0.1% level; $p < 0.001$; critical value = 10.83
- 99.99th percentile; 0.01% level; $p < 0.0001$; critical value = 15.13

Figure 14 Log likelihood significance level (Rayson 2003)

Only results from the very highest order of log likelihood and Bayes' effect size are reported. All values which have been reported achieve significance at $p<0.05$ and above. The formulae for calculating both measures are fully described in Rayson (Rayson 2003).

4.2.5 Effect Size measure

The second measure, the Bayes factor (BIC) is recommended for interpretation as all positive scores indicate evidence against the result occurring by chance. Measures from 6 – 10 are strong evidence for significance, >10 are very strong evidence. An effect size metric measures the size of the difference between the normalised frequency of a word in two corpora, not the statistical significance of that difference. The LL values show the level of statistical significance of the difference observed. Effect size estimates, such as Bayes, are not dependent on sample size and provide information allowing comparison of relationships across studies, (Rosenfeld and Penrod 2011) (Gabrielatos and Marchi 2011). Their significance is summarised in the following figure:

Bayes Factor (BIC)

You can interpret the approximate Bayes Factor as degrees of evidence against the null hypothesis (H0) as follows:

0-2: not worth more than a bare mention

2-6: positive evidence against H0

6-10: strong evidence against H0

> 10: very strong evidence against H0

Figure 15 Summary of Bayes' effect size interpretation (Wilson 2013)

The corpora are compared in terms of word frequency, parts of speech and finally by semantic category (McEnerie 2011).

4.2.6 Analysis of frequency of word occurrence in PBL corpus v BNC spoken corpus

Table 24 Terms used significantly more frequently in the PBL corpus

Word	PBL Freq	Relative occurrence	BNC Ed	Relative occurrence	Log Likelihood	Bayes
is	3745	3.86	885	1.02	1591.34	1579.22
not	1558	1.61	311	0.36	772.99	760.87
heart_failure	369	0.38	0	0	470.88	458.76
heart	338	0.35	4	0	393.79	381.67
the	4769	4.92	2666	3.08	388.44	376.32
oedema	296	0.31	0	0	377.73	365.61
inaudible	277	0.29	0	0	353.48	341.36
yes	962	0.99	284	0.33	316.83	304.71
pressure	235	0.24	1	0	288.46	276.34
are	1045	1.08	368	0.43	265.85	253.73
blood	294	0.3	22	0.03	248.56	236.44
she	571	0.59	143	0.17	228.44	216.32
because	549	0.57	160	0.18	183.85	206.34

All the listed terms occur very significantly more frequently in the PBL corpus when compared to the BNC spoken educational corpus. This frequency list demonstrates that common terms, 'is', 'not', the', 'yes, 'are' and 'she' significantly over-occur in the PBL corpus in comparison with the BNC corpus. The PBL scenario related to a female patient with suspected heart problems, which may account for the high frequency of the female third person pronoun. The term 'Inaudible' also over-occurs but is of course the term added by the transcriber when a speaker is too quiet or too many participants are talking at once to be able to discern a reliable transcription. This term was retained in the corpus in order to identify patches of particularly intense over-talking, or quiet contributors, as that might be worthy of further investigation. The remaining terms, heart failure, heart, oedema, pressure and blood, are all technical terms relating to the topic of this PBL cycle. The term 'because' occurs significantly more frequently in the PBL corpus, suggesting intense concern with explaining causal sequence and accounting for observations.

Table 25 Terms used significantly less frequently in the PBL corpus in comparison to the BNC spoken

Word	PBL Freq	Relative occurrence	BNC Ed	Relative occurrence	Log Likelihood	Bayes
yeah	1	0	654	0.76	969.21	957.09
erm	5	0.01	674	0.78	960.27	948.15
er	7	0.01	363	0.42	485.08	472.96
now	79	0.08	375	0.43	244.74	232.62
okay	123	0.13	433	0.50	220.08	207.96

All the listed terms occur significantly **less** frequently in the PBL corpus when compared to the BNC corpus. The term 'yeah' is again related to a choice of the transcriber; occurring only once in the PBL corpus, with 654 occurrences in the BNC corpus. If 'yeah' and 'yes' were to be combined, this difference would be neutralised. The term 'erm', usually signifying a vocalised hesitation, again, is

chosen by the transcriber. It may signify that there is slightly more precision and fluency of speech in the PBL corpus. The term 'now' was also significantly underused. 'Okay' is also significantly underused in the PBL corpus when compared, perhaps signifying a more formal style of speech, or a change in the current norms for expressing agreement, such as the increased use of the term 'cool' to denote agreement.

4.2.7 The PBL corpus was examined with respect to the parts of speech (POS) to identify significant differences.

Table 26 Parts of speech (POS) used significantly more frequently in the PBL corpus

POS term	POS Code	PBL	Relative Frequency	BNC	Relative Frequency	LL	Bayes
Noun (s)	**NN1**	**12878**	**13.28**	**6370**	**7.36**	**1568.24**	**1556.12**
adjective	JJ	6515	6.72	2588	2.99	1334.92	1322.8
Adverb	RR21	374	0.39	0	0	477.26	376.43
Article, the	AT	4885	5.04	2748	3.17	388.55	297.8
Preposition	**II21**	**265**	**0.27**	**0**	**0**	**338.17**	**228.82**
Verb (pres)	VVZ	1417	1.46	544	0.63	309.92	213.75
is	VBZ	3754	3.87	2258	2.61	225.87	133.98
He/She	PPHS1	678	0.7	258	0.3	150.74	104.95
Subordinating conjunction*	**CS**	**2031**	**2.09**	**1172**	**1.35**	**146.1**	**87.33**
preposition	**II31**	**87**	**0.09**	**0**	**0**	**111.02**	**78.67**

The PBL corpus was compared to the BNC by Parts of speech (POS) analysis, which is inevitably going to relate closely to the frequency lists already reported so can play a valuable role in amplifying or triangulating observations identified through frequency observations.

For the PBL corpus, there is a significant over-occurrence of nouns in general, which have previously been shown to include many technical terms related to the PBL

case. There is also an over-occurrence of adjectives, of which the most frequent example was 'inaudible' as previously discussed, as well as terms such as 'swelling' and 'painful' which gave more detailed descriptions of processes. There was also over-use of adverbs, i.e. words describing the action of a verb, such as 'anymore', 'just', 'either' 'even' which might occur when precise descriptions of actions or processes are strongly featured. Prepositions were also over-used, again, 'in' 'on', 'through' etc. are strongly associated with detailed explanations where position is important, such as when describing the function of the heart.

As previously reported, subordinating conjunctions (CS) were noted by da Silva and Dennick (Da Silva and Dennick 2010) as varying between the phases of the PBL cycle, as these terms include causative reasoning terms such as, 'because', 'unless', 'so' and 'for'. The PBL corpus also shows these terms are significantly over represented in the cycle as a whole.

Table 27 Parts of speech (POS) used significantly less frequently in the PBL corpus

POS	POS	PBL	Relative Frequency	BNC	Relative Frequency	LL	Bayes
Unclass, e.g. Er, um	**FU**	**84**	**0.09**	**1462**	**1.69**	**1651.78**	**1639.66**
Single letter	ZZ1	124	0.13	833	0.96	672.18	660.06
Number	MC	699	0.72	1772	2.05	611.43	599.31
Interjection	**UH**	**1302**	**1.34**	**2005**	**2.32**	**240.94**	**138.62**
verb	VV0	2226	2.3	2706	3.13	117.07	80.34
adverb	RR	3689	3.8	4096	4.73	92.46	1556.12

Firstly, unclassified utterances, (FU) such as 'um' and 'er, were significantly less likely to occur than in the comparator corpus. Interjections, (UH), such as 'yes' or 'oh', were also significantly less likely to occur. This may suggest a higher fluency of speech across the corpus as a whole.

Table 28 Semantic categories: PBL corpus v BNC Ed Spoken: Subjects which occurred more frequently in the PBL corpus

Semantic category	Code	PBL Freq	Rel	BNC Freq	Rel	Log likelihood	Bayes
Anatomy	B1	3669	3.78	441	0.51	2543.18	2531.06
Disease	B2-	2246	2.32	133	0.15	2040.43	2028.31
Cause & Effect	A2.2	1254	1.29	106	0.12	1015.05	1002.93
Medicines	B3	563	0.58	88	0.1	334.97	322.85
Change	A2.1+	724	0.75	208	0.24	246.91	234.79
Quantities:	N5++	516	0.53	116	0.13	230.23	218.11

The PBL corpus was compared to the BNC Speech, Education corpus in order to identify significant differences in the frequency of occurrence of pre-defined semantic categories, i.e. the topics of discussion.

Unsurprisingly, the top two categories are the categories, B1 and B2, which are concerned with anatomy and physiology, and disease. Both categories appear with highly significantly greater frequency. The highest occurrence is for words associated with anatomy, rather than disease (B2,) or disease management and treatment (B3) which might reflect an appropriate focus for a first year PBL scenario, where the anatomical features and associated language are chief subjects of study. There are significantly over-occurring instances of medical technical vocabulary categories B1, B2 and B3, as well as talk related to cause and effect, and change, which have already been identified in stand-alone frequency analysis of the corpus. Comparison, particularly in understanding what is usual is strongly represented, as are discussion of substances, and liquids in particular, which on inspection, often relate to the discussion of blood, flow and pressure an area associated with the PBL

topic – the mechanisms of the heart. Measures are extremely high on both the significance measure of Log likelihood and the effect size measure Bayes.

When semantic frequency analysis was carried out, the following table of results identified the categories which occurred with very significantly **lower** frequency in the PBL corpus.

Table 29 PBL corpus v BNC Ed Spoken: Semantic analysis. Less likely to occur in PBL corpus

Semantic	Code	PBL Freq	Rel	BNC Freq	Rel	Log Likelihood	Bayes
Unmatched	Z99	2158	2.23	10819	12.5	7335.51	7323.39
Numbers	N1	789	0.81	2095	2.42	770.86	758.74
Discourse bin	Z4	2313	2.39	3993	4.61	663.8	651.68
Paper and writing	Q1.2	191	0.2	659	0.76	328.39	316.27
Evaluation: good	A5.1+	479	0.49	999	1.15	250.64	238.52
Time: asynchronous	T1.1.2	179	0.18	483	0.56	181.58	169.46
Evaluation: real	A5.4+	66	0.07	213	0.25	99.07	86.95

In comparison to the BNC speech corpus, numbers, the present moment of time, evaluation of what is good, authentic evaluation and the discourse bin Z4 (see discussion below), are all under-represented to a highly significant degree by all effect measures. The highest category was the unmatched class, Z99.

4.2.7.1 A note on the Z99 unmatched Category

Z99 is the semantic category which describes terms which cannot be identified by the semantic tagger. In this semantic frequency analysis it shows that the PBL corpus had significantly fewer of these terms than the comparator speech corpus. The analysis of the PBL corpus identified that unmatched terms accounted for 2.23 in relative frequency terms for the corpus as a whole as opposed to 12.5 for the BNC corpus. Given that the reliability of the semantic tagger is given as approaching

92% (Rayson 2003) using frequency analysis, the top 20 terms within the Z99 unmatched category were identified and further examined. For clarification, the top five most frequently occurring examples of utterances in this category are listed below with their actual and relative frequencies:

Table 30 List of Z99 word frequency occurrences

Word Frequency	Semantic tag	Frequency	Relative
mitral	Z99	46	0.13
I'm	Z99	22	0.06
thrombin	Z99	18	0.05
You've	Z99	17	0.05
angiotensin	Z99	14	0.04

As the unmatched terms occur significantly less frequently overall than in the comparator, this perhaps suggests that the corpus has achieved relatively high levels of semantic tagging, in comparison with the averages quoted of 92% (Rayson 2008).

4.2.7.2Z4: A note on the Z4, Discourse Bin

Z4, the discourse bin category refers to words which we use to 'fill in' speech but which are not obviously attributable to any other semantic category. These are very evident in spoken corpora and are analysed through the frequency and parts of speech analyses. For example, the most frequently occurring examples of utterances in this category are listed below with their actual and relative frequencies:

Table 31 Table of occurrences and relative frequencies for Semantic category, Z4

Utterance	Semantic Tag	Frequencies	Rel Freq
yes	Z4	962	0.99
no	Z4	213	0.22
sorry	Z4	126	0.13
right	Z4	106	0.11

4.2.8 Summary

This concludes the basic descriptive analysis of the PBL corpus either 'standing alone' or in comparison with the BNC spoken educational corpus. The corpus has been described in terms of frequency, noun, conjunctive phrases, parts of speech and semantic subject categories.

4.2.9 Summary: PBL corpus, stand alone and in comparison with BNC spoken educational corpus

In summary, the frequency analysis shows that the PBL corpus contains word frequency results which show that nouns over-occur and that they correlate with the subject of the PBL scenario, heart failure, and focus on the female patient. The transcribers' use of 'inaudible', which is shown to significantly over-occur, is likely to be associated with the challenges of quieter contributors and fast paced discussion between eight or nine people, who are likely to speak over each other far more than is customary even in other educational settings such as seminars. 'Not' and 'yes' also over-occur, again, as might be expected in an educational discussion where a PBL group seeks to establish individual and mutual understanding, but which may vary by individual type of student. A cluster of terms associate with informal agreement, hesitancy in speech and interjection, significantly under-occur, in frequency, and POS (FU unclassified) and prepositions and adverbs over occur, together perhaps suggesting a more formal and precise approach to speech overall. The semantic analysis confirmed the finding that the most frequently over-occurring

semantic categories were the three medical categories, B1, B2, B3, cause and effect and change and quantities, demonstrating close focus on the task. The results of analysis show the coordinating conjunction being significantly over-represented for the PBL corpus suggesting that the PBL discussion was operating successfully to generate an inquiry driven approach in comparison to the educational comparator corpus. Under-use was reported for terms which were unmatched, demonstrating acceptable operation of the semantic tagger, even for specific medical terms, and also for discourse 'filler' type language.

4.2.9.1 Comparators for the remaining Corpus Analysis

The remainder of the corpus analysis rests on comparing sub sections of the PBL corpus **with each other** rather than with a normative speech corpus as a means of exploring the factors which may be associated with learning differences for individuals and individual factors through PBL speech.

The corpora were created simultaneously by students from the same, selective cohort, discussing the same problem scenario, at the same stage in their course, facilitated by staff who are similarly trained and meet weekly to ensure their approaches to the scenarios support an equitable experience for each PBL group. This allows 'normal' speech behaviours between a small group of people who know each other, to be sampled, whilst controlling for much of the variety of influence which would usually cause speech to vary.

Corpus analysis allows us to identify significant ways in which speech varies, and will be presented in order of gender, personality type, for each of the five factor measures and by learning style, and finally by whether they are in the highest or lowest quartile for the measures at the beginning and close of the first year of PBL study.

The remainder of this chapter presents these results. Results will also be discussed in combination with the results of the statistical analysis, in the discussion chapter.

Each sub corpus has been analysed using the frequency, parts of speech and semantic analyses which have already been described. Only results from the very highest order of log likelihood and Bayes' effect size are reported. Full tabular analyses have been conducted but space limitations preclude full reporting.

4.3 A comparison using corpus analysis of the Female and Male Corpora

There is ample evidence that female and male speech behaviours tend to differ. The PBL corpus was subdivided into one female and one male sub-corpus. These corpora were then analysed against each other in order to determine significant differences between them which might have a bearing on how an individual differs in effectiveness in learning through PBL discussion, according to gender.

4.3.1 Analysis of Frequency of noun occurrence: female and male corpora

When analysing the frequency of nouns, the comparisons provided the following significant results.

Table 32 Terms used significantly more frequently in the female corpus and significantly less frequently in the male corpus

Word	Female Freq	Rel	Male Freq	Rel	Log Likelihood	Bayes
do	499	1.38	491	0.81	70.69	59.21
iron	63	0.17	18	0.03	55.36	43.88
okay	82	0.23	41	0.07	43.55	32.07
you	1069	2.96	1397	2.3	38.38	26.89
any	112	0.31	80	0.13	34.92	23.44
not	681	1.88	877	1.44	27.17	15.69
know	154	0.43	149	0.24	23.03	11.55
yes	431	1.19	531	0.87	22.88	11.4

The female corpus is more likely to over-use terms of agreement: 'okay', 'any', 'yes', but also disagreement; such as 'not'. It is also more likely to contain references to 'you' and 'know' perhaps implying awareness of other people and their thought processes. Only one technical term, iron, is identified as being significantly over-used in the female corpus.

The male corpus significantly underuses terms of agreement as listed above, mentioning iron significantly less frequently and is less likely to express disagreement using, 'not'. This may be a consequence of the commonly noted style of male communication where comments are framed as clear statements of fact and expected to be challenged in order to stimulate discussion (Tannen 2001). A female style of communication has often been observed to frame statements more hesitantly, or as questions, and are likely to be associated with more frequent appeals for agreement 'okay', 'what do you think?' in order to avoid triggering open disagreement (Baker 2014).

Table 33 Terms used significantly less frequently in the female corpus and significantly more frequently in the male corpus

Word	Female Freq	Rel	Male Freq	Rel	Log Likelihood	Bayes
pressure	49	0.14	186	0.31	29.54	18.05
water	28	0.08	129	0.21	28.33	16.85

Relatively few technical terms were identified to occur significantly less frequently in the female corpus in comparison with the male. Terms relating to the discussion of the physical forces influencing heart operation, such as 'pressure' and 'water', were significantly underused in the female corpus, in comparison with the male corpus.

4.3.2 Analysis of parts of speech occurrence: female and male corpora

*Table 34 Parts of speech terms used significantly **more** frequently in the female corpus and significantly **less** frequently in the male corpus*

Part of Speech	Part of Speech Code	Female Frequency	Rel	Male Frequency	Rel	Log Likelihood	Bayes
Do	VD0	388	1.07	378	0.62	56.76	45.28
Interjection	UH	602	1.67	700	1.15	43.72	32.23
You	**PPY**	**1085**	**3**	**1413**	**2.32**	**39.96**	**28.48**
Infinitive	VVI	1381	3.82	1900	3.12	32.25	20.76
Not n't	XX	728	2.01	949	1.56	26.62	15.14
He, she	**PPHS1**	**312**	**0.86**	**366**	**0.6**	**21.68**	**10.2**
I	**PPIS1**	**715**	**1.98**	**964**	**1.58**	**20.02**	**8.54**
Me	**PPIO1**	**55**	**0.15**	**40**	**0.07**	**16.57**	**5.09**
We	**PPIS2**	**327**	**0.91**	**408**	**0.67**	**16.08**	**4.59**

POS analysis supplies further evidence for significant differences in the use of pronouns between the genders. All parts of speech associated with personal pronouns (shown in bold) are over represented significantly in the female corpus. This analysis allows us to identify that this is particularly the case for PPY (you), PPHS1 (he/she), PPISI, (I), PPIO1, (me), and PPH01, (him/her). This perhaps shows evidence in this corpus of females demonstrating concern for identifying who is being discussed and the impact of events on the person and that female speech is more specific about the agent in the sentence, preferring a pronoun rather than focussing on the article in describing a process. Here is a concordance illustrating the use of 'you' in the female corpus, suggesting that the speaker is herself in the place of the sick person.

f she does not have COPD though , if **you** have heart failure you might still ge

D though , if you have heart failure **you** might still get a bit of pulmonary oe

ures they take of people ? If **you** have heart failure you do not like be

e ? If you have heart failure **you** do not like being in a bed , do you ?

you do not like being in a bed , do **you** ? **You** do not like being supine

Figure 16 Concordance illustrating use of 'You' in female corpus

Below is an extract from the male corpus, demonstrating that the use of 'you' is often from the perspective of discussing a system.

ic example. When you get an infarct **you** have a reduced outflow of blood . If

r the contractions the more adequate **you** can get. That will be a compensatory

lace within the body . Normally , if **you** constrict a vessel it should reduce t

ase that flow to the left heart when **you** constrict the veins leading up to the

heart ? It is because normally when **you** get such a network they start poolin

Figure 17 Concordance illustrating use of 'You' in male corpus

The second most significant difference was that the female corpus evidenced significantly more interjections (UH) such as 'yes', 'oh', etc. This may demonstrate a communication skill of demonstrating active listening through encouraging vocalisation, particularly expressing agreement, as 'yes' was the most frequently occurring interjection, and is used strongly in a consensus-building style of discussion. However, females also used the negative 'not' (POS category XX) more frequently. 'Yes' or 'no' are often responses triggered by the previous speaker making a statement; and information giving is often associated with male speech.

Unfortunately, this analysis cannot link statements with the previous speaker and so cannot examine this aspect more fully, but the frequencies within this POS is given below.

Table 35 Parts of Speech analysis. A comparison between Female and Male Interjections - (UH)

Word	Female Frequency	Female Relative Frequency	Male Frequency	Male Relative Frequency
Yes	431	1.19	531	0.87
No	110	0.30	103	0.17
Oh	32	0.09	31	0.05

Table 36 Parts of speech terms used significantly less frequently in the female corpus and significantly more frequently in the male corpus

Part of Speech	Part of Speech Code	Female Freq	Rel	Male Freq	Rel	Log Likelihood	Bayes
Article, the	AT	1650	4.57	3235	5.32	25.7	14.22
noun	NN1	4537	12.56	8340	13.71	22.8	11.32
preposition	II	1397	3.87	2735	4.5	21.33	9.85
of	IO	531	1.47	1116	1.83	18.12	6.64

There is a significant over use of single nouns, together with the definite article 'the' in male speech, as well as general prepositions and 'of' which may be associated with explaining systems.

4.3.3 Question word Use

There were no significant differences reported in the incidence of question words between the corpora, or with the relative use of the subordinating conjunctive statements (CS), given below, showing remarkable similarity in relative frequency of occurrence.

Table 37: A comparison between female and male conjunctive statement (CS) use

Word	Female Frequency	Female Relative Frequency	Male Frequency	Male Relative Frequency
if	240	0.66	383	0.63
because	203	0.56	346	0.57
so	153	0.42	219	0.36
when	64	0.18	105	0.17
cos	25	0.07	59	0.10
where	15	0.04	43	0.07

4.3.4 Semantic analysis: a comparison between the female and male corpora

Table 38 *Semantic categories used significantly **more** frequently in the female corpus and significantly **less** frequently in the male corpus*

Semantic	Code	Female Freq	Rel	Male Freq	Rel	Log Likelihood	Bayes
Pronouns	Z8	6140	16.99	8951	14.71	74.93	63.45
Discourse	Z4	1020	2.82	1293	2.13	45.31	33.82
Substance: solid	O1.1	116	0.32	79	0.13	39.43	27.95
Evaluation:	A5.1+	237	0.66	242	0.4	29.6	18.12
Negative	Z6	746	2.06	975	1.6	26.76	15.28

Females were very much more likely to use pronouns, conversational discourse filler terms, to mention solid substances (iron), to evaluate things positively, but also use negatives more frequently.

4.3.4.1 Comparison of pronoun use: female to male

The following table compares the use of the top ten most frequently occurring pronouns for each gender. A column remains blank where the term does not feature in the top category for that gender.

Table 39 Comparison of pronoun use: female to male

	Female Frequency	**Female Relative Frequency**	**Male Frequency**	**Male Rel Freq**
they			430	0.71
it	1209	3.35	383	0.63
you	1069	2.96	346	0.57
that	677	1.87	219	0.36
I	650	1.80	105	0.17
we	327	0.91	59	0.10
what	313	0.87	43	0.07
she	262	0.73	309	0.51
they	233	0.64	430	0.71
which			283	0.47
your	182	0.50	217	0.36

'Which' and 'they' did not occur in the top categories for the female corpus. 'What' occurred much more frequently in the female corpus, in many cases referring to the matter under discussion, for example: 'from what we can tell…'. The male corpus identifies 'what' with specific questions to other people. 'What is the blood pressure?' The pronoun use, particularly for 'you', 'I' and 'we', is strikingly high for the female corpus, as already identified in the initial frequency analysis.

*Table 40 Semantic categories used significantly **less** frequently in the female corpus and significantly **more** frequently in the male corpus*

Semantic	Code	Female Freq	Rel	Male Freq	Rel	Log Likelihood	Bayes
Grammatical	Z5	8894	24.62	16323	26.83	43.05	31.57
Unmatched	Z99	668	1.85	1489	2.45	37.5	26.01
Location	M6	441	1.22	1013	1.67	30.76	19.28
Measurement	N3.5	64	0.18	212	0.35	25.09	13.61
Substance, liquid	O1.2	73	0.2	232	0.38	24.74	13.26
Putting,	M2	123	0.34	342	0.56	24.44	12.96
Warfare,	G3	6	0.02	48	0.08	18.93	7.44
Change	A2.1+	222	0.61	502	0.83	13.86	2.38
Sailing, flow	M4	21	0.06	80	0.13	12.79	1.31
Anatomy	**B1**	**1271**	**3.52**	**2398**	**3.94**	**10.87**	**-0.61**

Females were much less likely to use grammatical terms (Z5), (such as 'the' 'a' etc. or to use terms related to location, measurement, liquid substances (particularly blood), warfare, change and sailing, which on inspection consists of terms, vessels, flow and flowing, which related to discussions of blood behaviour. They were also statistically significantly less likely to mention anatomical terms as defined in the B1 category. Although Z99, the unmatched category, lacks some of the necessary terms in the specialist dictionary for semantic tagging, male speakers did seem to more frequently use certain highly technical words, defined in the Anatomy (B1) category as a significant log likelihood (LL10.87). Below is a brief analysis of the Z99 category, to examine how unmatched terms may differ between the male and female corpora.

Table 41 An analysis of significant differences between occurrence of some highly technical terms, over occurring in the male corpus and under occurring in the female corpus

Frequent tech term	Female corpus	Rel freq	Male corpus	Rel freq	Log likelihood
thrombus	3	0.01	38	0.06	19.89
venous	8	0.02	48	0.08	14.62
osmotic	2	0.01	23	0.04	11.45
hypertrophy	10	0.03	46	0.08	10.08
interstitial	3	0.03	24	0.04	9.46

This does suggest that terms of a highly specific technical nature occur less frequently in the female corpus.

4.3.5 In summary

Male and female corpora differed in statistically significant ways. Females used more personal pronouns particularly 'you', often signifying identification with the patient, 'I' and 'we' and made more interjections, especially yes and no. Males used some highly technical terminology, particularly with respect to certain physically related terms and anatomical terms more frequently.

4.4 Corpora divided according to academic performance

It is of interest to explore the differences in speech between those students scoring high grades and those scoring lower grades.

The first measurement taken is the GAMSAT on entry score. Two corpora were created to allow comparison of speech behaviours between those in the high scoring quartile and those in the lowest quartile. The middle 50% of the corpus was excluded. There were equal numbers of speakers, 11, contributing to each corpus, but by chance, there was a difference in the number of sessions which were contributed by speakers to build the corpus. This led to a difference in speech sample of approximately 30% between these two sub corpora; the low GAMSAT scoring corpus being larger at, 28,506 words, the higher scoring GAMSAT corpus, being 20,184 words.

4.4.1 A comparison of low and high GAMSAT scoring learners

4.4.1.1 Frequency analysis

*Table 42 Frequency analysis showing categories used significantly **more** frequently in the low GAMSAT score corpus and significantly **less** frequently in the high GAMSAT score corpus*

Word	Low GAMSAT Frequency	Relative frequency	High GAMSAT Frequency	Relative frequency	Log Likelihood	Bayes
you	770	3.03	401	2.23	25.41	14.73
Thrombin	22	0.09	0	0.00	25.41	12.86
i	502	1.98	260	1.45	17.12	6.44
Factor	16	0.06	0	0.00	17.12	6.44
Clotting	16	0.06	0	0.00	17.12	6.44
Stroke	27	0.11	3	0.02	14.67	3.99

The low GAMSAT scorers mention 'you' and 'I' significantly more frequently than the high scorers, together with the technical terms, 'thrombin' 'factor', 'clotting' related to the chemical operations of 'blood clotting', and 'stroke'.

*Table 43 Frequency analysis showing categories used significantly **less** frequently in the low GAMSAT score corpus and significantly **more** frequently in the high GAMSAT score corpus*

Word	Low GAMSAT Frequency	Relative frequency	High GAMSAT Frequency	Relative frequency	Log Likelihood	Bayes
Thrombus	2	0.01	31	0.17	41.69	31.01
embolus	2	0.01	19	0.11	22.41	11.74
Water	27	0.11	50	0.28	17.25	6.56
Formation	2	0.01	14	0.08	14.76	4.08
Fatty	1	0.00	11	0.06	13.57	2.90
Going to	59	0.23	76	0.42	12.07	1.39
Atherosclerosis	2	0.01	12	0.07	11.81	1.13

When compared, the low scoring GAMSAT speakers mention 'embolus', 'water', 'formation' 'fatty' 'going to' and 'atherosclerosis' significantly less frequently than the high GAMSAT scorers. The closely related terms, thrombus, the term relating to dangerous movement of blood clots, and thrombin when combined still occur very significantly less frequently in the low GAMSAT speaker corpus, and more frequently in the high GAMSAT corpus.

4.4.1.2 Parts of speech analysis

*Table 44 POS analysis showing categories used significantly **more** frequently in the low GAMSAT score corpus and significantly **less** frequently in the high GAMSAT score corpus*

POS	POS	Low GAMSAT Frequency	Relative frequency	High GAMSAT Frequency	Relative frequency	Log Likelihood	Bayes
You	PPY	776	3.05	405	2.25	25.28	14.6
S/he	PPHS1	193	0.76	80	0.45	17.21	6.54
I	PPIS1	547	2.15	289	1.61	16.52	5.84
Formulae	FO	35	0.14	7	0.04	11.93	1.25
To do	VD0	245	0.96	119	0.66	11.75	1.07

Low scoring GAMSAT speakers were far more likely to use several pronouns more frequently: 'you' (PPY, 2nd person singular), (PPHS1, 'he', 'she'), and (PPIS1)'I'. Also, low scorers use the verb 'to do' (VD0) more frequently, often seeking confirmation, in conjunction with 'know'.

not think you do but **you maybe do** . It was not on the list . I wen
ix was C ; you can get both **you do not** just get one or the other .
was C , mixed or both of them **I do not** know if mixed is the right t
keletal muscle is activated . **I do not** even know if
one is adrenaline , although I **do not know** , noradrenaline is that
Inaudible I **do not know** the cytoplasmic

Figure 18 Concordance illustrating use of pronouns in low GAMSAT scorers

*Table 45 POS analysis showing categories used significantly **less** frequently in the low GAMSAT score corpus and significantly **more** frequently in the high GAMSAT score corpus*

POS	POS	Low GAMSAT Frequency	Relative frequency	High GAMSAT Frequency	Relative frequency	Log Likelihood	Bayes
alongside	RL	82	0.32	105	0.58	16.39	5.71
Comparison	RRR	22	0.09	42	0.23	15.19	4.52

The low GAMSAT speakers are significantly less likely than the high scorers, to use locative verbs (RL, alongside) and terms of comparison, more, longer etc. This might suggest they spend less time discussing specific details to do with the operation of the heart and its operation under different conditions.

4.4.1.3 Semantic analysis

*Table 46 Semantic analysis showing categories used significantly **more** frequently in the low GAMSAT score corpus and significantly **less** frequently in the high GAMSAT score corpus*

Semantic category	Code	Low GAMSAT Frequency	Relative frequency	High GAMSAT Frequency	Relative frequency	Log Likelihood	Bayes
Measurement: speed	N3.8	30	0.12	4	0.02	14.51	3.83
Cause & Effect	A2.2	349	1.37	181	1.01	11.81	1.13

Low scoring GAMSAT speakers were much more likely to mention measuring speed, in most instances, the term 'rate', also terms related to determining cause and effect. This suggests a close focus in the discussion on the PBL circulation process under discussion in this PBL scenario.

The specific word use in the cause and effect category was compared.

Table 47 Most frequently used words in cause and effect category

Word	Low GAMSAT Frequency	Low GAMSAT Relative Frequency	High GAMSAT Frequency	High GAMSAT Relative Frequency
why	58	0.23	31	0.17
cause	33	0.13	28	0.16
Causes	19	0.07	18	0.10
To do with	16	0.06		
factor	16	0.06		
Because of	14	0.06	9	0.05
effect	11	0.04	11	0.05
effects	11	0.04	8	0.04
Due to	11	0.04	9	0.05
Reason	10	0.04	11	0.06
Leads to	10	0.04		
Lead to	10	0.04	7	0.04
Depends			7	0.04

High scorers do not frequently use the rather vague terms, 'to do with' or 'factor' which identify a need for causality but do not determine it clearly. They ask 'why' significantly less frequently, and mention cause and causes, because of, due to and lead to, less frequently.

4.4.1.4 Question word Use

Table 48 A comparison of the occurrence of question words between the GAMSAT corpora

Question	Low GAMSAT Frequency	Relative Frequency	High GAMSAT Frequency	Relative frequency
why	58	0.23	31	0.17
How	31	0.12	29	0.16
when	34	0.13	29	0.16
where	25	0.10	29	0.16

The low GAMSAT scorers ask more 'why' questions, relative to the size of the corpus, and fewer how and where questions. Overall, there was a similar frequency of questions posed

.

4.4.1.5 Summary of the findings for high and low scorers on GAMSAT corpora

There were significant differences between the corpora for low and high scorers by the GAMSAT measure, which may relate to their ability to flourish during PBL in the early years of their medical studies. Low scorers used significantly fewer key technical terms related to the clinical aspects of the PBL scenario, and focused on rather general or biochemical level processes. Low scorers also more frequently used the pronouns, 'I', 'you' and 'she', where the subject of the PBL scenario was a woman and reminded themselves of formulae. High scorers were more likely to use POS related to comparisons and positions, suggesting they were able to discuss the overall system operation as opposed to dwelling on detailed aspects.

Although there was no significant overall difference in terms of semantic categories closely related to medicine, B1, B2, B3, low scorers were more likely to mention flow

rate and cause and effect, suggesting they were still focussed on discussing the details of the circulation process. Low scorers continued to ask 'why' questions and consider cause and effect more frequently than the high scorers, suggesting that they have already mastered the process and or have no need to verbalise their questions.

4.4.2 A comparison of high and low scoring learners in the PPD Personal and Professional Development test

This section of the analysis seeks to identify patterns of speech behaviour which may be associated with success in assessment which is not primarily knowledge based. The PPD assessment seeks to assess the personal and professional development of students, their insight into bio-psycho-social issues and how health relates to the individual as opposed to the pure knowledge assessments more commonly identified through the MCQ test, and knowledge and reasoning identified through the GAMSAT test.

The following corpora compare the top quartile of speakers, according to the personal and professional development assessment at the end of year one, with the lowest scoring quartile. The middle 50% of speakers have been removed in order to intensify any difference. There are 10 low PPD scoring speakers included contributing a corpus of 21,597 words, and 11 high PPD scorers contributing a corpus of 27,291 words.

4.4.2.1 Word frequency analysis

*Table 49 Frequency analysis: nouns occurring **more** frequently in the low PPD corpus*

Word	Low PPD Frequency	Relative frequency	High PPD Frequency	Relative frequency	Log Likelihood	Bayes
Her	**76**	**0.40**	**32**	**0.13**	**29.89**	**19.22**
she	**165**	**0.87**	**107**	**0.45**	**29.21**	**18.54**
Diuretic	15	0.08	1	0.00	18.10	7.43
Who	33	0.17	11	0.05	17.10	6.43
Symptoms	**26**	**0.14**	**7**	**0.03**	**16.40**	**5.73**
Chloride	10	0.05	0.00	0.0	16.27	5.61
Wrong	**18**	**0.09**	**3**	**0.01**	**15.58**	**4.91**

The terms most significantly over-used in the low score PPD corpus are 'her' and 'she' closely followed by 'who', suggesting focus on the individual. From the concordance, observations mostly relate to the female patient in the PBL scenario.

not remember You would put **her** on beta agonists to try to relieve th
tly impaired renal function. **Her** leg will retain more water . Y
mach ulcer she will be bleeding into **her** stomach , but you would also get mele
ou would be able to diagnose it from **her** faeces It depends on the ext
ation Loop . Measure **her** blood pressure . No , lose wat

Figure 19 Concordance illustrating pronoun use in low score PPD corpus

Low PPD scorer talk frequently mentions technical terms and processes, such as diuretic, symptoms, chloride and whether their conclusion is wrong. This might suggest that low PPD scorers are learners who are still focussed on processes, symptoms and facts rather than being yet capable of considering the full, more complex setting, but who do make more frequent observations about the patient.

*Table 50 Frequency of words used significantly **less** frequently in the low PPD score corpus and significantly **more** frequently in the high PPD score corpus*

Word	Low PPD Freq	Rel	High PPD Freq	Rel	Log likelihood	Bayes
pressure	**21**	**0.11**	**120**	**0.50**	**56.05**	**45.38**
its	5	0.03	52	0.22	35.16	24.50
proteins	**2**	**0.01**	**28**	**0.12**	**21.36**	**10.69**
protein	**1**	**0.01**	**21**	**0.09**	**18.09**	**7.42**

When compared, the low scoring PPD speakers mention the above terms, 'pressure', 'its', and protein/s', significantly less frequently than the high scorers. Again, this suggests that high scores may also be associated with significant discussion of technical issues related to blood pressure and protein mechanisms.

4.4.2.2 Parts of speech

*Table 51 POS used significantly **more** frequently in the low PPD score corpus and significantly **less** frequently in the high PPD score corpus*

POS	POS	Low PPD Freq	Relative frequency	High PPD Freq	Relative frequency	Log Likelihood	Bayes
He, she	PPHS1	172	0.91	141	0.59	14.2	3.53
Any, some	DD	87	0.46	63	0.26	11.27	0.6
name	NP1	140	0.74	120	0.5	9.47	-1.2

Low scoring PPD speakers were thus far more likely to use the 3rd person singular, (PPHS1, she, he), determiners, (DD, any, some), and singular proper nouns, (NP1, such as, London, Jane), although the Bayes measure for names was low. High PPD scoring speakers were significantly less likely to use 'he' or 'she', 'any' or 'some', or names. This appears counter-intuitive when high scorers for personal and professional development might be expected to pay more attention to individuals, and possibly colleagues, and to mention names more frequently.

*Table 52 Frequency of Parts of Speech used significantly less frequently in the **low** PPD score corpus and significantly **more** frequently in the high PPD score corpus*

POS	POS Code	Low PPD Freq	Relative frequency	High PPD Freq	Relative frequency	Log Likelihood	Bayes
Who	PNQS	33	0.17	11	0.05	17.10	6.43
About, in	RP	52	0.27	126	0.53	17.16	6.49

The low PPD speakers are significantly less likely than the high scorers, to use the singular determiner (DD1, this, that, another), and prepositional adverbs (RP, about, in).

4.4.2.3 Semantic analysis

*Table 53 Frequency of semantic categories which are used significantly more frequently in the low PPD score corpus and significantly **less** frequently in the high PPD score corpus*

Semantic Category	Sem Code	Freq	Rel Freq low PPD	Freq	Rel Freq High PPD	LL	Bayes
Disease	**B2**	**529**	**2.78**	**453**	**1.90**	**35.94**	**25.27**
Medicines	**B3**	**124**	**0.65**	**79**	**0.33**	**22.95**	**12.28**
Waves	W3	22	0.12	4	0.02	18.16	7.49
Female	S2.1	18	0.09	3	0.01	15.58	4.91
Time	T3	52	0.27	27	0.11	14.77	4.11

Low PPD scorers were much more likely to mention disease and medicines and the action of waves and pressure and the female patient, in comparison with the higher scoring PPD students. Again, this may appear counter-intuitive, where the impact of disease and medicine is likely to be considered in relation to the individual and in the

bio-psycho-social context. However, it does suggest that this group feels the need to discuss these categories more frequently.

*Table 54 Frequency of semantic categories which are used significantly **less** frequently in the low PPD score corpus and significantly **more** frequently in the high PPD score corpus*

Semantic Category	Sem Code	Freq	Rel Freq low PPD	Freq	Rel Freq High PPD	LL	Bayes
Weight	N3.5	31	0.16	134	0.56	47.97	37.30
Location	M6	243	1.28	451	1.89	24.89	14.23
Time Begins	T2+	31	0.16	93	0.39	19.92	9.25
Substances	O1	66	0.35	155	0.65	19.46	8.79
Comparing: Unusual	A6.2-	6	0.03	35	0.15	16.62	5.96

Low scoring PPD speakers were much less likely to mention the weight, the location, and substances in general. Low PPD scorers less frequently referred to the beginning of a sequence of events and unusual features.

4.4.2.4 Summary of low PPD and high scoring PPD corpora differences

From the analyses, there is evidence for significant difference between the corpora.

Low PPD scorers over-use 'her' and 'she' closely followed by 'who', suggesting focus on the individual. However low PPD scorers also over-use technical terms and processes, such as diuretic, symptoms, chloride and whether they are wrong. This might suggest learners who are more focussed on processes, symptoms and facts rather than a more complex setting. Yet, some highly relevant technical terms, 'pressure', 'its', protein/s' and 'that', are mentioned significantly less frequently than the high scorers. Again this suggests that high scores may be associated with significant discussion of technical issues related to blood pressure and protein mechanisms.

The POS analysis triangulates the frequency analysis showing that low scoring PPD speakers were thus far more likely to use the 3rd person singular, (PPHS1, she, he), determiners, (DD, any, some), and singular proper nouns, (NP1, London, Jane). This appears counter-intuitive when personal and professional development might be expected to pay more attention to individuals, in the case and to mention names more frequently. Lower PPD scorers were much more likely to mention disease and medicines and the action of waves and pressure in comparison with the higher scoring PPD students. Again, this may appear counter-intuitive, where the impact of disease and medicine is likely to be considered in relation to the individual and in the psycho social context. However, it does suggest that this group feels the need to discuss these categories more frequently. Low PPD scorers were however, more likely to mention terms associated with the female, perhaps suggesting focus on the female patient in this case. Low scoring PPD speakers were much less likely to mention the weight, the location, substances and materials in general. High PPD scorers referred to the beginning of a sequence of events and unusual features.

4.4.2.5A comparison of high and low scoring learners in the MCQ knowledge test

The following corpora compare the top quartile of speakers, according to the MCQ knowledge assessment at the end of year one, with the lowest scoring quartile. The middle 50% of speakers have been removed. There are 11 low MCQ speakers, and 10 high MCQ speakers included. By chance, the low MCQ corpus was also a larger sample size, due to more sessions being recorded for learners in that group. The low scorers contributed speech a corpus of 30,187 words. The high scorers contributed a corpus of 23,406 words.

4.4.2.5.1 Frequency analysis

Table 55 Frequency of words used significantly more frequently in the low MCQ score corpus and significantly less frequently in the high MCQ score corpus

Word	Low MCQ Frequency	Rel	High MCQ Frequency	Rel	Log Likelihood	Bayes
Iron	64	0.24	7	0.03	40.24	29.49
I	**502**	**1.91**	**253**	**1.22**	**35.41**	**24.66**
Within	25	0.10	1	0.00	22.27	11.51
Em	17	0.06	0	0.00	19.791	9.03
are	342	1.3	201	0.97	11.31	0.55
My	**37**	**0.14**	**7**	**0.03**	**15.97**	**5.21**
Palliative	12	0.05	0	0.00	13.97	3.21
Me	**36**	**0.14**	**8**	**0.04**	**13.28**	**2.52**
Care	16	0.06	1	0.00	12.66	1.90

When compared, the low scoring MCQ speakers mention the above terms significantly more frequently than the high scorers. 'Iron' and 'palliative' as technical terms may relate to particularly detailed discussions. However, the over use of 'I', 'my' and 'me' by low MCQ scorers appears, from examining the concordance, to relate to tentative observations about the PBL scenario in virtually every case. For example: 'I think', or 'I might be wrong?' In addition, 'are' is also significantly more frequently occurring, from the concordance, again, often linked with tentativeness.

d everybody else as well . So no , **I do not think** that is equivalent to yel
ing fire in a crowded theatre **I would not rule out** the 95% of cases in
and this is a complete guess because **I do not know** the oedema side of things
ng to take a while to set **in . I would imagine**, if it is mid-calf , it
, it is not going to happen overnight **I do not think** . Yes , so if we a
is progressive or along those lines . **I do not know** It is not as if h
t she had collapsed suddenly . **I thought collapse** was quite often to do
tely going to be about blood pressure **I would think**

Figure 20 Concordance illustrating pronoun use in low MCQ scorers

Low MCQ scorers are also more likely to use 'em', although this relates to a choice by transcribers, but may also suggest lack of fluency in discussion by lower MCQ scorers. Low scorers also used the term, 'care' significantly more frequently, which on inspection through concordance and collocation related to one long discussion of palliative care and the Liverpool care pathway.

Table 56 Frequency of words used significantly **less** frequently in the low MCQ score corpus and significantly **more** frequently in the high MCQ score corpus

Word	Low MCQ Frequency	Rel	High MCQ Frequency	Rel	Log Likelihood	Bayes
volume	5	0.02	34	0.16	31.58	20.82
pressure	39	0.15	78	0.38	24.08	13.32
aortic	19	3.27	50	0.24	22.71	11.95
plasma	1	0.00	17	0.08	10.69	0.07
cause	29	0.11	60	0.29	19.58	8.82
regurgitation	16	0.06	42	0.20	19.02	8.26

The high scoring MCQ speakers are much more likely to mention specific terms related to understanding heart failure processes: volume, pressure, aortic, plasma, regurgitation. High scorers are also much more likely to mention, 'cause', which suggests greater concern for causal links. Both these behaviours might be expected in learners who gain the highest scores in the MCQ knowledge test.

4.4.2.5.2 Parts of speech

*Table 57 POS analysis showing POS used significantly **more** frequently in the low MCQ score corpus and significantly **less** frequently in the high MCQ score corpus*

POS	POS	Low MCQ Frequency	Rel	High MCQ Frequency	Rel	Log Likelihood	Bayes
I	PPIS1	539	2.05	281	1.35	33.09	22.33
That	CST	262	1	139	0.67	14.89	4.13
Me	PPIO1	36	0.14	8	0.04	13.28	2.52
Are	VBR	343	1.31	200	0.96	11.91	1.15

Low scoring MCQ speakers were far more likely to use the 1st person singular, (PPIS1, 'I'), 'that' (CST), 1st person pronoun, (PPIO1, 'me'), and are, (VBR). The high MCQ scorers are much less likely to use these parts of speech.

*Table 58 POS analysis used significantly **less** frequently in the low MCQ score corpus and significantly **more** frequently in the high MCQ score corpus*

POS	POS	Low MCQ Frequency	Rel	High MCQ Frequency	Rel	Log Likelihood	Bayes
Can, will	VM	586	2.23	582	2.8	15.29	4.53
We	PPIS2	182	0.69	207	1	12.91	2.15

The low MCQ speakers are significantly less likely than the high scorers, to use the modal auxiliary (VM, can, will would), and 'we' (PPIS2), the verb, 'does', and locative verbs such as, alongside, numbers and the third person neuter, (PPH1, it).

4.4.2.5.3 Coordinating conjunction use

There were no significant differences reported in the incidence of question words between the corpora with the relative use of coordinating conjunctive statements (CS) given below, showing similarity in relative frequency of occurrence.

Table 59 Low to high comparison of question use in MCQ

Word	Low MCQ Frequency	Low MCQ Relative frequency	High MCQ Frequency	High MCQ Relative Frequency
If	169	0.64	140	0.67
Because	157	0.60	107	0.52
So	**64**	**0.24**	**87**	**0.42**
When	58	0.22	32	0.15
Where	16	0.06	26	0.13
Cos	12	0.05	14	0.07
where			11	0.05
before			8	0.04

High MCQ scorers used 'so', with higher relative frequency. The following concordances illustrate use by low scorers of 'so' as an intensifier, or to pose a possible conclusion.

the reason why it affected **her so badly** and she died was because her fr
supposed to be a pint an hour , **so you should** keep your water
not improve as much with it **, so you end up** with a big baggy heart

Figure 21 Concordance illustrating use of 'so' from low scorer MCQ corpus

The use in the higher scoring GAMSAT corpus has a more explanatory tone where 'so' emphasises the next stage in a reasoning sequence.

ten they are completely reversible , **so you just** have a blockage of a bit
tle dye and you can see it flowing , **so you can see** whether it's narrowing
en because it has a poor prognosis **, so it is only** used when carotid endo
e no one had done carotid stenosis , **so I thought** I would just leave the

Figure 22 Concordance illustrating use of 'so' from high scorer MCQ corpus

The following table gives the comparative occurrence of question words.

Table 60 Question words - occurrence in high to low MCQ scorers

Question	**Low** MCQ Frequency	Relative Frequency	**High** MCQ Frequency	Relative frequency
why	62	**0.24**	39	0.19
How	26	0.10	28	**0.13**
when	33	0.13	29	**0.14**
where	31	**0.12**	19	0.09

The low MCQ scorers ask more questions of the 'why' and 'where' variety, relative to the size of the corpus, whereas, the high MCQ scorers ask very slightly more 'how' and 'when' questions.

4.4.2.5.4 Semantic analysis

*Table 61 Semantic analysis showing categories used significantly **more** frequently in the low MCQ score corpus and significantly **less** frequently in the high MCQ score corpus*

Semantic category	**Code**	**Low MCQ Frequency**	**Rel**	**High MCQ Freq**	**Rel**	**Log Likelihood**	**Bayes**
Solids	O1.1	92	0.35	28	0.13	22.54	11.78
Geographical	Z2	49	0.19	9	0.04	21.71	10.95
Food	F1	61	0.23	15	0.07	20.06	9.30
Danger	A15	25	0.10	4	0.02	12.38	1.62
Dead	L1	48	0.18	15	0.07	11.27	0.51
Speech acts	Q2.2	204	0.78	109	0.53	11.23	0.47

Low scoring MCQ speakers were much more likely to mention solid substances and geographical names in comparison with the higher scoring MCQ students. One of the flaws in using semantic tagging is revealed when examining the concordances for geographical locations, where it states that MI, an acronym for a myocardial infarction, and the upper sternal border are geographical places! Low scorers are also more likely to mention food, found to refer most frequently to 'salt', which is relevant to the technical discussion, but also to discuss risk and risk factors, here under the semantic category of danger. Low scorers spoke more of death and dying and also about speaking.

*Table 62 Semantic analysis showing categories used significantly **less** frequently in the low MCQ score corpus and significantly **more** frequently in the high MCQ score corpus*

Semantic category	Code	Low MCQ Frequency	Rel	High MCQ Freq	Rel	Log Likelihood	Bayes
Volume	**N3.4**	**6**	**0.01**	**40**	**0.19**	**36.81**	**26.05**
Location	M6	320	1.22	365	1.76	23.08	12.33
Weight	N3.5	53	0.20	90	0.43	20.40	9.64
Unimportant	A11.1	13	0.05	32	0.15	13.39	2.63
Numbers	N1	174	0.66	199	0.96	12.75	1.99
Defence	G3	8	0.03	23	0.11	11.54	0.78
Putting,	M2	104	0.4	127	0.61	10.93	0.17
Anatomy	**B1**	**956**	**3.64**	**869**	**4.10**	**8.96**	**-1.80**

Low scoring MCQ speakers were much less likely than high MCQ scorers to mention, measurements of volume or weight, location numbers, defence, putting, or anatomy and physiology. On inspection, the unimportant category consisted of the term 'peripheral' which is a technical term in the context of discussing circulation.

However, this occurred significantly more frequently in the high MCQ scorers speech, suggesting more focus on details related to the PBL topic. Similarly, the terms occurring in the anatomy category, occurred significantly more frequently in high MCQ speech, although receiving a lower Bayes, effect size score.

When the top ten most frequent terms in this category were compared the following results were shown.

Table 63 Most frequently used words in Anatomy category: low MCQ scorers v high MCQ scorers

Word	Low MCQ Frequency	Low MCQ Relative Frequency	High MCQ Frequency	High MCQ Relative Frequency
Heart	99	0.38	53	0.63
Blood	68	0.26	72	0.57
Body	35	0.13		
Pulmonary	34	0.13	31	0.15
Cardiac	27	0.10	26	0.13
diuretics	22	0.08		
Blood pressure	21	0.08	37	0.18
lungs	19	0.07	19	0.09
aortic	19	0.07	50	0.24
chest	19	0.07		
regurgitation			42	0.20
renal			18	0.09
plasma			17	0.08

High MCQ scorers do not frequently use the general terms, 'chest' and 'body'. They do not use the technical term 'diuretics' but use a variety of technical terms such as, 'regurgitation', 'renal' and 'plasma' as terms to discuss the case and also use 'blood pressure' more frequently.

4.4.2.6 In summary: low MCQ scorers v high MCQ scorers

From all three analyses there is evidence of significant difference between the speech recorded for low MCQ scorers as opposed to high MCQ scorers.

When compared, the low scoring MCQ speakers used terms associated with tentativeness, and used some more advanced and specific technical terms much less frequently, for example, blood pressure', 'regurgitation', 'aortic', whilst using more general terms, such as 'body' and 'chest' more frequently. Low scorers also over-used 'I', 'my' and 'me' which related to tentativeness: 'I think', or 'I might be wrong?' The high MCQ speakers are significantly more likely than the high scorers to talk about 'we'.

Low scoring MCQ speakers were much more likely discuss risk and risk factors and spoke more of death and dying and also about speaking itself.

Low scoring MCQ speakers were much less likely than high MCQ scorers to mention, measurements or anatomy and physiology, suggesting a lack of detailed discussion related to the PBL topic. Similarly, specialist, more complex technical terms in the anatomy category, occurred significantly more frequently in high MCQ speech.

4.5 Personality: analysis by FFM

4.5.1 Analysis: Introvert: Extravert

The full PBL corpus was sub divided into two sub corpora, one of introvert speech and one of extravert speech. These were then subjected to the same analyses as have already been observed for the main and gender corpora in order to identify any highly significant differences.

As PBL necessitates intensive discussion it might be expected that extraverts and introverts may behave differently in their speech. The quantitative analysis did not find a significant difference in the overall quantity of speech, but did identify that extraverts took significantly more turns to speak than did introverts (Model 8), presented in the following results chapter. Corpus analysis of their separated corpora also identified differences in speech behaviour which may be relevant for enhancing learning behaviours.

4.5.1.1 Frequency of nouns: introvert v extravert

Table 64 Frequency of words used significantly more frequently in the introvert corpus and significantly less frequently in the extravert corpus

Word	Introvert Frequency	Relative Frequency	Extravert Frequency	Relative Frequency	Log Likelihood	Bayes
circulation	41	0.08	4	0.01	33.12	21.62
water	129	0.25	51	0.11	30.25	18.75
heart	231	0.46	126	0.26	25.13	13.63
liver	42	0.08	8	0.02	23.27	11.78
thrombus	36	0.07	5	0.01	24.53	6.06
pressure	160	0.32	87	0.18	17.58	6.08

This table shows the words which were significantly over-used by introvert students, when compared with extravert speakers, in order of log likelihood. The technical terms, circulation, water, heart, liver and pressure and the highly technical thrombus

appear significantly more frequently in introvert speech. All these terms are central to the PBL case.

*Table 65 Frequency of words used significantly **less** frequently in the introvert corpus and significantly **more** frequently in the extravert corpus*

Word	Introvert Frequency	Relative Frequency	Extravert Frequency	Relative Frequency	Log Likelihood	Bayes
Factor	0.00	0.00	16	0.03	23.20	11.7
It's	79	0.16	142	0.3	22.37	10.87
yes	433	0.85	544	1.14	20.53	9.04
so	373	0.74	464	0.97	16.42	4.92
cool	2	0.00	22	0.05	9.28	20.78

There is significantly lower word frequency in the introvert corpus of 'factor' 'it's' 'yes' and 'so' and 'cool'. This suggests that extraverts may be more likely to express agreement, and to link comments to the next, through using conjunctions such as, 'it's' and 'so'. There is also over-use of the informal term of agreement 'cool' in the extravert corpus.

*Table 66 Parts of Speech: categories used significantly **less** frequently in the introvert corpus and significantly **more** frequently in the extravert corpus*

Part of Speech	POS Code	Introvert Frequency	Relative Frequency	Extravert Frequency	Relative frequency	Log Likelihood	Bayes
Base verb	VV0	999	1.97	1193	2.51	31.46	19.97
General adv	RR	1723	3.4	1885	3.96	20.93	9.43
Interjection	**UH**	**592**	**1.17**	**716**	**1.5**	**20.8**	**9.31**

Introvert speech contains significantly fewer occurrences of the parts of speech concerned with base verbs such as 'give' 'work', (VV0), general adverbs, (RR) and interjections (UH).

Interjections, interrupt speech and often contain, terms such as 'yes' to express agreement, and 'oh' to express a response to what the listener has heard. This

seems to accord with what is theorised about extravert speech; that they might vocalise their internal states more frequently. Introverts might be more likely not to verbalise their inner states in this way. Below is a selection of interjections made by one extravert speaker

d like an echo . Oh I see **aha** , We still have to decide w
nervous mood so New areas , **bravo** So one of you could like em
t , somebody else could do **Yes** there's a few gaps There's an
That was my steer if it's wrong . **No** unfortunately I got so far with
So did I failed ! **Yes**

Figure 23 Interjections by extraverts

There were no significant differences reported in the incidence of subordinating conjunctions (CS) between the corpora, with the relative use, given below showing similar levels of frequency.

Table 67 Question word Use introvert v extravert

Word	**Introvert Frequency**	**Introvert Relative Frequency**	**Extravert Frequency**	**Extravert Relative Frequency**
if	327	0.65	314	0.66
because	306	0.60	264	0.55
so	178	0.35	198	0.42
when	107	0.21	77	0.16
cos	17	0.03	51	0.11
where	39	0.08	15	0.03

However, the extravert corpus includes many more occurrences of the informal 'cos' as opposed to 'because'. When those categories are added together, the extravert corpus includes a higher relative frequency of 'because/cos' than the introvert corpus. 'Where' and 'when' are more frequently used by the introvert speakers.

4.5.2 Semantic analysis: introverts compared to extraverts

*Table 68 Semantic analysis: categories used significantly **more** frequently in the introvert corpus and significantly **less** frequently in the extravert corpus*

Semantic category	Code	Introvert Frequency	Relative Frequency	Extravert Frequency	Relative frequency	Log Likelihood	Bayes
Anatomy	**B1**	**2136**	**4.22**	**1680**	**3.53**	**29.84**	**18.35**
Substance, liq'd	O1.2	219	0.43	119	0.25	24.11	12.62

In relation to extraverts, the introvert corpus significantly over-uses anatomical terms and the action of liquids, already identified as being about blood in this instance. These terms attract the highest log likelihood significance, and Bayes, effect size score and are highly relevant to the PBL scenario.

The following table illustrates the very wide range of semantic topics evident in extravert speech, in comparison with the two categories noted for introverts.

*Table 69 Semantic analysis: categories used significantly **less** frequently in the introvert corpus and significantly **more** frequently in the extravert corpus*

Semantic category	Code	Intr Freq	Rel Freq	Extr Freq	Rel freq	Log Likelihood	Bayes
Discourse	Z4	1065	2.1	1233	2.59	25.04	13.55
Medicines	**B3**	**223**	**0.44**	**315**	**0.66**	**22.09**	**10.59**
Evaluation: positive	A5.1+	192	0.38	273	0.57	19.7	8.2
Temperature:	O4.6-	4	0.01	26	0.05	19.43	7.94
People: male	S2.2	10	0.02	38	0.08	19.21	7.72
General	A1.1.1	527	1.04	636	1.34	18.18	6.69
Mental	X4.1	54	0.11	94	0.2	13.59	2.1
Open, showing	A10+	52	0.1	91	0.19	13.35	1.85
Knowledgeable	X2.2+	202	0.4	263	0.55	12.29	0.8
Pronouns	Z8	7645	15.09	7595	15.95	11.93	0.43
Psychological	X1	1	0	12	0.03	11.67	0.18
Time: beginning	T2++	74	0.15	114	0.24	11.26	-0.24
People: female	S2.1	15	0.03	37	0.08	11.03	-0.46
Likely	A7+++	5	0.01	20	0.04	10.6	-0.9
Parts	H2	41	0.08	69	0.14	9.06	-2.43
Vehicles	M3	17	0.03	37	0.08	8.89	-2.61
Personal name	Z1	98	0.19	135	0.28	8.44	-3.05
Comparing:	A6.2+	148	0.29	190	0.4	8.19	-3.31
Confident	E6+	1	0	9	0.02	7.87	-3.62
Degree: boosters	A13.3	289	0.57	339	0.71	7.72	-3.77
Comparing:	A6.2-	24	0.05	44	0.09	7.29	-4.21
Games	K5.2	11	0.02	26	0.05	7.23	-4.26
Participating	S1.1.3+	1	0	8	0.02	6.64	-4.85

I have chosen to include all measures of greater significance than the $p<0.01$ level for log likelihood in order to demonstrate the very significant differences in breadth of the topics of conversation between introvert and extravert students. The very diversity of topics is an objective demonstration of the wide range of topics often cited as typical in extravert speech. In relation to introvert speakers, the extravert corpus also significantly over uses semantic topics related to the following themes:

- PBL related tasks

Comparing positively, comparing unusual occurrences, likelihood, medicines and medical treatments.

- Related to learning and motivation and relating to others

Games, confidence, participating, being knowledgeable, showing, awareness of mental and psychological states, awareness of males and females, using personal names and pronouns.

- Category B3: Medicine and treatment

Extraverts are significantly more likely to discuss medicines, the way treatment effects the individual. When the differences in use in category, B3, medicines and treatments is examined, the following differences occur between introverts and extraverts in their most frequent usage.

Table 70 Differences in use of B3 category words: a comparison between introverts and extraverts

Word	Introvert Frequency	Relative Frequency	Extravert Frequency	Relative Frequency
treatment	14	0.03	25	0.05
clinical	23	0.05	23	0.05
drugs			20	0.04
palliative			14	0.03
ECG			13	0.03
systemic	12	0.03	14	0.03
doctor	11	0.02		
auscultate	11	0.02		

Introverts only discuss two terms with more relative frequency than the extraverts: the doctor and the term auscultate. Extraverts differ, mentioning the treatment, drugs, palliative care ECGs and the systemic impact with greater relative frequency.

*Table 71 Semantic analysis: categories used significantly **more** frequently in the introvert corpus and significantly **less** frequently in the extravert corpus*

Semantic category	Code	Extravert Frequency	Relative Frequency	Introvert Frequency	Relative frequency	Log Likelihood	Bayes
Anatomy, physiology	**B1**	**1680**	**3.53**	**2136**	**4.22**	**29.84**	**18.35**
Substances: Liq'd	O1.2	119	0.25	219	0.43	24.11	12.62
Selling	I2.2	10	0.02	47	0.09	23.82	12.62

The introvert corpus more frequently discusses speech classified within the anatomy and physiology category, blood, selling – which on inspection refers to 'circulation', here related to discussions of blood circulation.

4.5.3 In summary

The introvert and extravert corpora differed in statistically significant ways with respect to the frequency, parts of speech and semantic categories. There was significant evidence for extravert speech being less formal, less technically oriented, particularly with respect to anatomy and physiology, including more interjections, and discussing a far broader range of topics, particularly relating to people and treatments. Extraverts used terms related to disease (B3) significantly more frequently. Introvert speech was more formal, with fewer mentions of personal pronouns, treatments, topics related to the learning group processes and included more technical terms, particularly for anatomy and physiology (B1).

4.5.4 Low Conscientiousness (adaptable) v high conscientiousness (structured)

PBL necessitates both prepared and unprepared intensive discussion. High conscientiousness scores have been associated with thoroughness and preparation (McCrae and Costa Jr 1999) and so it might be expected that learners who are identified with lower conscientiousness (high adaptability) and those with higher conscientiousness, (structured) may behave differently in their speech. High conscientiousness has been repeatedly identified in the literature as being associated with higher ratings in the workplace (Doherty and Nugent 2011) and considered as a successful predictor for foundation medical studies (Finn, Carter et al. 2012). The descriptive statistics, will show that within this study, there were fewer learners identified as conscientious than might be expected by chance or by comparison with Quintax norms on this measure among professional workers (Robertson, Wilkie et al.). Overall, the proportion of individual learner speech did not significantly vary in connection with whether a learner was adaptable or structured. However, the sample sizes differed due to the low proportion of structured speakers within the sample. The sub corpus for high conscientiousness (structured) speakers was contributed by eight students, creating a corpus of 19,119 words compared to 33 low conscientiousness (adaptable) students contributing a corpus of 85,541 words.

4.5.4.1 Frequency corpus analysis

Frequency analysis did identify that the adaptable corpus did significantly under-use both medical terms, such as 'symptoms' 'aortic' and words associated with reasoning, such as 'cause' as shown in the table below.

*Table 72 Frequency analysis: terms used significantly **less** frequently in the adaptable corpus and significantly **more** frequently in the structured corpus*

Word	Adaptable Frequency	Relative Frequency	Structured Frequency	Relative Frequency	Log Likelihood	Bayes
Cause	88	0.12	52	0.3	26.98	15.54
symptoms	51	0.07	34	0.2	21.34	9.91
aortic	113	0.15	53	0.31	17.33	5.90

4.5.4.2 Parts of Speech analysis

In comparison to adaptable speech, structured talk significantly over-used 'would' as a way of discussing what 'would' happen, as well as using 'you', which was under used in the adaptable corpus and is shown in the table below.

*Table 73 Parts of Speech: categories used significantly **less** frequently in the adaptable corpus and significantly **more** frequently in the structured corpus*

Part of Speech	POS Code	Adaptable Frequency	Relative Frequency	Structured Frequency	Relative frequency	Log Likelihood	Bayes
Can, will	VM	1634	2.17	492	2.88	29.43	17.99
You	PPY	1872	2.48	528	3.10	19.30	7.86

*Table 74 Parts of Speech: categories used significantly **more** frequently in the adaptable corpus and significantly **less** frequently in the structured corpus*

Part of Speech	POS Code	Adaptable Frequency	Relative Frequency	Structured Frequency	Relative frequency	Log Likelihood	Bayes
Present verb	VVZ	1128	1.50	177	1.04	22.29	10.85
labelling	ZZ1	106	0.14	6	0.04	16.71	5.27

Adaptable talk featured the present tense of the verb and individual letters. This use of letters is illustrated through a concordance below which on inspection shows it to be related to intensive discussion about X-Rays, and a quiz, which used letters as labels to identify anatomical points.

Can you write that on please ? **W** Can we name it ? So are
p is the one on your left , **and then A** , Other way round A and lp . <
just drew . So the top left is **just B** If you do not name it seven t
ms a bit rigid But would **d be** your response to painful stimulus
Aortic regurgitation . **An x ray** . Shall we get a blood t

Figure 24 Concordance illustrating use of letters in quiz

4.5.4.3 Semantic analysis: adaptable v structured

There were no categories which were significantly over-represented in the adaptable corpus which had a high Bayes score in addition to the Log likelihood score. However, in relation to adaptable speech, the structured corpus significantly over-uses categories related to disease, (B2) and likelihood, two highly relevant semantic categories to PBL learning.

Table 75 Semantic analysis: categories used significantly less frequently in the adaptable corpus and significantly more frequently in the structured corpus

Semantic category	Code	Extravert Frequency	Relative Frequency	Introvert Frequency	Relative frequency	Log Likelihood	Bayes
Disease	B2	1680	3.53	2136	4.22	29.84	18.35
Likely	A7+	119	0.25	219	0.43	24.11	12.62

4.5.5 In summary

The structured and adaptable corpora differed in statistically significant ways with respect to the frequency, parts of speech and semantic categories. However, due to the structured corpus being the smallest, by sample size, there is limited evidence available. However, there was still significant evidence for structured speech being more technically oriented, particularly with respect to the disease and likelihood categories: close to the heart of effective PBL talk. In addition, there was significant over use by adaptable speakers of initial letters, which brought to light an extended 'learning game'.

4.5.6 A comparison between personable speakers and logical speakers – FFM Agreeableness

As PBL necessitates intensive discussion and interdependence within the PBL group, learners who are identified with lower agreeableness and those with higher agreeableness may behave differently in their speech.

4.5.6.1 Frequency corpus analysis

Frequency analysis did identify that the personable corpus significantly over-used both 'em' and 'OK', as shown in the table below.

*Table 76 Frequency analysis: terms used significantly **more** frequently in the personable corpus and significantly **less** frequently in the logical corpus*

Word	Personable Frequency	Relative Frequency	Logical Frequency	Relative Frequency	Log Likelihood	Bayes
em	52	0.13	0	0	78.71	67.35
Ok	36	0.18	16	0.14	37.14	25.78

Logical speakers significantly over-used various technical terms, such as those shown below.

*Table 77 Frequency analysis: terms used significantly **less** frequently in the personable corpus and significantly **more** frequently in the logical corpus*

Word	Personable Frequency	Relative Frequency	Logical Frequency	Relative Frequency	Log Likelihood	Bayes
Pressure	53	0.13	164	0.36	46.66	35.30
Thrombus	3	0.01	33	0.07	25.63	14.32

4.5.6.2 Parts of Speech analysis

Personable talk featured a highly significant occurrence of unclassified speech, er, um etc.

Table 78 Parts of Speech: categories used significantly more frequently in the personable corpus and significantly less frequently in the logical corpus

Part of Speech	POS Code	Personable Frequency	Relative Frequency	Logical Frequency	Relative frequency	Log Likelihood	Bayes
Er, Um	FU	66	0.16	14	0.03	43.44	32.08

This is briefly illustrated in the concordance below.

dietary intake , **em** , normal dietary intake is approx. 20
rmal person , we do have quite a few **em** , areas of iron storage so iron is st
o thank you . It was a test . **Em** which is probably why when you are gi
globin , haem , and that's within the **em** , er , ferrous , sorry ,

Figure 25: Concordance illustrating interjections in personable speech

4.5.6.3 Semantic analysis

In relation to highly logical speech, the personable corpus significantly under-uses categories related to disease, (B2) and (B1) Anatomy.

*Table 79 Semantic analysis: categories used significantly **less** frequently in the personable corpus and significantly **more** frequently in the logical corpus*

Semantic category	Code	Personable Frequency	Relative Frequency	Logical Frequency	Relative frequency	Log Likelihood	Bayes
Disease	B2	1680	3.53	2136	4.22	29.84	18.35
Anatomy	B1	1145	3.58	1867	4.09	14.39	3.03

4.5.6.4 In summary

There was significant evidence for logical speakers to be more technically oriented, particularly with respect to the disease and anatomy categories: close to the heart of effective PBL talk. In addition, personable speakers over-used hesitant terms such as 'em' and expressed signs of agreement, particularly 'ok'. This may be evidence

that logical speakers orient their focus towards facts and ideas as opposed to responding to the person who is speaking to them.

4.5.7 A comparison between Grounded speakers and theoretical speakers – FFM openness to experience

As PBL necessitates intensive discussion and the ability for learners to set their own topics and learning objectives to some extent, learners who are identified with lower openness to experience (incurious, grounded) and those with higher openness to experience, (with great interest in theoretical ideas, and high curiosity), may behave differently in their speech.

4.5.7.1 Frequency corpus analysis

Frequency analysis did identify that the grounded corpus significantly over-used the terms 'at the moment', and 'disease', as shown in the table below.

Table 80 Frequency analysis: terms used significantly more frequently in the grounded corpus and significantly less frequently in the theoretical corpus

Word	grounded Frequency	Relative Frequency	theoretical Frequency	Relative Frequency	Log Likelihood	Bayes
At the moment	9	0.04	1	0.00	18.45	7.05
Disease	28	0.12	26	0.04	16.52	5.13

The following concordance illustrates the use of the phrase 'at the moment' in context.

e middle , it's on regurgitation **at the moment** , what should we be hearing then
k . She is alert and orientated **at the moment** , no chest discomfort , severely
.5> She is not on any diuretics **at the moment** You would be quite cau
lying on a bed flat on her back **at the moment** and maybe that is why there is d
you have any medical conditions **at the moment** that we should be aware of ? Any
Are you still taking the spray **at the moment** ? Is that the only medic
e only medication that you have **at the moment** ? Was it an ACE inhibito
we're having problems breathing **at the moment** , though you seem to be talking
I ask you how you are sleeping **at the moment** ? Do you need to be propped up a

Figure 26 Grounded speech concordances showing awareness of time 'at the moment'

4.5.7.2 Semantic analysis

In relation to highly theoretical speech, the grounded corpus significantly under-uses categories related to constraint (pressure), (A1.7+) and Cause and Effect (A2.2); both of which are highly relevant to the PBL scenario.

Table 81 Semantic analysis: categories used significantly less frequently in the grounded corpus and significantly more frequently in the theoretical corpus

Semantic category	Code	Grounded Frequency	Relative Frequency	Theoretical Frequency	Relative frequency	Log Likelihood	Bayes
Constraint	A1.7+	8	0.03	84	0.13	17.53	6.14
Cause, effect	A2.2	242	1.05	925	1.40	16.70	5.31

4.5.7.3 In summary

There was significant evidence for grounded speakers to be more focussed on aspects of disease which are relevant to the current situation, but to under-use technical terms, such as related to pressure, which relate to the scientific principles, and how they may influence human systems in other circumstances.

4.5.8 A comparison between Calm speakers and Volatile speakers – FFM neuroticism

As PBL necessitates intensive discussion it might be expected that calm students and more volatile students may behave differently in their speech. The quantitative analysis identified that calm and volatile speakers were equal in number in this cohort, and did not show any significant difference in the proportion of their speech contribution.

4.5.8.1 Frequency corpus analysis

Table 82 Frequency of nouns used significantly more frequently in the calm corpus and significantly less frequently in the volatile corpus

Word	Calm Frequency	Relative Frequency	Volatile Frequency	Relative Frequency	Log Likelihood	Bayes
regurgitation	86	0.17	17	0.04	31.97	20.57
thrombus	34	0.07	2	0.01	25.59	14.19
dyspnoea	64	0.12	12	0.03	25.05	13.65
mitral	81	0.16	20	0.05	23.33	11.92
leads_to	31	0.06	2	0.01	22.63	11.23
systolic	41	0.08	5	0.01	22.29	10.89
murmur	43	0.08	7	0.02	19.06	7.66
angiotensin	43	0.08	7	0.02	19.06	7.66
embolus	23	0.04	1	0	18.84	7.44
causes	90	0.18	28	0.07	18.19	6.79

The calm corpus contains significantly more instances of discussion of technical terms such as 'regurgitation', 'thrombus', 'dyspnoea', 'mitral', 'murmur', 'angiotensin', and 'embolus'. The calm corpus also includes many more uses of 'leads to' and 'causes', suggesting a close attention to cause and effect and the order of processes.

Table 83 Frequency of nouns used significantly less frequently in the calm corpus and significantly more frequently in the volatile corpus

Word	Calm Frequency	Relative Frequency	Volatile Frequency	Relative Frequency	Log Likelihood	Bayes
iron	8	0.02	71	0.19	78.6	67.2
em	2	0	50	0.13	70.86	59.46
It's	82	0.16	138	0.36	36.4	25
ok	5	0.01	34	0.09	33.87	22.47
do	454	0.88	480	1.26	29.93	18.53
sodium	19	0.04	48	0.13	23.29	11.89
your	169	0.33	202	0.53	21.41	10.01

The calm corpus contains significantly fewer mentions of the term for iron, and sodium and terms 'it's' and 'do'. The term, 'em', transcribing hesitation, occurs very significantly more frequently in the volatile corpus, as does the term 'ok' perhaps expressing agreement, or vocalising that the listener is following the discussion or understands the term. 'Your' also occurs more frequently for volatiles, perhaps demonstrating awareness of the 'other', perhaps out of concern for how the other is perceiving them or out of continual comparison of others' performance with their own.

4.5.8.2 Parts of speech: a comparison between calm and volatile speakers

Table 84 Frequency of nouns used significantly more frequently in the calm corpus and significantly less frequently in the volatile corpus

Part of Speech	POS Code	Calm Frequency	Relative Frequency	Volatile Frequency	Relative frequency	Log Likelihood	Bayes
General adj	JJ	3646	7.09	2370	6.24	23.69	12.29
There (is)	EX	314	0.61	156	0.41	17.05	5.65

Calm speech contains significantly more occurrences of the parts of speech concerned with general adjectives (JJ), existential there, i.e. 'there is'.

When examining the JJ category, the highest number of instances related to the addition of 'inaudible' by the transcriber. Calm speakers were more likely to have the term 'inaudible' ascribed by the transcriber; volatile speakers were slightly less likely to be labelled inaudible, perhaps suggesting that they did not choose to compete and speak when other were speaking. Volatile speakers were significantly more likely to say 'sorry' and to say 'good'. This might suggest that volatile speakers are monitoring their utterances and their impact on others more frequently and interpreting events as either good or necessitating an apology – in common English usage not necessarily implying fault but simply an acknowledgement of something less than good having happened, which might occasion regret.

Table 85 Examination of most frequently occurring general adjectives JJ category, from POS analysis

Word	Calm Frequency	Relative Frequency	Volatile Frequency	Relative Frequency
inaudible	166	0.32	102	0.27
aortic	23	0.05		
Pulmonary	98	1.19		
left	89	0.17		
other	82	0.16		
mitral	73	0.14		
right	68	0.13		
good	50	0.10	54	0.14
Sorry	48	0.009	58	0.15

*Table 86 Frequency of nouns used significantly **less** frequently in the calm corpus and significantly **more** frequently in the volatile corpus*

Part of Speech	POS Code	Calm Frequency	Relative Frequency	Volatile Frequency	Relative frequency	Log Likelihood	Bayes
my your	APPGE	413	0.8	480	1.26	45.72	34.32
Unclassified	FU	20	0.04	62	0.16	37.16	25.76
Gen adverb	RR	1802	3.51	1548	4.08	18.87	7.47

By far the most significant measured difference relates to the category, APPGE, the possessive pro-noun (my, your, our, etc.) which is very significantly under-used in the calm corpus as opposed to the volatile corpus. The FU category consisted in the main of transcription of hesitation in speaking as 'erm' or 'um'. This was significantly more frequently occurring in the volatile corpus. In addition, volatile speakers were significantly more likely to use general adverbs, (RR), often qualifying what they said, for example using 'almost' or 'just' in frequent combination with statements.

There were again no significant differences between the use of question words or subordinating conjunctions.

4.5.8.3 Semantic analysis: comparison between calm and volatile speakers

*Table 87 Semantic analysis: categories used significantly **more** frequently in the calm corpus and significantly **less** frequently in the volatile corpus*

Semantic	Code	Calm Frequency	Relative frequency	Volatile Frequency	Relative frequency	Log Likelihood	Bayes
Objects	O2	516	1	258	0.68	27.21	15.81
Restriction	A1.7	73	0.14	19	0.05	19.57	15.81
Flow	M4	76	0.15	20	0.05	20.06	8.66
Disease	**B2-**	**1263**	**2.46**	**763**	**2.01**	**19.54**	**8.13**
Flow	M4	76	0.15	20	0.05	20.06	8.66
Sensory: inaudible	**X3.2**	**134**	**1.41**	**58**	**0.15**	**12.29**	**0.89**
Cause & Effect	**A2.2**	**726**	**0.33**	**441**	**1.16**	**10.66**	**-0.74**
Anatomy, physiology	B1	2044	3.98	1364	3.59	8.51	-2.89

Calm speakers are significantly more likely to discuss objects in general, and categories related to blood pressure, the category of disease, to be labelled as 'inaudible' and to discuss cause and effect and anatomy. This table also indicates that volatile speakers were significantly less likely to discuss two out of three medically specialist topics, B1, B2-, as well as less discussion of categories linked to early clinical reasoning, cause and effect.

Table 88 A comparison of terms most frequently mentioned in the Disease category

Word	Calm Frequency	Relative Frequency	Volatile Frequency	Relative Frequency
Heart failure	177	0.34	142	0.37
oedema	170	0.33	115	0.30
stenosis	117	0.23	48	0.13
dyspnoea	64	0.12	12	0.03
symptoms	44	0.09	29	0.08
disease	31	0.06	23	0.06
patient	**26**	**0.05**	**23**	**0.06**
Heart disease	22	0.04		
pain	**20**	**0.04**	**29**	**0.08**

The table above compares the incidence of technical terms categorised as B2, Disease. The same terms are mentioned in each corpus, but the overall incidence is lower for volatile speakers, except for the terms 'patient' and 'pain', perhaps suggesting a higher awareness of emotions.

*Table 89 Semantic analysis: categories used significantly **less** frequently in the calm corpus and significantly **more** frequently in the volatile corpus*

Semantic	Code	Calm Frequency	Relative frequency	Volatile Frequency	Relative frequency	Log Likelihood	Bayes
Solid	O1.1	51	0.1	136	0.36	70.1	58.7
Evaluation: good	A5.1+	205	0.4	232	0.61	19.83	8.43
Knowledge	X2.2+	212	0.41	223	0.59	13.55	2.14
obligation	S6+	206	0.4	215	0.57	12.53	1.13

The calm corpus is significantly less likely than the volatile corpus to discuss solid substances, but on inspection this is due to using a more varied vocabulary to describe 'things' including, terms such as valves, mechanism, pitting and plaque, rather than simply, 'things'. It seems therefore that the calm corpus may be making finer distinctions rather than speaking generally.

The volatile speakers are more likely to evaluate things positively using terms such as 'okay' well' 'fine'; to be concerned with being knowledgeable, using terms such as 'know' or remember or to be concerned over having a strong obligation.

Table 90 A Comparison of semantic category S6+; obligation between the calm and volatile speaker

Word	Calm Frequency	Relative Frequency	Volatile Frequency	Relative Frequency
should	40	0.11	61	0.12
need	58	0.15	48	0.09
Have to	40	0.11	32	0.06
Has to	11	0.03	12	0.02
must			10	0.02

The calm speakers are less likely to mention obligation in terms of what they 'should do' and 'need' to do, rather than the volatile speakers who express obligation

significantly more frequently and also in terms of their 'need' and 'having to' do something and that they 'must' do something.

4.5.8.4 Summary of calm and volatile corpora differences

The style of speaking differed, as calm speakers were found to be more frequently inaudible and mentioned technical terms related to the PBL case, disease and anatomy, more frequently and with a more diverse vocabulary. Volatile speakers were more often audible but expressed hesitancy through 'em' more frequently. Volatile speakers used possessive pronouns (my, your, our) and terms related to obligation more frequently and discussed the topics of anatomy and disease, cause and effect less frequently.

4.6 Quintax temperament Types: Four-way corpus analysis

When the PBL corpus is divided into four, as opposed to two, the results take more space to report and it might be argued to be less robust than two way analyses. For the following analyses, the corpus of interest has been compared to a larger corpus created from all speakers who are not of that type, with the intention of identifying relevant significant differences. It is also important to note that the sample size for the SG corpus was approximately 20% of the size of the other corpora (see section 4.1.4).

4.6.1 The AG Type: adaptable and grounded – the trouble shooter

4.6.1.1 AG v all other temperament types, word frequency comparison

Table 91 Words over-used by AG as opposed to other types

Word	AG Frequency	Relative frequency	Rest Frequency	Relative frequency	Log Likelihood	Bayes
problems	23	0.11	17	0.03	20.60	9.24
lie	7	0.03	0	0.00	20.05	8.69
At the moment	9	0.04	1	0.0	19.82	8.46
sign	13	0.06	5	0.01	18.69	7.33

The AG temperament type significantly over-uses 'problems', 'lie', 'at the moment' and 'sign' in comparison with the all the other temperament types combined.

Table 92 Words under-used by AG as opposed to other types

Word	AG Freq	Relative frequency	Rest Frequency	Relative frequency	Log Likelihood	Bayes
system	5	0.02	61	0.09	12.20	73.88

The AG type significantly underuses 'system' when compared to all the other temperaments combined.

4.6.1.2 AG v Rest of the corpus (LT, PT, SG combined) semantic category comparison

Table 93 Semantic categories over-used by AG as opposed to other temperaments combined

Semantic	Code	AG Freq	Relative frequency	Rest Frequency	Relative frequency	Log Likelihood	Bayes
Location	M6	382	1.87	926	1.42	19.44	8.08
Liking	E2+	41	0.20	54	0.08	16.98	5.62
Difficulty	A12	52	0.25	84	0.13	13.83	2.48

AGs are overusing terms related to location, liking, and difficulty relative to the rest of the temperament types.

Table 94 Semantic categories under-used by AG as opposed to other temperaments combined

Semantic	Code	AG Freq	Relative frequency	Rest Frequency	Relative frequency	Log Likelihood	Bayes
Change	A2.1+	111	0.54	523	0.8	15.21	3.85
Cause&Effect	A2.2	218	1.07	897	1.38	12.05	0.7

The 'trouble-shooter' AG temperament type is shown to be underusing terms related to change, cause and effect, in comparison with the rest of the temperament types combined.

4.6.1.3 Overall summary of the characteristics of AG speech

AG types focus on 'problems' 'in the moment' and consider the semantic domains of location, liking and difficulty more strongly than other types. However, they under use terms related to changing circumstances and cause and effect and the term

'system'. This profile strongly reflects the grounded corpus, but also is in line with the interpretation of this type as being likely to be strong at 'trouble-shooting'.

4.6.2 The SG Type: structured and grounded – the organiser and doer

The structured, highly conscientious type was least represented in the sample (18%, reported in section 5.2.4), and so contributes a far smaller corpus on which to base analyses.

4.6.2.1 SG v all other temperament types, word frequency comparison

Table 95 Words over-used by SG as opposed to other types

Word	SG Frequency	Relative frequency	Rest Frequency	Relative frequency	Log Likelihood	Bayes
Disease	12	0.35	41	0.05	23.64	12.28
Quiz	7	0.2	11	0.01	21.74	10.38

The SG temperament type significantly over-uses 'disease', in comparison with the all the other temperament types combined.

Table 96 Words under-used by SG as opposed to other types

Word	SG Frequency	Relative frequency	Rest Frequency	Relative frequency	Log Likelihood	Bayes
you	52	1.50	2113	2.58	17.66	6.31
if	8	0.23	598	0.73	15.63	4.27

The SG type significantly underuses 'you' 'if', when compared to all the other temperaments combined.

4.6.2.2 SG v Rest of the corpus (LT, PT, AG combined) POS comparison

Table 97 Parts of speech under-used by SG as opposed to other temperaments combined

Part of Speech	POS Code	SG Freq	Rel Freq	LT/PT/AG Freq	Rel freq	Log Likelihood	Bayes
You	PPY	52	1.50	2143	2.62	18.69	7.34
Now/tomorrow	RT	10	0.29	622	0.76	12.80	1.45
I	PPIS1	40	1.16	1431	1.75	7.57	-3.79

Comparison identifies a significant under-use of parts of speech for the second person personal pronoun 'you', adverbs relating to time and the first person singular pronoun, 'I' in comparison with the rest of the temperament types.

4.6.2.3 SG v Rest of the corpus (LT, PT, AG combined) semantic category comparison

Table 98 Semantic categories over-used by SG as opposed to other temperaments combined

Semantic	Code	SG Freq	Relative frequency	Rest Freq	Relative frequency	Log Likelihood	Bayes
Time, age	T3	21	0.61	102	0.12	30.66	19.31

SGs are overusing terms related to time and age relative to the rest of the temperaments combined.

Table 99 Semantic categories under-used by SG as opposed to other temperaments combined

Semantic	Code	SG Freq	Relative frequency	Rest Frequency	Relative frequency	Log Likelihood	Bayes
If	Z7	9	0.26	606	0.74	13.93	2.57

SGs are underusing 'if' significantly less frequently than other types.

4.6.2.4 Overall summary of the characteristics of SG speech

The SG corpus focusses on time now rather than tomorrow, and underuses pronouns, 'I' and 'you' but uses 'disease' significantly more frequently than all other speakers combined. The sample of structured personality types is small (, and again, the influence of the grounded speaker is evidence, within the role of the 'organiser, doer'.

4.6.3 Logical and theoretical Temperament Types – logical strategist

Table 100 Words over-used by LT as opposed to other types

Word	LT	Relative frequency	PT/AG/ SG	Relative frequency	Log Likelihood	Bayes
Thrombus	31	0.09	5	0.01	31.79	20.44
Embolus	22	0.06	2	0.00	27.76	16.41
System	47	0.13	19	0.04	24.94	13.58
thrombi	12	0.03	0	0.00	21.51	10.15

LT types seem significantly more comfortable with using specific technical terms, 'thrombus' 'embolus' 'system and 'thrombi'.

Table 101 Words under-used by LT as opposed to other types

Word	LT	Relative frequency	PT/AG/ SG	Relative frequency	Log Likelihood	Bayes
It's	49	0.14	170	0.34	33.30	21.95
Sodium	6	0.02	58	0.11	31.77	20.41
no	70	0.20	201	0.40	25.67	15.32

LT types underuse 'it's' 'sodium' 'no' relative to the other types, very significantly according to log likelihood measures.

4.6.3.1 Parts of speech Analysis

Table 102 Parts of speech under-used by LT as opposed to other temperaments combined

Parts of speech	POS	LT	Relative frequency	PT/AG/ SG	Relative frequency	Log Likelihood	Bayes
Em, erm	FU	12	0.03	69	0.14	25.93	14.57
Oh, yes,	UH	397	1.14	782	1.55	25.44	14.09

LT underuses unclassified terms, which in the main prove to be hesitations, rendered as 'er' or 'erm', together with interjections (UH), in the main, 'yes'.

4.6.3.2 LT v Rest of the corpus (AG, PT, SG combined) semantic category comparison

Table 103 Semantic categories over-used by LT as opposed to other temperaments combined

Semantic	Code	LT Frequency	Relative frequency	Rest Frequency	Relative frequency	Log Likelihood	Bayes
Sailing, swimming	M4	58	0.17	33	0.07	19.37	8.02
Means, method	X4.2	87	0.25	74	0.15	11.41	0.06
Anatomy	B1	1436	4.12	1852	3.66	11.06	-0.29

Terms such as vessels and waves, were categorised semantically, as sailing and swimming, when they were terms being used to describe cardiovascular processes and were significantly overused, together with methods and the anatomy category opposed to the rest of the temperament types, although the Bayes effect size was shown to be low for B1.

Table 104 Semantic categories under-used by LT as opposed to other temperaments combined

Semantic	Code	LT Freq	Relative freq	Rest	Relative freq	Log Likelihood	Bayes
Materials, solid	O1.1	23	0.07	160	0.32	70.67	59.32
Discourse	Z4	747	2.14	1335	2.64	21.29	9.94
Knowledgeable	X2.2+	128	0.37	290	0.57	18.6	7.25

LTs underuse some terms describing materials, in this case the term 'sodium'. In addition, LTs under-use phrases from the discourse bin (yes, sorry, bless her etc.) and mention knowledgeability (know, remember) significantly less frequently.

4.6.3.3 Overall summary of the characteristics of LT speech

The LT corpus overused technical terms which were closely related to the ideas behind the PBL scenario, and underused terms relating to hesitant speech or acknowledging the contributions of others by signalling their agreement or disagreement with others.

4.6.4 PT Temperament Types – passionate idealist

4.6.4.1 PT v Rest of the corpus (AG, LT, SG combined) frequency category comparison

Table 105 Words over-used by PT as opposed to other types

Word	PT	Relative frequency	LT/AG/SG	Relative frequency	Log Likelihood	Bayes
Iron	66	0.25	12	0.02	95.57	84.21
Em	47	0.18	5	0.01	80.17	68.82
ok	32	0.12	7	0.01	42.99	31.63
know	123	0.46	146	0.25	24.62	13.26

The summary position when comparing PT types with the rest of the corpus shows that the term 'iron' is overused. 'Em' and 'OK' are also overused by PTs relative to the other types. In addition, 'know' is over represented to a highly significant degree.

Table 106 Words under-used by PT as opposed to other types

Word	PT	Relative frequency	LT/AG/SG	Relative frequency	Log Likelihood	Bayes
Pressure	37	0.14	177	0.30	21.63	10.28
Water	23	0.09	122	0.21	18.08	6.73

PT types underuse the specific technical term, pressure, and water in comparison with the other types.

4.6.4.2 PT v Rest of the corpus (AG, LT, SG combined) POS category comparison

Table 107 Words over-used by PT as opposed to other types

Part of Speech	POS	PT	Relative frequency	LT/AG/SG	Relative frequency	Log Likelihood	Bayes
Unclass, e.g. Em	FU	60	0.22	21	0.04	62.62	51.26
General adverb	RR	1148	4.3	2048	3.49	31.7	20.34

PT types are over-using the term 'em' whilst speaking, and general adverbs.

PT v Rest of the corpus (AG, LT, SG combined) semantic category comparison

Table 108 Semantic categories over-used by PT as opposed to other temperaments combined

Semantic category	Code	PT Frequency	Relative frequency	LT/SG/AG Frequency	Relative frequency	Log Likelihood	Bayes
Solids (sodium, iron)	O1.1	114	0.43	69	0.12	74.41	63.06
Knowledgeable	X2.2+	187	0.7	231	0.39	33.32	21.96
Discourse bin	Z4	752	2.82	1330	2.27	22.42	11.07
Evaluation: +	A5.1+	173	0.65	248	0.42	18.12	6.76

PT temperament types overuse phrases related to solids, primarily (sodium and iron) knowledgeability (know, remember), phrases from the discourse bin (yes, sorry, bless her etc.), and vocalising positive evaluation (good, ok, fine).

Table 109 Semantic categories under-used by PT as opposed to other temperaments combined

Semantic category	Code	PT Frequency	Relative frequency	LT/SG/AG Frequency	Relative frequency	Log Likelihood	Bayes
Location	M6	316	1.18	992	1.69	32.07	20.71
Weight: pressure	N3.5	46	0.17	207	0.35	22.21	10.85
Disease	**B2-**	**513**	**1.92**	**1428**	**2.43**	**21.62**	**10.26**

Compared to the other temperament types, PT types mention location and pressure (related to discussions of blood pressure) and disease significantly less frequently. Overall, this does suggest a significant reduction in close discussion of the processes related to the cardiovascular system in comparison with the other types.

4.6.4.3 Overall summary of the characteristics of PT speech

The PT corpus underused technical terms, particularly associated with disease, which were closely related to the content of the PBL scenario. This group overused terms relating to hesitant speech and acknowledging the contributions of others by signalling their agreement or disagreement.

4.7 Kolb Learning Styles analysis

As noted in the literature review, the evidence for learning styles has been summarised as being weak and conflicting (Coffield, Moseley et al. 2004). The corpora for speakers whose learning styles were known, (thirty-eight) ranged from six accommodating learners contributing 17,681 words, to thirteen assimilating learners, who contributed a 31,589 word corpus. In the interests of brevity only the most salient results from corpus analysis will be reported for each of the four types: Convergent, Accommodating, Assimilating, Divergent. Each corpus is analysed in comparison with a corpus made up of the three other learner types. For example, convergent learners are compared against a combined corpus of accommodating, assimilating, and divergent types combined.

4.7.1 A comparison of convergent learners – the devil's advocate, utilitarian or concrete learner

4.7.1.1 Frequency corpus analysis

*Table 110 Words **over**-used by convergent types as opposed to other types*

Word	Con-vergent	Relative frequency	Ass/Acc/Div	Relative frequency	Log Likelihood	Bayes
oedema	104	0.5	161	0.27	24.08	12.78
you	618	3	1489	2.47	15.95	4.65
we	202	0.98	418	0.69	15.6	4.3
Pressure	61	0.30	96	0.16	13.50	4.30

Convergers used the technical terms 'oedema' and 'pressure' significantly more often than other types. Pronouns 'you' and 'we' are also over represented.

Table 111 Words under-used by convergent types as opposed to other types

Word	Con-verger	Relative frequency	Ass/Acc/Div	Relative frequency	Log Likelihood	Bayes
That's	13	0.06	120	0.20	20.99	9.69
Iron	5	0.02	72	0.12	19.01	7.71
they	106	0.51	461	0.76	14.7	3.4

Convergers under-used the technical term, 'iron' compared to the other types, and also terms 'that's' and 'they'.

4.7.1.2 POS category comparison

*Table 112 Parts of Speech, **over**-used by convergent types as opposed to other types*

POS	POS	Converger	Relative frequency	Div/Acco/Assim	Relative frequency	Log Likelihood	Bayes
Em, um	FU	43	0.21	39	0.06	27.04	15.74
You	PPY	628	3.04	1507	2.5	16.87	5.56
We	PPIS2	202	0.98	418	0.69	15.6	4.3
For example	REX21	14	0.07	9	0.01	12.79	1.49

Convergent types over-use unclassified phrases such as 'erm' or 'um' as well as the pronouns, 'you', and 'we'. They also over-use the specific phrase, 'for example' significantly more frequently.

*Table 113 Parts of Speech, **under**-used by convergent types as opposed to other types*

POS	POS	Con-verger	Relative frequency	Div/Assim/Accom	Relative frequency	Log Likelihood	Bayes
that	CST	138	0.67	596	0.99	18.42	7.11
The	AT	907	4.4	3108	5.15	18.13	6.83
They	PPHS2	106	0.51	467	0.77	15.76	4.46

The parts of speech which are significantly under-used by Convergent types, are 'that' 'the' and 'they'.

4.7.1.3 semantic category comparison

*Table 114 Semantic categories, **over**-used by convergent types as opposed to other types*

Semantic	Code	Converger Frequency	Relative frequency	Div/Assim/ Accom Frequency	Relative frequency	Log Likelihood	Bayes
Comparing: usual	A6.2+	113	0.55	158	0.26	33.76	22.46
Disease	B2-	528	2.56	1306	2.16	10.3	18.24

Convergent types are over-using terms of comparison relating to usualness, and disease.

4.7.1.4 Summary of the differences between convergent types and the other learning style types

Convergent types over-use specific technical terms, 'pressure' and 'oedema', unclassified phrases such as 'erm' or 'um' as well as the pronouns, 'you', and 'we'. They also over-use the specific phrase, 'for example' significantly more frequently, perhaps indicating their purported 'utilitarian' or concrete approach to learning. They discuss what is usual and the disease category (B2) more frequently than other types combined.

4.7.2 A comparison of Accommodators –hands-on learners, cooperative, concrete learner

4.7.2.1 Frequency corpus analysis

*Table 115 Terms **over**-used by accommodator types as opposed to other types*

Word	Accom	Rel	Assim/ Con/ Div	Rel	Log Likelihood	Bayes
factor	16	0.10	1	0.00	45.33	34.03
thrombin	18	0.11	4	0.01	39.93	28.63
platelets	14	0.09	4	0.01	28.60	17.30
extrinsic	8	0.05	0	0.0	26.25	17.30
Afterload	8	0.05	0	0.0	26.25	14.95
Clotting	14	0.09	7	0.01	22.23	10.93
Intrinsic	8	0.05	1	0.00	20.41	9.10

Accommodators significantly over-use the technical terms, 'factor', 'thrombin', 'platelets', 'extrinsic', 'afterload', 'clotting' and 'intrinsic'.

Table 116 Terms under-used by accommodator types as opposed to other types

Word	Accom	Rel	Assim/ Con/ Div	Rel	Log Likelihood	Bayes
Iron	1	0.01	76	0.12	25.35	14.05
kidneys	0	0.00	37	0.06	15.94	4.64
em	1	0.01	50	0.08	14.98	3.68
if	87	0.55	510	0.78	9.47	1.83

Accommodators under-use the technical terms iron, kidneys. Accommodators use the filler 'erm' and the reasoning term, 'if' significantly less frequently.

4.7.2.2 Parts of speech analysis

*Table 117 Parts of Speech analysis, showing **under**-use by accommodator types as opposed to other types*

POS Category	POS	Accom	Relative frequency	Div/ Con/ Assim	Relative frequency	Log Likelihood	Bayes
General adj	JJ	908	5.79	4534	6.95	26.28	14.98
Er	FU	3	0.02	79	0.12	18.15	6.84

Accommodators significantly under-use general terms, of which many were instances of inaudibility, as also signified by the under-use of vocal filler, 'er'.

4.7.2.3 semantic category analysis

*Table 118 Semantic analysis, showing **over**-use by accommodator types as opposed to other types*

Semantic	Code	Accom Freq	Relative freq	Div, Assim/Con Frequency	Relative frequency	Log Likelihood	Bayes
Unusual	A6.2	26	0.17	36	0.06	16.51	5.20
Time begins	T2+	66	0.42	148	0.23	15.94	4.63

Accommodators over-use comparison about whether something is usual or not and beginning processes.

4.7.2.4 Summary of Accommodator types

Accommodators over-use certain highly technical terms relating to the heart, but underuse some related to the kidneys and iron. They avoid vocal fillers such as 'er' and avoid inaudibility, but also the reasoning term 'if'. Their discussion focuses on what is unusual and the beginning of processes.

4.7.3 A comparison of assimilators – the fact-oriented, utilitarian, abstract learner

4.7.3.1 word frequency category

*Table 119 Word frequency analysis, showing **over**-use by accommodator types as opposed to other types*

Word	Assim	Relative frequency	Accom/Con /Div	Relative frequency	Log Likelihood	Bayes
thrombus	31	0.11	5	0.01	40.89	29.59
inaudible	114	0.41	93	0.18	35.81	24.51
embolus	21	0.07	3	0.01	28.94	17.64
atheroma	13	0.05	0	0.0	27.53	16.23
kidneys	28	0.10	9	0.02	25.91	14.60
thrombi	12	0.04	0	0.0	25.41	14.11

A variety of technical terms, shown above, are significantly over-used by assimilators when compared to the rest. Inaudibility is significantly higher for this type.

Table 120 Word frequency analysis, showing under-use by accommodator types as opposed to other types

Word	Assim	Relative frequency	Accom/ Con/ Div	Relative frequency	Log Likelihood	Bayes
Obviously	7	0.02	71	0.13	28.21	16.91
Em	3	0.01	48	0.09	24.43	13.13
Iron	8	0.03	69	0.13	24.36	13.06
We	164	0.58	456	0.86	19.40	8.10

The term 'obviously', is comparatively under-used, as is the 'em' filler, the technical term, iron, and the pronoun, 'we'.

4.7.3.2 POS analysis

Table 121 Parts of Speech analysis, showing under-use by assimilator types as opposed to other types

POS	POS	Assim	Relative frequency	Div/ Con/ Accom	Relative frequency	Log Likelihood	Bayes
we	PPIS2	164	0.58	456	0.86	19.4	8.1

Assimilators under-use the 'we' pronoun.

4.7.3.3semantic analysis

*Table 122 Semantic analysis, showing **over**-use by assimilator types as opposed to other types*

Semantic	Code	Assim Freq	Relative frequency	Con/ Accom/ Div Frequency	Rel Freq	Log Likelihood	Bayes
Inaudible	X3.2	116	0.41	96	0.18	35.42	24.12
Size: big	N3.2+	65	0.23	58	0.11	16.94	5.64
Anatomy	B1	1169	4.16	1888	3.57	16.78	5.48

Assimilators over-use speech which is inaudible, comment on the large size of things and anatomical terminology.

*Table 123 Semantic analysis, showing **under**-use by accommodator types as opposed to other types*

Semantic	Code	Assim Freq	Relative frequency	Con/ Accom/ Div Frequency	Relative freq	Log Likelihood	Bayes
Materials: solid	O1.1	28	0.10	123	0.23	19.27	7.97
Comparing: usual	A6.2+	65	0.23	206	0.39	14.56	3.25
Pronouns	Z8	4206	14.97	8500	16.07	14.19	2.89

Assimilators under-use terms related to solids (sodium and iron), comparing and general pronouns.

4.7.3.4 Summary of Assimilator type

Assimilators over-use relevant highly technical terms, comparative size and anatomical terms, but avoid 'sodium' and 'iron', comparisons and pronouns in general, especially 'we'. They are more frequently inaudible. This profile suggests an orientation towards facts, particularly anatomical terms.

4.7.4 A comparison of divergers – the social comfort, cooperative, abstract learner

4.7.4.1 word frequency category

*Table 124 Word frequency, terms **over**-used by divergent types as opposed to other types*

Word	Divergent frequency	Relative frequency	Accom/Con/ ass frequency	Relative frequency	Log Likelihood	Bayes
Iron	63	0.38	14	0.02	133.36	122.05
Death	11	0.07	6	0.01	15.59	4.28

Divergers mention iron significantly more frequently than other types, also death.

*Table 125 Word frequency, terms **under**-used by divergent types as opposed to other types*

Word	Divergent frequency	Relative frequency	Accom/ Con/ Ass/ frequency	Relative frequency	Log Likelihood	Bayes
Oedema	19	0.11	246	0.38	36.14	24.84
Pressure	19	0.11	199	0.31	22.36	11.06
inaudible	24	0.14	183	0.28	11.40	0.10

Divergent types mention the specific, relevant technical terms, oedema and pressure significantly less frequently than other types. This type is also significantly less likely to be judged inaudible than other types.

4.7.4.2 POS analysis

Table 126 Parts of Speech; terms under-used by divergent types as opposed to other types

POS	POS	Divergent	Relative frequency	Rest	Relative frequency	Log Likelihood	Bayes
Can, will	VM	309	1.87	1579	2.45	20.46	9.16

Divergent learners significantly under-use modal auxiliaries expressing can, will or would, in comparison.

4.7.4.3 Semantic analysis

*Table 127 Semantic categories: terms **over**-used by divergent types as opposed to other types*

Semantic	Code	Diverger Frequency	Relative frequency	Rest Frequency	Relative Frequency	Log Likelihood	Bayes
Solid materials	O1.1	79	0.48	72	0.11	74.70	63.39
Care	B4	27	0.16	33	0.05	18.23	6.92
States	X1	8	0.05	3	0.00	13.87	2.57

Divergers significantly over-use terms related to solids, (iron and sodium), care, (of the body and the home) and considering the psychological state.

*Table 128 Semantic categories: terms **under**-used by divergent types as opposed to other types*

Semantic	Code	Diverger Frequency	Relative frequency	Rest Frequency	Relative Frequency	Log Likelihood	Bayes
Weight: pressure	N3.5	21	0.13	232	0.36	28.09	16.69
Likely	A7+	272	1.64	1443	2.24	23.65	12.35
Liquid	O1.2	23	0.14	225	0.35	22.79	11.49

Divergers under use terms related to pressure, probability, or liquid (blood or water): all closely related to the detailed discussions of cardio vascular processes.

4.7.5 Summary of Diverger type

Divergers over-use iron and sodium and terms related to care of the body and the home and the psychological state. Notably, they are also the only group to mention death in connection with a heart failure scenario. They under-use highly relevant technical terms, oedema and pressure and the semantic categories related to pressure, likelihood and blood and water, and also forms of the verb 'can' and 'will'. They are less frequently inaudible.

4.7.6 Chapter Summary

Corpus analysis has been successful in demonstrating significant difference between all speech corpora when sub-divided according to gender, personality and learning style and performance. In many cases, the significant differences were in line with expectations from the literature, as will be further considered in the discussion chapter. Evidence for differentiation between speakers on the grounds of learning style alone was less clear than that for gender, personality FFM and performance.

The following chapter will present descriptive statistics for key variables for the study, before presenting the significant models identified through linear and multiple regression analysis.

5 Statistical Results

5.1 Introduction to Chapter

This chapter will present the descriptive and inferential statistical results of the study. The first section will describe the frequency and distribution of key data points and will illustrate with appropriate graphical methods where necessary. The appropriate data assumptions and tests which apply for performing linear regression analysis will be discussed. Finally, the linear regression analysis results will present the models and relationships of significant association which were demonstrated between key variables, and the interpretations which may reasonably be based upon them. This will lead to the final discussion chapter which will consider the findings in the context of the study, the literature and future research.

5.1.1.1 A note on missing data

Three subjects did not provide personality type data. In total, eight subjects did not provide learning style data items. One subject provided neither. On statistical advice, the missing data items were not substituted, but are excluded from relevant analyses. In the following descriptive and inferential statistics, where data items are missing, this is shown.

5.1.1.2 A note on data types

Data was analysed using the IBM SPSS 22 package. Some data was graphically presented using Microsoft Excel, 2016. The methods chapter presented the types of variables, in considering the results, these variables and the statistical procedures which need to be performed determine how the data is usefully described and secured for analysis. For categorical data, it is sufficient to describe the distribution. It is not meaningful to perform calculations on such data but a proportion or percentage figure can be reported. For metrical data it is usual to present the

results of frequency and distribution in a table, and to identify whether the data conforms to normal distribution in order to determine suitable statistical methods. However, dichotomous variables, where there are only two choices, are considered a special case of metric variables. Gender, and the FFM data are dichotomous, in that there are only two options possible. The Kolb learning styles is a four way variable but there is no undisputed order for them to occur, so they are essentially nominal. The interview grade does have an implied order of merit, and so is treated in the same way as the other test result scores, as it is rarely necessary to distinguish between them for statistical testing purposes.

5.2 Descriptive Statistics

This section sets out the descriptive statistics for the data points which were collected in order to address the research questions associated with individual difference through PBL discussions.

5.2.1 Gender distribution

Although each PBL group is arranged so that there is a gender balance, overall, for the PBL groups participating in this sample, the gender balance within the six PBL group sample, was split 60% male and 40% female.

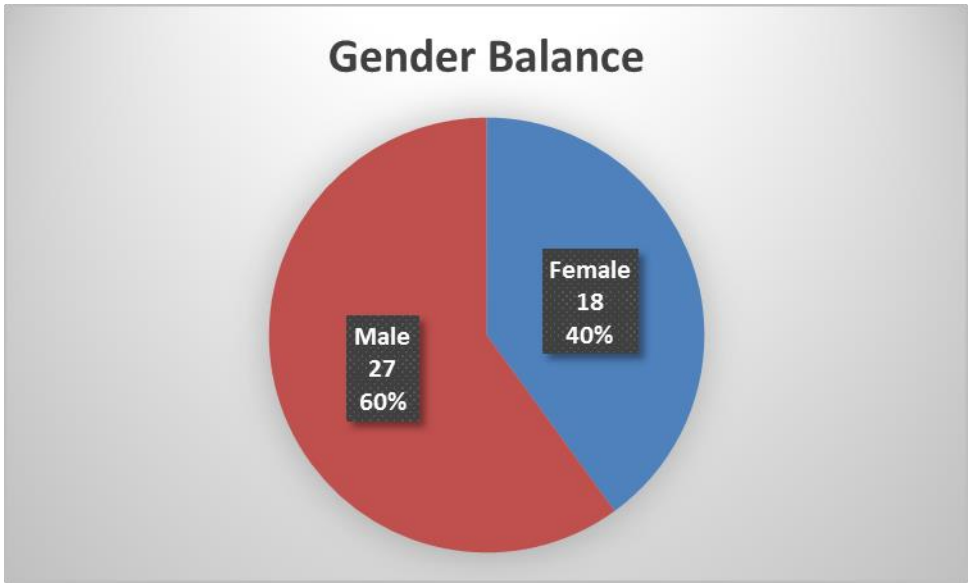


Figure 27 Gender Balance for PBL student sample

5.2.2 Distribution of FFM Personality measures

The balance for all participants between introversion and extraversion was fairly equal, when accounting for the three cases where the measure was not known.

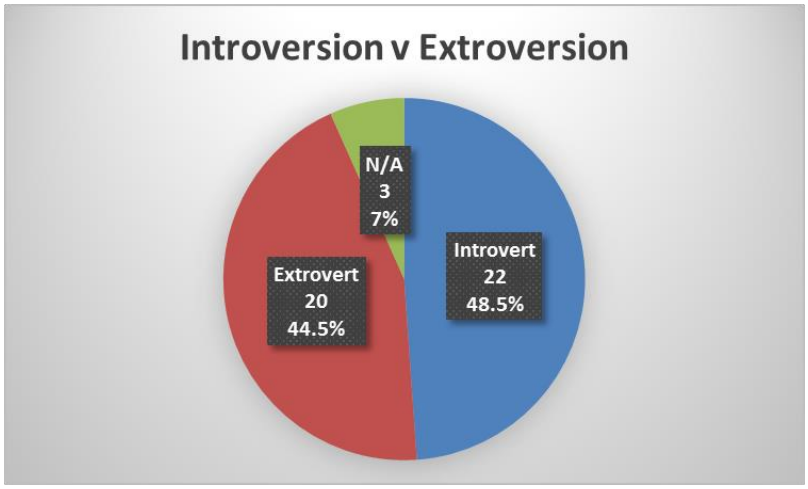


Figure 28 Introversion v extraversion for PBL student sample

The distribution between those who were noted as personable versus those who were logical was also equally distributed, where known.

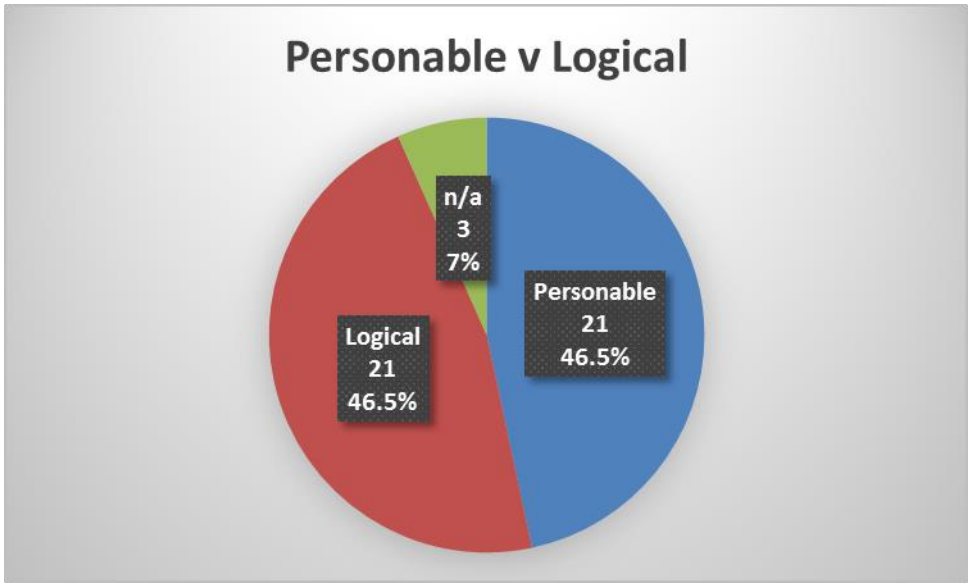


Figure 29 Personable v logical (agreeableness) for PBL student sample

However, a marked difference appears when comparing subjects judged to be adaptable, in the 75% majority, as opposed to 18% whose type was structured.

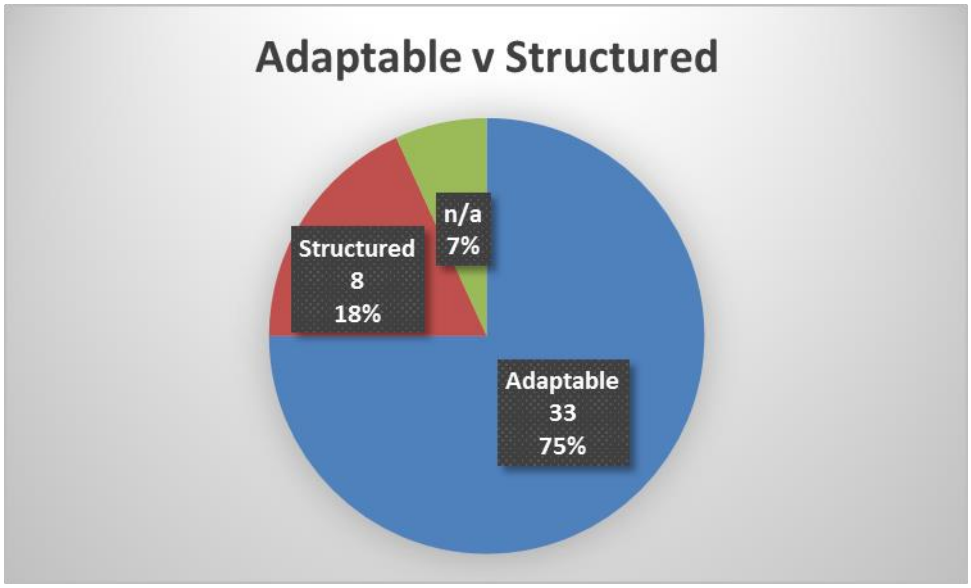


Figure 30 structured v adaptable (conscientiousness) for PBL student sample

A further marked split occurs when considering the proportion of subjects who were judged to be theoretically orientated, 60% as opposed to being grounded 33%.

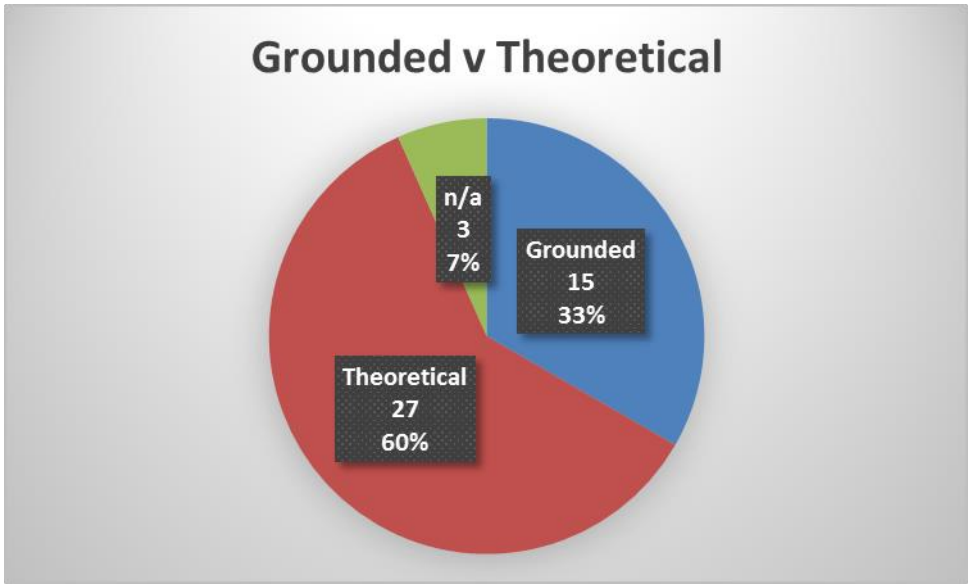


Figure 31 grounded v theoretical (openness to experience) for PBL student sample

The distribution of Calm as opposed to volatile, was equally distributed within the group, where the value was known.

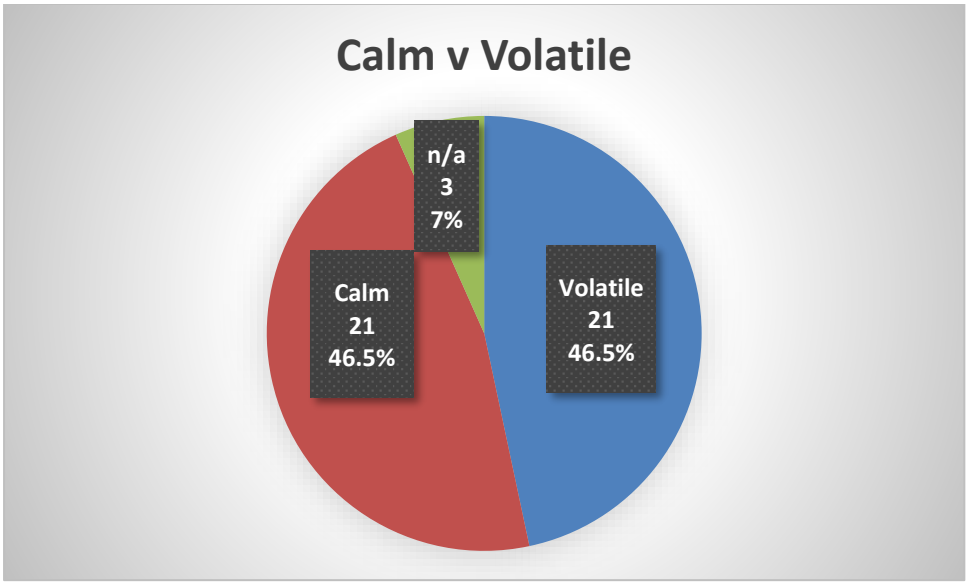


Figure 32 Calm and Volatile frequency (neurosis) for the PBL student sample

5.2.3 Quintax Temperament Groupings

When these types are combined into four dominant temperament groupings: either adaptable & grounded, Structured & grounded, Logical & theoretical or personable and theoretical, there was a marked lack of structured and grounded temperaments, given the low prevalence of the structured type among the sample. For three subjects, the temperament type was unknown.

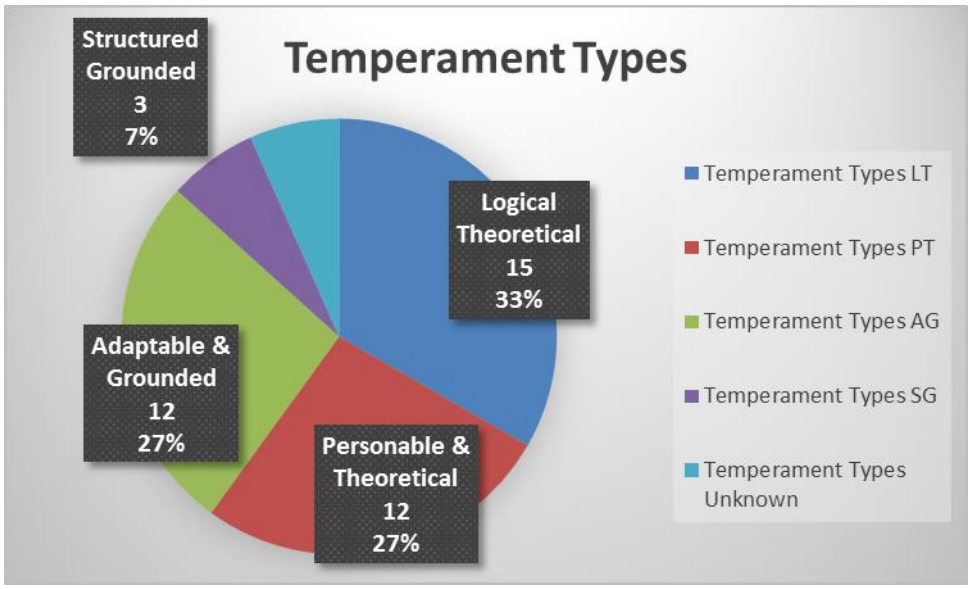


Figure 33 Quintax Temperament type frequency for PBL student sample

5.2.4 Kolb Learning Styles

When the subjects are described according to their dominant Kolb learning style, where known, there is a marked preference (54% of known types) for assimilating and converging learning styles.

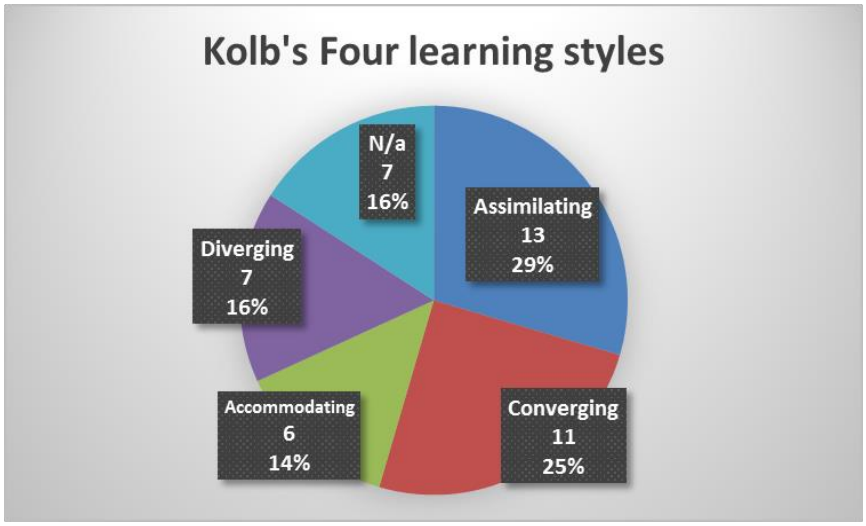


Figure 34 Kolb's Four learning styles for PBL student sample

5.2.5 Measures of Performance

It is meaningful to consider the measures of performance and percentage measures of test scores and measures of speech behaviour according to the usual tests for numerical data. Below is a table giving descriptive statistical measures of central tendency. Full data is available for each of these measures for 45 subjects. The following table sets out the standard descriptive statistics for these eight variables.

Table 129 Basic descriptive statistics for dispersion of key continuous variables

		Pre Entry		End year One		Speech		Behaviour Measures	
Variables		**Int Grade**	**GAMSAT**	**PPD grade**	**MCQ**	**Pcent totspeech**	**Pcent totturn**	**Pcentsent quest**	**Simplicity index**
N	Valid	45	45	45	45	45	45	45	45
	Missing	0	0	0	0	0	0	0	0
Mean		1.8	66.7	66.0	76.3	13.3	13.3	27.4	77.1
Median		2.0	65.0	63.8	76.0	12.8	13.2	23.6	78.0
Mode		2	63	59.8	69.5	2.43	10.4	9.8	78.0
Std. Deviation		.6	5.1	9.8	7.5	5.8	4.4	10.7	6.4
Variance		.4	26.8	97.9	57.1	34.6	19.5	115.4	42.1
Skewness		.1	**.9**	.1	.0	**.8**	**.7**	.6	-.1
Std. Error of Skewness		.3	.3	.3	.3	.3	.3	.3	.3
Kurtosis		-.7	-.0	-.8	-.6	1.6	2.0	-.4	-.3
Std. Error of Kurtosis		.6	.6	.6	.6	.6	.6	.6	.6
Range		2	20	39.6	28.5	29.5	24.3	43.7	28.8
Minimum		1	61	47.8	62.0	2.4	4.2	9.8	62.2
Maximum		1	81	87.4	90.5	31.9	28.6	53.5	91.0
Percentiles	25		62.5	59.3	70.2	9.1	10.2	20.1	71.9
	50		65.0	63.8	76.0	12.8	13.2	23.6	78.0
	75		70.0	75.0	81.5	16.1	15.8	35.5	81.6

Bold text is used to identify variables not found to be normally distributed.

5.2.5.1 A note on relevant statistical tests for the inferential statistical analyses in this study

The purpose of describing variables is in order to secure them and prepare them for the type of analysis which they will undergo. In this case, this is simple linear and multiple linear regression analysis. For this analysis, particularly where most variables are dichotomous, normality is not a crucial qualifying factor and so box plots, histograms and q plots, although produced, are not presented. Statisticians have discussed the issue of normality tests in relation to linear and multiple linear regression at length, and the lasting opinion appears to be that special tests apply to the normality of the residuals, which will be discussed below. Thus data does not need to be normally distributed in order to apply linear regression (Cook 1982).

5.2.6 Inferential statistics

Further tests are necessary in order to establish the suitability for linear and multiple linear regression on the data before the statistical tests to explore any predictive value and associations can be presented. However, a linear relationship is assumed with dichotomous variables, which applies to all FFM and gender. Linearity signals the strength of association, however, it is still possible to demonstrate association without linearity being present (Abrams 2007). Overall the aim of the statistical processes for this study is to determine whether differences in various independent variables, such as personality, can help to explain differences in dependent variables, such as the MCQ knowledge test score. This approach is common in theory building where there is a proposal that several independent variables may have influence (Laerd 2016).

Simple and multiple linear regression can be used to help to quantify the degree to which the independent variable explains differences in the dependent variable. For example, a simple linear regression can be used to determine the relationship

between the interview grade as an independent variable (i.v.), allocated on entry, and the performance in the PPD test at the close of year one, the dependent variable. Both the interview and the PPD test aims to measure the development of similar personal and professional clinical competences and so, if valid and reliable, are likely to be closely related.

Multiple regression is used when several independent variables, or predictor variables are thought to be influential on one dependent variable. For example, personality measures of neuroticism and conscientiousness might well be associated with scores on knowledge tests such as the GAMSAT on entry, and the MCQ at the close of year one (Lievens, Coessier et al. 2002, Ferguson, James et al. 2003). The measured scores can become the dependent variable, and personal attributes, the independent, or predictor variables.

As discussed in the Methods chapter, naturally, this method would only ever be able to identify predictor and independent variables which together might provide a useful estimate of a participant's likely achievement or success. It could never identify causality. It is subject to much noise or error, in common with all measures of human behaviour where a combination of factors is always likely to operate in concert. Some authors have suggested reasonable reliability can be mathematically modelled at five (Vittinghoff and McCulloch 2007), or even at two subjects per variable, (Austin and Steyerberg 2015) and in general, larger numbers of subjects provide more robust linear regression analysis. However, for this study, 45 participants were available and as the most extensive multiple regression model computed, related four independent variables, (two of which were dichotomous) in relation to the dependent variable, this suggests the minimum of subjects per variable was comfortably exceeded.

Model analysis identified which variables were likely to be accounting for variance and analysis was then restricted to these.

The predictor variables were all di- or trichotomous and were:

- Calmness or volatility
- Adaptable or structured
- Grounded or theoretical
- Gender
- Interview grade
    - (1 for the top group, 2 for the middle and, 3 for the lowest group).

The dependent variables were all percentage measures and were:

- MCQ percentage score
- GAMSAT % score
- PPD % score
- % of total speech
- % of total turns taken
- % of sentences that were questions
- Flesch measure of complexity of utterance – or a simplicity index

5.2.6.1 Qualifying tests for the use of linear and multiple linear regression

In each of the models where significant relationships are demonstrated, scatterplots were used to demonstrate that a linear relationship existed between the variables. Scatterplots of residuals also demonstrated that there was homoscedasticity and normality of the residuals, meaning that the error levels were acceptable to qualify

for multiple regression analysis. For each model where an association is demonstrated, the PP Plot will be included in order to visually demonstrate the linear relationship between the predictor and the independent variable. This satisfies one of the key assumptions on which linear regression is based.

5.2.7 Interpretation of results

The beta value measures how strongly each independent or predictor variable influences the dependent variable. The higher the beta value the greater the impact, measured in standard deviations. The strength of the correlation is generally considered weak when below 0.3; moderate above 0.5 and strong above 0.7. A positive result signifies an increase in the dependent variable, a negative score, a decrease.

R is a measure of the correlation between the observed value and the predicted value; in this study, the extent to which the predictor variable accounted for the measure. Adjusted R squared is usually lower as it takes account of the number of variables and reports a more reliable figure. It will be noted that where the adjusted r squared is low, this shows that the independent or predictor variable can only account for small proportions of the variance. However, it is common for all models concerned with predicting human behaviour to give low r squared scores; frequently below 25% or 50%. It must be judged together with the F test value with significance quoted as a p value, with significance occurring below p=0.05.

For the multiple linear regression model, it is important to demonstrate that multiple collinearity is not an issue, in other words, that there is overlap between the variables which causes concern for the model's power in explaining variance. The collinearity figures are presented in this model.

In each case the 'enter method' of linear regression was used. This means that the independent variables are entered into the linear regression in one step, rather than using forward or backwards steps to manipulate the variables. This is the simplest method of identifying where associations may exist and is useful when a limited number of variables have been identified which are likely to have an influence on the dependent variable.

The significant associations will now be presented as 8 models.

5.3 Associations between entry measures (dependent variables)

5.3.1 Model 1: PPD and interview grade

Inspection of the scatterplot identified a linear relationship between the PPD score and the interview grade, as demonstrated in the PP Plot below.

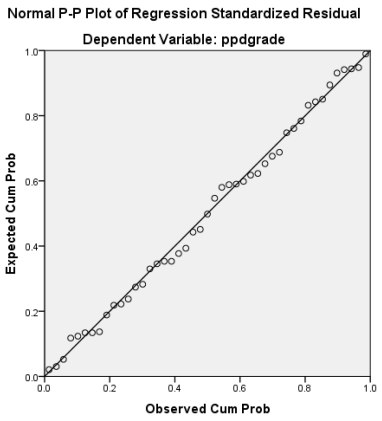


Figure 32 PPD scores were suitable for performing linear regression

Using the enter method, a significant model predicting performance in the end of year PPD test from the interview grade emerged (F= 5.088, p <0.029). Adjusted R square = 0.085).

Overall, the predictor variable, the interview grade, accounted for **8.5% of variance in PPD scores.**

Predictor (i.v.) Variable	Beta	p
Interview grade	**-.325**	**<0.029**

A weak to moderate (-0.325) but significant negative relationship is demonstrated between the interview grade and performance on the PPD test, suggesting that

allocation to a higher interview group on entry is associated with, or weakly predictive of a higher score on the PPD test. The highest grade of interview group is 1, and so a negative Beta value demonstrates that the lower value is associated with the higher PPD score.

5.3.2 Model 2. GAMSAT and interview grade

Inspection of the scatterplot identified a linear relationship between the variables as demonstrated in the PP Plot below.

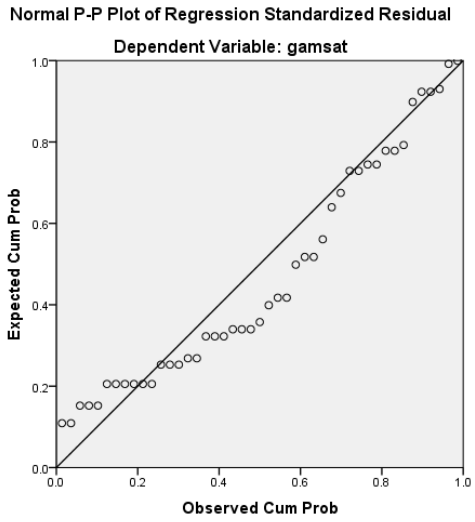


Figure 35 GAMSAT scores were suitable for performing linear regression

Using the enter method, a significant model predicting performance in the GAMSAT from the interview grade emerged (F= 6.552, p <0.014).

Adjusted R square = 0.112, showing that the interview grade accounted for **11% of variance in GAMSAT scores.**

Predictor (i.v.) Variable	Beta	p
Interview grade	**.364**	**<0.014**

A weak to moderate (0.364) but significant positive relationship is suggested between the interview grade and performance on the GAMSAT test, suggesting that allocation to a lower interview group on entry is associated with, or weakly predictive of a higher score on the GAMSAT test. The highest grade of interview group is 1, and so a positive Beta value demonstrates that the higher interview groups i.e. the poorer performing groups, are associated with the higher GAMSAT score.

5.3.3 Model 3: GAMSAT and Flesch, (simplicity index measure)

Inspection of the scatterplot identified a linear relationship between the variables as demonstrated in the PP Plot below.

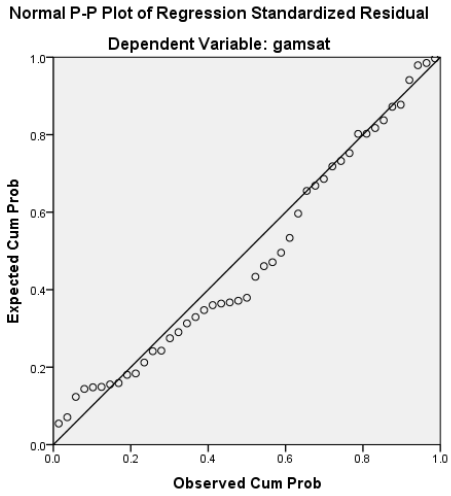


Figure 36 GAMSAT scores were suitable for performing linear regression

Using the enter method, a significant model emerged predicting performance in the GAMSAT from the simplicity index measure (Flesch) (F= 5.272, p <0.027). Adjusted R square = 0.088.

Overall, the predictor variable, simplicity index, in this model accounted for **8.8% of variance in GAMSAT scores.**

Predictor Variable	**Beta**	**p**
Flesch simplicity	.330	<0.027

A weak to moderate but significant relationship is suggested between the GAMSAT grade and the Flesch simplicity measure of speech, suggesting that a higher score

on the GAMSAT test is associated with lower complexity of utterance in speech, as the higher the Flesch score, the simpler the utterance, according to the measure and the beta value is positive.

5.4 Knowledge test at year end as dependent variable

5.4.1 Model 4: Multiple Linear Regression Analysis Predicting performance through scores on the MCQ knowledge test

Inspection of the scatterplot identified a linear relationship between the variables as demonstrated in the PP Plot below.

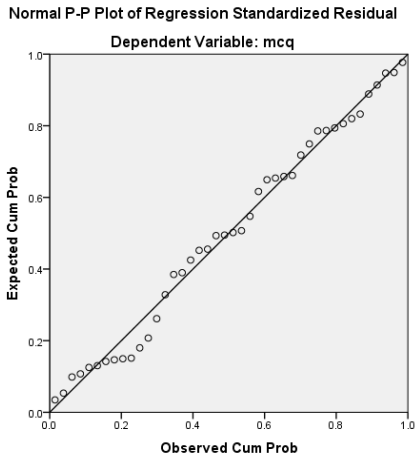


Figure 37 MCQ scores were suitable for performing multiple linear regression

For multiple linear regression analysis a further test was performed, the Durbin Watson test, which demonstrated satisfactory independence of the residual errors, with a Durbin Watson statistic of 2.309. The Durbin-Watson statistic is always between 0 and 4. A value of 2 means that there is no autocorrelation in the sample.

Table 130 showing acceptable Collinearity Statistics for multiple regression analysis model 4

Tolerance	VIF
.985	1.01
.876	1.141
.990	1.010
.874	1.144

A further test of collinearity statistics was performed, shown in the table above. Where tolerance is below 0.2 and VIF is shown to be greater than five, it is considered problematic for collinearity, i.e. there is too much overlap (Belsley, Kuh et al. 2005). No measure demonstrates that issue for model 4, multiple regression analysis supporting the valid consideration of the model.

Using the enter method, a significant model predicting performance in MCQ knowledge tests emerged (F= 8.700, p <0.0005). Adjusted R square = 0.412.

Overall, the predictor variables in this model accounted for **41% of variance in MCQ scores.**

Significant i.v. predictor variables are shown below, with their respective contribution to variance in MCQ scores.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	34.336	12.607		2.723	.010
	GAMSAT	.690	.169	.473	4.074	.000
	calmvol	-4.596	1.817	-.307	-2.529	.015
	adapstruct	7.982	2.183	.427	3.657	.001
	groundtheo	-4.229	1.864	-.274	-2.268	.029

a. Dependent Variable: MCQ

Figure 38 SPSS Coefficients table for Model 4, multiple linear regression predicting MCQ scores

Predictor Variable	**Beta**	**p**
GAMSAT	.473	<0.0005
Calm/vol	-.307	<0.015
Adapt/struct	.427	<0.001
Grounded/theo	-.274	<0.029

This demonstrates a moderately strong relationship to a highly significant degree, between the MCQ and the GAMSAT score on entry. The positive beta value means that the higher the GAMSAT score the higher the MCQ score will be predicted to be, more than is accounted for by chance.

A weak negative but significant relationship is suggested between the degree of volatility and performance, suggesting that higher volatility is associated with lower scores on the MCQ test.

There appears to be a moderate, positive and significant relationship between the adaptable and structured dimension and MCQ scores. This suggests that the structured (high conscientious) personality type is associated with a significantly higher MCQ scores than by chance. This accords with many results associating higher levels of conscientiousness with higher performance on tests and in workplace measures.

Similarly, a weak significant negative relationship is suggested between the grounded and theoretical personality type: theoretical types being associated with lower performance on the MCQ test.

Measures of complexity, Flesch simplicity index

5.4.2 Model 5: Percentage of total speech and complexity

Inspection of the scatterplot identified a linear relationship between the percentage of total student speech, against the Flesch simplicity index, as demonstrated in the PP Plot below.

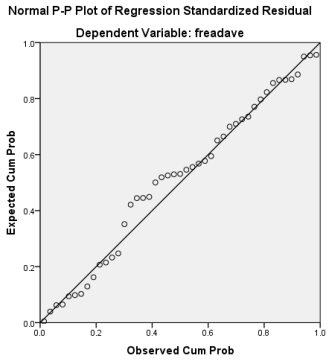


Figure 39 Simplicity index – Flesch complexity measures were suitable for performing linear regression

Using the enter method, a significant model predicting performance measured by Flesch from the total amount of speech emerged (F= 14.897, p <0.0005). Adjusted R square = 0.240. Overall, the predictor variable, the percentage of total speech, in this model accounted for **24% of variance in Flesch scores.**

Significant predictor (i.v.) variables are shown below:

Predictor Variable	**Beta**	**p**
Percentage of total speech	-.507	<0.0005

A moderate, significant negative relationship is suggested between the proportion of total speech and the Flesch simplicity index, suggesting that as the proportion of total speech increases, the Flesch simplicity score decreases, i.e. becomes more

complicated. This suggests that speaking more is associated with a more complex sentence structure and vocabulary. Thus, contributing a lower proportion of speech is associated with simpler speech utterances.

5.4.3 Model 6: Flesch and proportion of questions in speech

Inspection of the scatterplot identified a linear relationship between the percentage of total student speech, against the Flesch simplicity index score, as demonstrated in the PP Plot below.

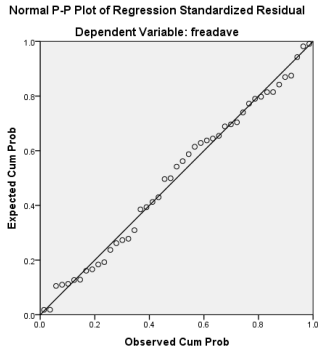


Figure 40 Simplicity index – Flesch complexity measures were suitable for performing linear regression

Using the enter method, a significant model predicting performance measured by Fleisch from the total amount of speech emerged (F= 4.801, p <0.034). Adjusted R square = 0.08 meaning that the percentage of total speech which are questions, in this model accounted for **8% of variance in Flesch scores.**

Significant variables are shown below:

Predictor (i.v.) Variable	**Beta**	**p**
Percentage of speech questions	.317	<0.034

A weak to moderate and significant positive relationship is suggested between the percentage of total sentences which are questions and the Flesch simplicity index, suggesting that a higher score on the Flesch test is associated with a higher

proportion of questions. Questions tend to have simpler sentence structures, and be shorter, and so are more likely to give a high score by the Flesch measure.

If Model 5 suggests that speaking more is likely to be associated with a more complex sentence structure and vocabulary, Model 6 reinforces this finding as it suggests that more questions are associated with fewer contributions and simpler sentence structures, such as tend to be found in questions.

5.5 Total Speech

5.5.1 Model 7: Total proportion of speech and gender

Inspection of the scatterplot identified a linear relationship between the % of total student speech, against the gender of speaker, as demonstrated in the PP Plot below.

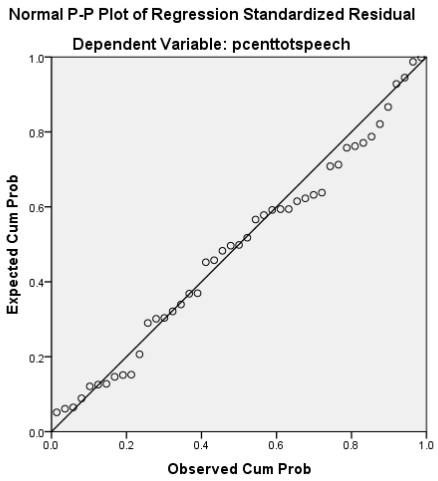


Figure 41 Simplicity index –percentage total speech and gender were suitable for performing linear regression

Using the enter method, a significant model predicting the total proportion of speech emerged (F= 5.631, p <0.022). Adjusted R square = 0.95. Overall, the predictor variable, the gender**, in this model accounted for 95% of variance in the total amount of speech.**

Significant variables are shown below:

Predictor Variable	**Beta**	**p**
Gender	-.340	<0.022

A weak to moderate but significant relationship is suggested between gender and the total proportion of speech. As females are coded as 1, and the slope is negative, this suggests that the male student (coded 2) is significantly more likely to present their views in group discussion when compared to the female student and that this accounted for 95% of the variance within the model. Male speech behaviour in groups is associated with more competitive and assertive behaviour, which is likely to lead to more speech in groups.

5.5.2 Model 8: Total proportion of speech turns taken and introversion or extraversion

Inspection of the scatterplot identified a linear relationship between the percentage of total student turns taken to speak and the introversion/extraversion of the speaker, as demonstrated in the PP Plot below.

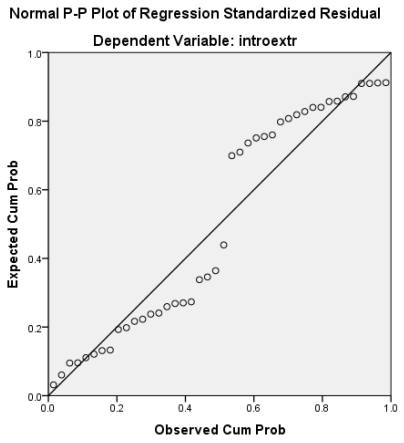


Figure 42 Introversion/extraversion and total proportion of turns to speak measures were suitable for performing linear regression

Using the enter method, a significant model predicting the total proportion of turns taken to speak emerged (F=5.835, p <0.020). Adjusted R square = 0.105. Overall, the predictor variable, the degree of extraversion, in this model accoun**ted for 10.5% of variance in the percentage of turns taken to speak.**

Significant variables are shown below:

Predictor Variable	**Beta**	**p**
Percentage total turns speaking	.357	<0.020

A weak to moderate but significant relationship is suggested between the degree of extraversion and the total proportion of speech. As introverts are coded as 1, and the slope is positive, this suggests that the extravert student (coded 2) is significantly more likely to take a turn to speak in group discussion when compared to the introvert student. Extravert speech behaviour in groups is associated with higher 'surgency' which makes it more likely that they would take a turn to speak. However, in this study it is not associated with a statistically significant difference with the total speech contributed, which suggests that extraverts make more frequent but shorter speech contributions.

5.5.3 Summary of the statistical results chapter

This chapter has presented descriptive and inferential statistical results of the study. The final discussion chapter will discuss these findings in the context of the whole study, the literature and future research.

6 Discussion

6.1 Introduction

This chapter will review the key findings of the study in the light of the research questions posed at the close of the literature review chapter. The discussion chapter begins with a table giving a summary of these questions. linking the question with the evidence stated in the corpus and statistical results chapters.

The discussion will then go on to discuss and evaluate which statements may be supported or rejected on the basis of the overall results of the study under the broad headings of gender differences, performance and personality differences. The discussion will consider the statistical findings, labelled as Models 1 – 8, and corpus analysis findings, in the light of existing work in the field.

Finally, a summary of conclusions will suggest how these findings may be applied within learning through PBL discussion and other discussion-based forms of learning. The chapter will close with a brief re-statement of the limitations of the study and suggest scope for future research in this field.

Table 131 Summarising table of results in relation to the research questions posed

Research Question	Method	Multiple or linear Results model	Conclusion
Gender differences			
Is there a difference in the amount of speech contributed by males or females?		Linear regression analysis model 7	Yes. Males are significantly more likely to contribute a higher proportion of student speech. ($F=5.631$, $p<0.022$, adjusted r square = 0.95, Beta -.340).
Is there a difference in the number of questions asked by males or females?	Multiple linear regression analysis		No. There is no statistically significant evidence for a difference in the number of questions asked according to gender.
Can corpus analysis demonstrate significant differences between the male and female corpora in line with previous studies?	Corpus analysis of male v female corpus		Yes. There is substantial evidence of a number of statistically significant differences in language use between the genders, which may impact on future clinical practice.
Performance and group differences			
Is there a statistically significant relationship identified between different PBL groups and PPD and MCQ measures for this cohort?	Multiple linear regression analysis		No. The PBL group was not shown to have any statistically significant relationship with PPD and MCQ measures.

Research Question	Method	Multiple or linear Results model	Conclusion
Is there a connection between interview grade and PPD score at the end of the first year?		Linear regression analysis model 1	Yes. High scores in the interview are significantly more likely to be associated with high PPD scores. (F=5.088, p<0.029, adjusted r square = 0.085, Beta -.325).
Is there a relationship between the interview grade and performance in GAMSAT pre-entry test?		Linear regression analysis model 2	Yes. High scores in GAMSAT are associated with allocation to a lower scoring interview group (F=6.552, p<0.014, adjusted r square = 0.112, Beta .364).
Is there a connection between speaking more words overall and the score students received on interview, in the GAMSAT; in their MCQ exam; in their professional development grade?	Multiple linear regression analysis		No. There is no statistically significant evidence between the proportion of student speech and higher scores in any test.
Is there a connection between the Flesch simplicity index (complexity measure) when applied to speech, use of questions and performance in knowledge tests?	Multiple linear regression analysis	Multiple Linear regression analysis model 3, model 5, model 6.	Yes. There is statistically significant evidence that a higher score on the GAMSAT test is associated with a lower complexity score in speech. However, more simple sentence structures were also associated with a lower proportion of total student speech and a higher proportion of sentences that were questions.

			Model 3: GAMSAT and Flesch: (F=5.2721, p<0.027, adjusted r square = 0.088, Beta .330). Model 5: Proportion of total speech and Flesch (F=14.897, p<0.0005, adjusted r square = 0.240, Beta -.507). Model 6: Proportion of questions and Flesch (F=4.801, p<0.034, adjusted r square = 0.08, Beta -.317).
Is there a difference in the frequency or type of questions asked or terminology used by higher and lower scoring learners?	Corpus analysis of high low scorer corpora		Yes. There is statistically significant evidence that lower scorers continue to ask 'why' questions and repeat simpler technical terms more frequently than higher scorers, who ask more 'how' and 'when' questions and use more advanced technical terms.
Is there a connection between personality type and performance in knowledge tests in year one?	Multiple linear regression analysis	Multiple Linear regression analysis model 4	Yes. There is statistically significant evidence of a predictive relationship between three dimensions of the FFM and both knowledge tests. In particular, higher scores in the GAMSAT test, lower neuroticism (volatility), higher conscientiousness (structuredness) and lower openness to experience (groundedness) were associated with higher scores in the MCQ test. Model 4: MCQ knowledge test and GAMSAT and personality factors (F= 8.700, p <0.0005, adjusted r square = 0.412)

Research Question	Method	Multiple or linear Results model	Conclusion
Personality type, learning styles and speech behaviour			
Is there a correlation between extraversion and the number of words, or number of utterances (turns taken) which a person makes during a group session?		Linear regression analysis model 8	Yes. Extravert speakers are significantly more likely to contribute a higher proportion of student speech turns, although they do not contribute a higher percentage of speech overall and so must be speaking more frequently at shorter length than introverts. Model 8: Turns taken to speak and introversion/extraversion ($F= 5.835$, $p <0.020$, adjusted r square = 0.105)
Can corpus analysis demonstrate significant differences in student speech between personality and learning style corpora created according to the FFM, quintax temperament and Kolb learning inventory measures?	Corpus analysis		Yes. For all five dimensions of the FFM there was substantial evidence of statistically significant differences in language use, in terms of frequency, parts of speech and semantic categories. For the KLSI and quintax temperaments, the evidence was less coherent, but could be interpreted as aligning to the category descriptors at times.

6.2 General Comment on Corpus Analysis

The corpus analysis data shows that when compared to a similar educational spoken corpus, (BNC 2007), PBL student speech substantially and significantly over-uses speech related to the semantic categories of anatomy, disease, cause and effect, medicines, change and quantity. For example, the occurrence of the category relating to anatomy in the PBL corpus carries a log likelihood (LL) = 2543.18, when LL=15.13 this is considered equivalent to $p<0.0001$ in terms of whether it could occur by chance alone. There was significantly higher use of subordinating conjunctions. Question words such as 'what', 'why' etc. over-occurred with a log likelihood of LL=146, (close to $p<00001$) in comparison to a similar educational spoken corpus, demonstrating an objective measure indicating the strength of PBL as an inquiry-driven form of learning.

In addition, all analyses comparing the sub-corpora with each other, identified a rich assortment of highly statistically significant differences in language use, which were broadly in keeping with existing work on personality and gendered speech.

Analysis of learning styles corpora were perhaps less compelling, in keeping with the current research doubt on their coherence (Coffield, Moseley et al. 2004). Temperament type corpora were also less clear in delineating what type of student may be at risk of doing poorly through observing their speech behaviour in PBL discussions alone.

Taken together, it is possible to identify speech behaviours which may be associated with students at risk of being less effective, without testing the personality, temperament or learning style. These will be summarised in the conclusion to the discussion (section 6.6).

6.3 Speech behaviour and gender

6.3.1 Gender and speech participation

Model 7 (Section 5.5.1) demonstrated that in this study, males were significantly more likely to contribute a higher proportion of student speech and that this factor accounted for 95% of variance in relation to gender for the proportion of total student speech. For many studies which were considered in the literature review, the gender of student participants was not reported or featured among the research questions. However, where gender was known and considered, there was strong evidence of gender difference. In a key meta-analytical study, female clinicians took longer to complete consultations and demonstrated a much wider range of communication strategies (Roter, Hall et al. 2002). However, this was associated with much higher use of open questions and a participatory style and so the distribution of speech was likely to be more client focussed and so not evidence of the female clinicians speaking more. The higher measures for the male clinicians, under the category for information giving, suggests that male clinicians spoke more, in order to give this information. Duek found that women tended to speak more when the facilitator was absent, and noted that most facilitators in the study were male (Duek 2000).

Where the proportion of speech contributed by gender has been studied, the results consistently show that males tend to speak for a higher proportion of the time than females across many cultural and professional settings (Tannen 1993, Tannen 2002, Schiffrin, Tannen et al. 2008). There is also consistent evidence that the genders are judged differently with respect to how much they speak: women being expected to speak less and being considered by both genders to have spoken more than males even when this is objectively not the case (Mast 2002, Brescoll 2011). In general, studies show that the more powerful the man, the more he speaks. Even in professional settings, where women are trained and expected to contribute

by virtue of their role, women still appear to be judged differently; with lower participation in speech than males being demonstrated (Brescoll 2011). It is possible that the transcription method may favour male contributions as opposed to female contributions and this will be discussed in the limitations section of this chapter. Evidence from this study appears to further support the findings of the literature review.

Some authors have argued that adapting to the demands of communicating between the genders is a specific communication skill and demonstrates a form of 'cultural awareness' (Tannen 2001, Schiffrin, Tannen et al. 2008) and that skilled communicators adapt conversations effectively within different 'cultures'. Gender was not a feature of the corpus study of GP qualifying exams, (Roberts 2014) but the ability to adapt similar phrases in response to the simulated patient's needs were associated with fewer misunderstandings and so with higher marks for the exam. Yet, there is no apparent acknowledgement of gender differences in speech in the Calgary Cambridge communication skills materials, currently recommended to healthcare professionals (Kurtz and Silverman 1996). Relatively recent articles providing an updated meta-analysis on the topic, continue to cite work from the last century (Jefferson, Bloor et al. 2013). There may be an apparent reluctance to discuss gendered communication, as discussions of general tendencies among groups in society run the risk of triggering misunderstanding and allegations of bias where it is assumed that differences are automatically being identified as 'better or worse' rather than simply different (Dow and Wood 2006) (Hickerson and Gastil 2008).

6.3.2 Gender and questions and agreement

Female speech in this sample was not found to reach levels of significance to indicate that they were more likely to use questions than males, ($F=3.241$, Beta .265, $p<0.079$, adjusted R square = 0.48), based on the question marks which were entered during transcription. Corpus analysis identified

remarkably similar frequency of use of question words when comparing the corpora. Other studies have identified a high proportion of female questions as 'tag' questions, adding a tag phrase to a statement such as 'isn't it?', where the function is probably seeking agreement for a statement, rather than an answer: a style consistently observed in female speech (Mulac and Bradac 1995). Evidence in this study did not support this to a level of statistical significance.

However, female speakers were significantly more likely to use pronouns (LL=74.93), make interjections (LL=43.72), especially 'yes', 'no' and okay', and to be concerned with positive evaluation generally (LL=29.6). This evidence supports a style of 'active listening', consensus building and awareness of individuals, as fellow learners and as the subject of discussion, which has often been associated with female speech (Baker 2014). Such behaviours support discussion as a mechanism for participative learning, as they encourage contributions.

6.3.3 Gender and specialist terminology

Males significantly over-used certain specific technical terms, 'pressure' (LL=29.54), measurements, (LL=25.09), and liquids (LL=24.74).

Differences in the use of technical terms between genders in this study supports theories proposing a general tendency for males to discuss technical systems and to over use nouns, whereas females over-used pronouns, 'I' and 'we' (Ruigrok, Salimi-Khorshidi et al. 2014) (Newman, Groom et al. 2008). In the study of Physics, there has been a consistent gender gap identified over the past thirty years; currently, 20% of Advanced Level Physics students in England are female (Physics 2017). This PBL scenario concerned blood pressure and it is likely that more males than females studied Physics prior to entry, and may be more fluent with technical systems related to this particular PBL topic. Whether learners had studied Physics was not available to this study.

There may be some reluctance by female medical students to use technical terms as there is strong evidence for females and female clinicians avoiding prolonged technical information giving in discussion (Roter, Hall et al. 2002) (Blanch, Hall et al. 2008). It may be that some female medical students might benefit from further, specific encouragement from facilitators within PBL to support them in developing fluency with giving highly technical explanations. It may suggest simply that this confidence may take some time to develop, especially as this study took place one semester into the first year of GEMS. The PPD module, is designed to foster fluency with the bio-psycho-social context of clinical practice and this may be of particular value to male students, as the focus is not on technical aspects.

There is strong evidence from this study for consistent differences in speech behaviour when learning through discussion for males and females. All PBL groups are actively gender balanced. This suggests that there is, in fact, an acknowledgement of gender difference. It is widely acknowledged that PBL provides opportunities to foster the development of communication skills. It is likely that gender differences are one aspect of the dynamic, complex PBL learning environment which may serve the future clinical professional in developing skills which will help in their future work with colleagues and diverse members of the public.

However, some differences, such as males tending to speak more and focus on technical terminology, and for females to speak less, and to focus less on technical terminology and more on agreement, may not currently be optimally addressed during the early clinical years. Males may continue to speak more often and interrupt female speakers and thereby avoid developing their skills in information gathering through use of open ended questions, and demonstrating empathy, through empathetic listening.

Females students may continue to speak less frequently and thereby avoid developing their skills in making confident, technical clinical contributions.

6.4 Performance and significant differences

6.4.1 No differences identified between PBL groups

There was no statistically significant relationship identified between different PBL groups and PPD or MCQ tests measured in this study. This means that in terms of the individual factors which were measured, groups were gender balanced, of the same size, and did not vary significantly in terms of the FFM, learning styles or test performance represented within each group.

This finding is interesting in the light of much concern in the literature about the specialist contribution of the facilitator from a knowledge perspective, and in terms of their skill as an effective facilitator and role model, both in face to face and online discussion environments. It may be that a larger scale study of student speech behaviour would be capable of identifying significant differences. It may be that as the focus of this study was not on the facilitator and their speech was excluded, crucial factors related to group performance were inadvertently excluded from the study. Where a study focuses tends to be where effects are observed, and this is a form of possible confirmation bias. It is possible that groups, who were not functioning to the satisfaction of their facilitator or learners, chose not to take part and thus skewed the study. Where significant associations were identified between factors, they accounted for relatively low proportions of variance, as might be expected when examining human behaviour and so the study may simply have failed to detect significant differences.

However, at face value, it is reassuring, as this finding may suggest that to some extent the study has been successful in focussing on individual difference whilst keeping other factors, such as time, place and scenario constant.

6.4.2 Statistical connection between interview grade and PPD score and interview grade and GAMSAT score

The interview grade allocated pre-entry was found to be statistically significantly associated with the high score on the PPD test at the close of year one (Model 1, Section 5.3.1). This suggests that the interview selection procedure has moderate alignment with the constructs which are tested through the PPD test.

However, model 2 (Section 5.3.2) identified a negative association between the GAMSAT score on entry and the interview group allocation. This suggests that those gaining a high GAMSAT knowledge test score, were more likely to be allocated to a lower interview group. This suggests that the GAMSAT knowledge test is not fully aligned with the constructs which are being assessed in the interview process. GAMSAT is significantly associated with the MCQ knowledge test (Model 4).

There have been several years of detailed psychometric analysis to monitor the relationship between the performance of selection methods and student performance as part of the process of establishing the new GEMS (Manning and Garrud 2009, Garrud 2011, Garrud and Yates 2012). The findings of this study support the view that the interview and GAMSAT processes are together identifying important factors which may both combine to be of importance in the future clinician.

6.4.3 No differences identified between speaking more words overall and the score on any test

There was no statistically significant relationship identified between the proportion of student speech contributed to discussion and to any other measure of attainment in this study. Simply put, saying more does not mean that an individual will learn more. This suggests that, contrary to Azer's encouragement (Azer 2009), attempts to motivate learners to

participate more frequently in PBL discussions purely through rewarding the quantity of speech may not be the most helpful route to enhancing the quality of learning through discussion, once a baseline for willing participation has been reached. This also supports the view that what a student may vocalise of their thoughts during discussions does not have a straight forward relationship with the quality of their thought, or the extent to which they or others are learning. It is the quality of the interaction between these factors, which are only partly verbalised, which seems likely to suggest determine the effectiveness of learning through PBL discussions. Learner participation clearly extends beyond the easier measures we can make of attendance and assessment scores on entry and exit or even 'time on task' (Schmidt and Moust 2000, Kamp, Dolmans et al. 2012, Ericsson 2016).

6.4.4 Performance in knowledge tests and simplicity of speech and use of questions

In Model 3 (section 5.3.3), Model 5 (section 5.4.2) and Model 6 (section 5.4.3) linear regression identified a cluster of significant associations related to the simplicity of statements to performance on the GAMSAT knowledge test, and to the proportion of sentences which were questions and to the proportion of overall speech. Higher knowledge scores as measured by GAMSAT, were in general identified with simpler sentence structures. In turn, a simpler sentence structure was associated with a lower proportion of contributions to the student speech overall. A higher proportion of sentences as questions was associated with a higher simplicity measure. The simplicity index formula is based on an average number of syllables and an average length of sentence. Sentences tend to use a simpler structure in terms of average length so this result is easy to account for.

However, questions have been considered as a possible indicator of clinical reasoning, and observed to fluctuate during PBL sessions (Da Silva and Dennick 2010, Da Silva 2014), so it is interesting that simpler sentence

structures were associated with speaking less overall. It seems likely that questioning may be linked with longer thought time and cognitive activity, both in posing the question, and then in listening to replies and so may somehow limit the overall contributions. Extraversion was found to be associated with taking more turns to speak (Model 8, section 5.5.2). For the simplicity measure to be linked with higher GAMSAT scores but not scores on other knowledge tests, particularly the MCQ, is surprising. Corpus analysis comparing the high and low scorers for GAMSAT, identified that low scorers asked more frequent 'why' questions, which was replicated when comparing high and low MCQ scorers.

The discussion will now consider corpus analysis results, in building supporting evidence for an emerging profile of the knowledgeable student being identified with speaking less frequently, but with more focus on technical content, particularly at the more advanced level, and asking a higher proportion of questions which probe how something is occuring, rather than seeking to describe or understand (why) at a more basic level.

6.4.5 Corpus analysis of the high and low scoring quartiles for each knowledge test demonstrated significant differences in frequency of technical terminology, types of questions

6.4.5.1 High and low scorers in the PPD test (section 4.4.3)

If PPD assessments are valid in targeting the awareness of health in its full bio-psycho-social context and assesses according to a model of reflective and professional practice, we might expect higher scorers to demonstrate a greater focus on the subject of the scenario ('she') and the progression of the disease and the medicines on this individual. Instead, high PPD scorers were significantly more frequently discussing the scientific and factual aspects of the case, (pressure and weight). This accords with the overall language hypothesis for which there is some evidence in this study that when technical language, context and terminology is unfamiliar, such as when one is mastering a new clinical vocabulary, it is used more frequently but perhaps with less skill. Once securely learned, the frequency reduces

and the apposite or more sophisticated and specific use increases (Bridge 1962, Roberts 2014). This may be occuring in the PPD curriculum as well as the more direct tests of clinical knowledge, the MCQ, and clinical reasoning, the GAMSAT.

High PPD scorers made more interjections, and significantly over-used the pronoun, 'we' in comparison to the lower scorers. This may suggest a focus on establishing their own professional identity and awareness of how they need to develop as a doctor, which is closely associated with the learning objectives of the PPD curriculum.

6.4.5.2 High and low scorers in the GAMSAT test (section 4.4.2)

High GAMSAT scoring speakers significantly over-used certain specific, highly technical terms. Low scorers were significantly more likely to use terms related to formulae, cause and effect and particularly 'why?' questions.

6.4.5.3 High and low scorers in the MCQ test (section 4.4.4)

High MCQ scoring speakers also significantly over-used certain specific technical terms., particularly anatomy and asked more 'If' and 'when' questions. However, low scorers were significantly more likely to refer to themselves, 'I' 'me' and ask more 'why?' questions.

6.4.5.4 Knowledge test results general discussion

High knowledge test scorers significantly over-use highly technical terms of relevance to this scenario. Low scorers, demonstrate greater concern in their speech behaviours for their own performance. As lower scorers ask more 'why' questions, this suggests more evidence for a hierarchy of questioning, where 'why' is associated with establishing basic understanding and is likely to co-occur with discussion of cause, whereas 'if', 'how' and 'when' questions signify a concern with establishing how this basic situation may be influenced in differing circumstances, in other words applying knowledge at a higher cognitive level within the clinical problem solving cycle (Bloom 1989).

6.4.6 Personality type and performance in knowledge tests

Model 4 (section 5.4.1), identified through multiple linear regression that the GAMSAT score, together with three of the five personality measures produced a significant model predicting performance in MCQ tests. GAMSAT scores on entry were moderately positively correlated with MCQ scores at the close of year. The higher the volatility, the lower the MCQ score suggests that anxiety may be impairing student performance. The higher the conscientiousness (structured type) measure the higher the MCQ score. Theoretical personality types were found to be negatively correlated with a high MCQ score

In summary, the calmer, more structured (conscientious), grounded (low openness to experience) learner, with a high GAMSAT score, tends to achieve significantly higher MCQ scores.

Overall, this model accords with the finding in this study that GAMSAT pre-entry and the MCQ knowledge test at the close of year one are aligned and so may be measuring the same constructs with high levels of validity and reliability. With respect to personality, higher volatility, often manifesting as anxiety, in this study is associated with lower scores, replicating findings elsewhere (McManus, Keeling et al. 2004, Tyssen, Dolatowski et al. 2007). The model also accords with general evidence that performance in controlled circumstances, such as knowledge tests in early medical study, is positively associated with higher conscientiousness scores (Lievens, Coessier et al. 2002, Ferguson, James et al. 2003) (McLachlan, Finn et al. 2009, Doherty and Nugent 2011).

The association between theoretical (higher openness to experience) learners and lower scores in knowledge tests is not strongly considered in the literature. A preference for deep learning, has been associated with the highest enjoyment of PBL in the St Georges study (Bigsby, McManus et al. 2013). Deep learning may have underlying constructs in common with high

openness to experience. and so might be over-represented in the type of student who is likely to be attracted to PBL study. For the participants in this study, there were indeed a higher proportion of theoretical learners (60%) as opposed to grounded learners (33%) but for the overall cohort the measure was not known. One might speculate that there might be a tension between high intellectual interest and the limitations of time and focus in revision and exam conditions in the early clinical years. Such students may benefit from awareness of their score on this factor and reinforcement of study skills designed to manage the huge volumes of study material requiring their attention in the early clinical years. Students with lower conscientiousness scores and higher volatility, may be particularly vulnerable to receiving test scores which are lower than their capability.

This might prove a fruitful area for future longitudinal research, particularly as higher openness to experience scores have been associated elsewhere with high intelligence and curiosity, (Moutafi, Furnham et al. 2006). This factor, together with conscientiousness may contribute, to different students being identified with success in later clinical practice when compared to those who do well in knowledge tests in the early clinical years (Doherty and Nugent 2011).

6.5 Personality type, learning styles and speech behaviour

6.5.1 Model 8: (section 5.5.2) Turns taken to speak and introversion/extraversion

Extraverts demonstrated a weak to moderate significant association with the proportion of turns taken to speak.

Greater 'surgency' or dominance in groups is associated with extraversion, but this group did not contribute a higher proportion of speech overall. In this study, there was evidence suggesting that extravert speakers are significantly more likely to take a higher proportion of student speech turns, although they do not contribute a significantly higher percentage of speech overall and so must be speaking more frequently, at shorter length than introverts. Corpus analysis identified further significant differences between the introvert and extravert speech corpora which may contribute to accounting for this difference. Extravert speakers spoke on a much wider number of semantic topics, particularly with evidence of interest in a wide range of topics, and significantly more interjections. This pattern may suggest that communication structures and roles within PBL may provide structures which encourage contributions by introverts, which then tend to favour their preference for preparation and speaking at greater length (Cain 2013)

6.5.2 Corpus analysis: discussion of highly significant key differences identified in student speech according to personality and learning style

6.5.2.1 Introversion and extraversion (section 4.5.1)

Extravert speakers interrupted and agreed more frequently than introverts, who used more technical terms related to anatomy and the operation of blood circulation. However, extraverts were significantly more likely to discuss medicines and a much wider range of relevant topics as well as the process of the PBL tutorial itself.

The study identifies significant differences in surgency in line with findings in the literature, with extraverts being more willing to express dominance in group settings, through taking a turn more frequently and discussing a very wide range of topics, but by necessity, not to as much depth (McCrae and Costa Jr 1999). Discussing medicines more frequently, may be linked to an awareness of and interest in other people, as this semantic category relates to how treatments affect the individual in many respects, as opposed to the specific, technical discipline of mastering anatomy.

By contrast, the introverts were found to be significantly more closely focussed on semantic categories related to the anatomy itself and to the operation of blood circulation. As noted through the statistical analysis, introverts were more likely to speak for longer when they did speak.

This invites speculation. This may suggest that mastering and using anatomical terminology is better suited to introvert temperaments, who are thought to be capable of sustained time working alone, enjoying structured tasks achieved through working quietly, alone (Cain 2013). Anatomy is also demanding of precision in applying terminology, in a rather static and predictable way. It may be that the discussion of anatomical knowledge in PBL settings, creating games and quizzes is a significant help to extravert learners in maintaining motivation for the very considerable task of mastering anatomical terminology in the first two years of medicine (Bridge 1962). Certainly, the key categories B1, B2 and B3 have different cognitive natures and have been shown in this study to differ between the different types of student corpora.

Extraverts are likely to have become more used to speaking without preparation, a communication skill which may be of particular benefit within consultation situations as well as within PBL. However, extraverts may benefit from a supportive group, which provides motivation and opportunity for developing their technical vocabulary, particularly in anatomy and physiology. It may be that the PBL group, a small and familiar group is particularly favourable to introverts, as these are the circumstances in which

they would be most willing to express their thoughts, in a known group and on an important topic (Cain 2013). For introverts, it may be that they might need to be encouraged to consider the reality of the clinical context more fully, especially through learning in a simulated environment, so that they can build confidence in unpredictable circumstances, such as consultations, which will allow them to transfer their skills into the clinical workplace with more confidence.

6.5.2.2 Personable and logical (high and low agreeableness) (section 4.5.6)

Corpus analysis supported theories that logical (low agreeableness) speakers were more technically oriented, speaking more frequently of anatomy and disease, whereas high agreeable speakers expressed more hesitance and said yes more frequently. Agreeable medical students have been shown in several studies to prefer PBL (Lievens, Coessier et al. 2002, Holen, Manandhar et al. 2015), and may benefit from contributing to effective group dynamics, which provide opportunities to listen to more technical aspects being discussed.

6.5.2.3 Adaptable and structured (Low and high conscientiousness) (section 4.5.4)

There was still significant evidence for structured (high conscientious) speech being more technically oriented, particularly with respect to the disease and likelihood categories: close to the heart of effective PBL talk. In addition, there was significant over-use by adaptable (low conscientious) speakers of initial letters, which brought to light an extended 'learning game'. One of the criticisms of lower conscientiousness scoring individuals in the workplace is that they can lack the discipline and focus to complete more boring tasks (O'Connor, Conner et al. 2009). Varying tasks and using games to motivate mastering repetitive tasks may be a beneficial strategy adopted by such learners and PBL is a suitable setting for this interaction.

The study did not support the findings for the over use of the first person singular 'I' reported by Iacobelli from written sources (Iacobelli, Gill et al. 2011).

6.5.2.4 Low and high volatility (low and high neuroticism) (section 4.5.8)

Corpus analysis suggests that learners who have high volatility (neuroticism) scores moderate their expression in response to their listeners and in so doing they reduce the extent to which their discussion uses technical terms, thus focus less on the PBL scenario. This may contribute to the situation identified by Model 4 (section 5.4.1) in the statistical evidence, for higher anxiety scores being significantly associated with lower scores in the MCQ and GAMSAT tests. It may well benefit all students to undertake training to become aware of their levels of volatility as the training for becoming a practising clinician extends over many years and encompasses many assessment opportunities. All students may benefit from the opportunity to identify and to practice more skilful management of their own motivating strategies through an approach such as Cognitive Behavioural Therapy (CBT). In particular, volatile learners may experience greater cost to their reserves of resilience and mental well-being through their fundamental temperament and reassurance alone wouldn't necessarily reduce someone's fundamental tendency to worry. Explicit skill development in basic CBT, particularly for volatile students, with low conscientiousness, may be extremely beneficial and assist in building and maintaining resilience (Stewart, Betson et al. 1997). In this study, 46.5% of participants where the measure was known, were volatile (neurotic) and 75% were adaptable (low conscientiousness) and high levels have been identified in other medical schools (Jadoon 2010).

For instance, volatile learners are far more likely to hesitate in speech, which may suggest that they may experience doubt or anticipate that they may be incorrect more frequently, as they are expressing their thoughts. They

express agreement more frequently, which together with their over-use of commenting on whether they are knowledgeable or not suggests that they would like to be agreed with themselves, and that they are keen to show others when they agree or approve of what others have said. The use of qualifying adverbs, or 'minimisers' has been associated with lack of confidence in speech, as a 'hedge' to forestall disagreement (Schiffrin, Tannen et al. 2008).

Volatile speakers also over-use pronouns relating to themselves, 'my' and others, 'yours' suggests that there is a higher awareness of their own performance and comparison with that of others. This might indicate a heightened sense of competitiveness, or a fear of being found inadequate in comparison to others. High volatile speakers also express a sense of obligation significantly more frequently, speaking of what they 'must' or 'ought' to do. This again is often linked to a desire for approval or agreement. Certainly, it indicates that in comparison with the calm speakers, there is considerable expression of speech which relates to self-monitoring associated with the high volatile speakers. At the very least, this restricts the amount of speech available for other purposes. By contrast, calm speakers are significantly more able to frequently discuss disease, cause and effect and anatomy: the chief purposes of the PBL group.

In the literature review, high volatility (neuroticism) was consistently associated with lower performance particularly in association with performance and in interaction with others. As Zajonc memorably demonstrated in experiments with chickens and cockroaches, (Zajonc, Heingartner et al. 1969) creatures are affected at the most primitive levels by awareness of others, which may have an encouraging influence on tasks completion when considered benign, but may also impair performance. An individual with higher neuroticism, may well be more aware of possible negative implications than their more calm colleagues, and this inflicts some 'process losses' on their ability to contribute speech during PBL discussions.

One study of PBL preference (Holen, Manandhar et al. 2015) suggested that individuals with high neurotic scores divided into two types: those who welcomed the reassurance of others in the PBL group, who therefore preferred that modality, and those who definitely preferred not to learn in the PBL group, perhaps to avoid what they experienced as unproductive stressors.

6.5.2.5 Groundedness versus theoretical (low or high openness to experience) (section 4.5.7)

Corpus analysis identified significant differences particularly relating to a focus on the present moment and priorities in low openness scorers, which supports the 'straightforwardness' reported in written work by Iacobelli (Iacobelli, Gill et al. 2011). However, this study did not report the use of longer words by those high in openness (theoretical) which would have been detected through the simplicity index. Overall, grounded learners demonstrated lower use of the technical terms, which were central to the PBL scenario, but this might be accounted for if they were already familiar with the terms and so did not need to practice their technical terminology (Bridge 1962).

6.5.2.6 Quintax Personality temperaments

The four Quintax temperament types were not normally distributed for the sample, varying between 7% for structured-grounded types, to 33% for logical-theoretical types. For 75% of the sample of professionals to be Adaptable, rather than Structured (18%) runs counter to the norms recorded by Quintax for professionals (Robertson, Wilkie et al.). The structured/adaptable measure is most closely associated with the conscientiousness scale, where adaptable scores have been associated with increased creativity and flexible open ended problem-solving, but lower industriousness and conformity (Barrick and Mount 1991). This finding is particularly interesting in the light of the adoption of conscientiousness indices discussed in sections 2.5.11 and 2.5.12. Flexible problem solving in changing environments becomes increasingly important as clinical training

progresses, which may account for the weakening of the link between conscientiousness and performance beyond the early years.

AG temperament speakers significantly over-used liking and difficulty and the phrase 'at the moment' when compared to all the other types combined. This perhaps suggests an enjoyment of practical problem solving which is particularly suited to PBL. The very small corpus for the SG temperament speakers limits conclusions. However, under-use of terms related to 'If', cause and effect and blood circulation when compared to all the other types combined suggests a disinclination to generalise beyond the specific case, which may be a limitation in PBL study where the scenario is the trigger for an area of study.

LT temperament speakers significantly over-used categories related to blood circulation, means and methods and anatomy and under-used terms such as 'sorry' 'yes' etc. and whether they know or remember something when compared to all the other types combined. This suggests that responding to the states and contributions of others is less important than discussing ideas.

PT temperament speakers significantly over-used expressions of hesitancy such as 'um' yet under-used terms related to blood-pressure, and disease when compared to all the other types combined. Together this suggests that there is a balance to be struck in discussion with others. Together with the strong data related to frequent speech by males, this may suggest that stronger facilitation may model and foster more effective discussion through fairer turn taking, time for reflection and demonstrating empathy to others; vital clinical communication skills.

6.5.2.7 Learning Styles

The distribution of the four learning styles was broadly reflective of other findings for the caring professions with lower proportions of accommodating and diverging style learners (section 2.4.9). Although the evidence for learning styles and performance are not compelling or coherent, significant differences were found when comparing one learning style corpus against the other corpora combined (Coffield, Moseley et al. 2004).

Assimilating learning style speakers (29%) (section 4.9.3) significantly over-used the technical term 'thrombus' and 'thrombi' and the pronoun 'they'. They over-used the anatomy category, and were significantly more likely to be found inaudible when compared to all the other types combined. They under-used the pronoun 'we'. Perhaps they sat further from the group and therefore the microphone. Zoghi terms the assimilator the fact oriented learner, who observes and theorises. There is some evidence that this group of learners focussed on factual content, in discussion of anatomy and technical terms and their underuse of 'we', preferring not to refer to themselves.

Convergent learning style speakers (25%) (section 4.9.1) significantly over-used the expression 'for example' and topics related to disease yet under-used terms related to speech, such as 's/he said' when compared to all the other types combined. There was not strong evidence for over use of technical terms in general, nor was this group significantly associated with high scores in the factually based tests, even though this has been identified in the literature (section 2.4). The corpus analysis evidence supports the differentiation of this learning style as being true to Zoghi's nickname of 'the devil's advocate learner' (Zoghi, Brown et al. 2010). The PBL discussion based on explanations of technical problems and problem solving with fellow learners may well suit those with this approach to learning.

Divergent learners (16%) (section 4.9.4) under-used some of the most relevant semantic categories for this PBL scenario, pressure and liquids, focussed on care of the body and the home, and were the only learners to mention death in a case related to heart failure. Zoghi terms divergers as social comfort learners as they enjoy collaborative work. They appear in higher proportions in studying the arts and humanities and their preference for imaginative approaches from multiple perspectives (Gurpinar, Alimoglu et al. 2010).

Accommodating learning style speakers (14%) (section 4.9.2), significantly over-used specific technical terms and the category relating to whether an occurrence was unusual but under-used terms from the anatomical category Zoghi terms accommodators, the 'hands on learner', who is good at adapting and learning through trial and error and enjoys being with others. The corpus analysis did not offer particular support for this theory.

6.6 The main implications of this study for learners and facilitators of PBL discussions

This study has provided objective evidence for individual factors which may indicate that students may be at risk of lower levels of achievement or have predictable development needs which relate to their future professional practice. It has offered evidence that simple typologies, such as temperament type or learning style are not sufficient to capture significant differences. However, there is evidence in this study to support the widespread view that neuroticism and conscientiousness are important factors in early student attainment. Particularly novel findings are that extraversion, gender and neuroticism may impact unhelpfully on PBL discussions. Once identified, characteristic patterns of speech behaviour for gender, extraversion and neuroticism are observable to the facilitator or learner themselves as an aspect of evaluating their effectiveness in learning through PBL. Whether interventions may mitigate these risks would be a fascinating field for future research.

In addition, the study has confirmed that there are objective clusters of behaviour which identify which students may be at particular risk without having access to personality or entry score data. Students who continue to rehearse the basic scientific facts, use more general or simpler technical vocabulary, avoid anatomy and use 'why' rather than 'what if' questions, are likely to be more at risk of lower attainment. If students tend to dwell on their own performance, using many personal pronouns, it is likely that they may also be experiencing anxiety, which may further impinge on their effectiveness. If students (particularly females) make few definite statements of fact, or only make statements of facts, interrupt others and seldom invite contributions (particularly males) then their communication skills may be ripe for development. PBL facilitators may become more attuned to identify this cluster of behaviours when they occur, for particular

topics or consistently for particular students and identify remedial assistance at an early stage of study. Experienced facilitators are likely to have built up similar insight as an aspect of their expertise. This study provides objective evidence for that insight.

All students may benefit from developing their study skills in the early clinical years and productive management of their emotions, particularly when preparing for high-stakes assessments. Where the literature and this study suggest that a particular profile of individual attributes may present higher vulnerability, for example, a student who is highly volatile (neurotic) and highly adaptable (low conscientiousness), and low in groundedness (openness to experience) this may be strongly advisable. It may be useful for all learners to be invited to share and critique their study strategies and that the PBL environment might be eminently suitable for fostering such understanding as an aspect of personal and professional development.

PBL has been credited with providing the opportunity to enhance communication skills and group working. This study suggests scope for modelling and developing specific professional norms whilst learning through PBL discussions which could reduce unproductive speech behaviour and may enhance learning opportunities for all. Specifically, facilitators and learners could draw attention to episodes of talking over others, especially dominance behaviours where 'holding the floor' may discourage others from equal speech participation. This adjustment alone may require a considerable shift away from what has previously felt 'normal' for all participants, male and female. It is likely that males could benefit from being encouraged to develop communication skills in addition to their likely information-giving and assertion-making strategies. Males could be encouraged to ask more open questions, to invite participation, particularly from quieter or female students. It is likely that many female students could be encouraged to make more clear statements of fact and knowledge rather than comment on

the assertions of others. Learners who are strongly extravert might be encouraged to monitor and reduce the number of turns they tend to take in discussions, introverts may be prompted to practise taking more turns. The PBL format provides the opportunity for 'teaching' fellow learners by giving a deeper, more carefully considered response, as well as spontaneous, informal discussion. All learners may benefit from ensuring that they recognise the challenge and benefits of both approaches for their individual personality type, particularly introversion/extraversion dimension.

The literature has long debated the benefits for the group of the facilitator's specialist knowledge base. This study suggests that, in this instance, the personality and knowledge of the facilitator may not be a dominant factor, so long as the learners have established an effective PBL group. The findings of this study inevitably focus on aspects of language and social process from within student speech and how these vary according to individual factors. However, the study does suggest that there is definite potential for developing effectiveness in learning through discussions in PBL through learners and facilitators becoming more aware of their own patterns of communication and personality. Although the literature suggests that many of these individual factors are innate and biologically determined, there is still scope for a professional to develop characteristic adaptations which will be of great benefit in developing their professional persona.

6.7 Re-statement of the known limitations of the study

As acknowledged throughout, this PhD study has chosen to focus only on student speech participation in relation to certain other measurable characteristics and is subject to, as is all research of human behaviour in the natural setting, many limitations.

Only one half of one cohort of GEMS was able to participate in the study, the number of subjects for which data was gained was 45. The GEMS at Derby is only representative of itself and so generalisability is extremely limited for

such data. It may be that this cohort differed greatly from other comparable graduate entry medical schools, or this particular cohort was also not representative of the usual cohort. For such specialist populations it is only possible to consider that other very similar, specialist populations may have similar characteristics.

Only whole PBL groups could participate. In several cases the facilitator chose not to participate, excluding the whole group. In others, one or more learners also withdrew. This pattern of participation may have skewed the sample to over-represent certain factors. For example, more extraverts or those with high openness to experience who were low in volatility may have chosen to take part, relative to the cohort as a whole, for whom these factors were not known. This personality profile has been associated with seeking new experiences and experiencing lower levels of anxiety which might pre-dispose them to participate. It is possible that only groups which were functioning well for their facilitator and participants chose to take part. However, within that number there were further reductions in the data collected, three subjects not providing FFM data; seven not providing KLSI data, of which two subjects provided neither. These circumstances led to a particularly small sample size being available particularly for conscientious (highly structured) students and for accommodating and diverging students. The study does not make strong claims on the basis of any of these areas as further research, with larger data sets, would be required to become more certain that both the distribution and characteristics associated with these aspects were significant as opposed to being determined by chance. In the case of the conscientiousness FFM, the findings replicate many other studies and it is on this basis that some interpretations have been made.

The linear regression statistical method is only ever able to identify possible association and account for relatively low proportions of variance in human behaviour at best. In this study, the findings of this method were triangulated and further explored in the main through the objective analysis

of corpora in order to build stronger confidence in possible inference. The method was used well within the recommended limitations for the number of cases which must be available relative to the number of variables. The method was applied by a psychometrician initially in order to determine the viability of likely statistical models being possibly identified. The entry and analysis was then independently and separately carried out in every respect by the researcher and calculated using both entry, forward and backward regression in order to establish the robustness of the findings.

The spoken word corpus itself was judged to be small in comparison to some in the field, but of at least a moderate size in comparison with other specialist, spoken word corpora, particularly as individual utterances were identified. There will be transcription error at play, such as in attributing utterance to speakers, giving the fullest account of what is said, particularly when more than one person is speaking at once, or where the microphone or position or quiet voice made it difficult to hear clearly. In such circumstances, it is likely that louder, deeper voices will be more audible which suggests that male speakers are more likely to be transcribed when others are speaking. In addition, there may be cultural factors which may favour one group above another in being transcribed fully. For example, if a speaker has an accent, speaks quickly or quietly, uses technical and unknown terms in close proximity and has a shy manner these may all contribute to error in transcribing their contribution to discussion. Where a subject had not provided data for all categories they were not included in the relevant sub corpus. However, simple human copying error may have been introduced at some stage during corpus creation, although corpus sizes were cross checked and corrected wherever error was found.

The Wmatrix3 package was used for corpus analysis. A problem with creating specialist dictionaries was identified during analysis, which slightly

restricted the extent to which specialist terminology could be fully explored. This was reported to the University of Lancashire Linguistics department. This problem is now acknowledged by the designer of the software and the website makes this limitation clear, pending development work to correct this. This package is particularly good at calculating log likelihood to determine relative frequency of language use and to compare on the basis of semantic categories, and parts of speech but there may still be unknown errors in the software. Two measures, log likelihood and Bayes' were both used in order to minimise the chance of over estimating the significance and prevalence of linguistic terms.

Although Wmatrix3 reports on objective measures, it is possible for confirmation bias to operate when considering which results to report, particularly when there are many results available, as was the case with the corpus analysis. Only statistically significant results at the highest level by two measures were reported in the results. However, judgement is inevitably involved in selecting which cases to illustrate, where to focus discussion, which conclusions to draw and which recommendations to make for future research. The researcher sought to focus on areas of practice for which there was strong evidence in the literature and practical scope to influence genuine practice in the effectiveness of offering learning through PBL discussions.

It would be an enhancement to subject the same corpus to analysis with different software, such as Wordsmith. It would be an interesting possibility to reconstitute the corpora in speaker sequence order and use the appropriate software to explore the interactions between speakers, for example in exploring gendered speech interaction.

6.8 Future research questions arising as a result of this study

Da Silva and Dennick called for a range of future investigations in their exploratory study in 2010 (Da Silva and Dennick 2010). This work has gone some way towards contributing to knowledge of learning through the PBL process, through identifying how individual students may behave differently during that process and how that difference may be associated with differing degrees of success. This has practical benefits. Because speech behaviour is readily observable, both by learners and the facilitator, it becomes possible to raise awareness of the links between speech behaviour, and internal mental processes, particularly in the cognitive and affective domains. For example, self-study behaviour, which has been shown to greatly influence effectiveness in learning especially in terms of time and focus on task, (Slavin 2003, Ericsson 2016) is not susceptible to observation by others; only through self-report and inference. Speech behaviour during discussion is an important means of 'sampling' the cognitive processes necessary for learning and providing access to benefits from and comparison with the comments of fellow learners and the input of a facilitator.

This study was restricted, due to feasibility, to one half of the cohort of first year students during one cycle of PBL discussion. It would be fascinating to be able to extend such a study to engage longitudinally with individual students and compare their speech behaviours and results as they progress through the early years and into clinical and specialist training. As technology progresses, methods of recording and transcribing and speech recognition may become easier and less labour intensive to employ. As blended learning becomes an increasingly common mode of learning, then online discussion will provide self-documenting corpora which may be more easily analysed, and provide identification of individuals with a very high level of accuracy, although interaction between students would be of a

different nature to that of face-to face discussion, perhaps favouring different skills.

A number of specific questions occur which explore to what extent it is possible to promote the development of an individual's range of behaviours relevant to study through discussion.

For example, the following section proposes several potential research questions which may be explored through comparing the educational outcomes for control and intervention PBL groups in response to different experiences.

- How does self-awareness of the personality type dimension, introversion-extraversion, impact on PBL discussions and the balance and range of discussion participation according to a pre-determined scale?
- In what ways does speech behaviour change in single gender PBL discussion groups and what impact does this have on student perception and assessment performance?
- Does a programme of behavioural de-sensitisation enhance the learning experience of highly volatile learners, in terms of student perception and assessment performance?
- How does gendered communication awareness training impact on learning through discussion for PBL groups and their facilitator, especially in terms of information giving and fostering more equally participative discussions?
- How might cost-effective personality testing be used in conjunction with activities to support and develop mentally robust health for students in the modern NHS?

6.9 Conclusions:

This study has further demonstrated the usefulness of corpus analysis in the analysis of individual speech behaviour in PBL and presented evidence for learners behaving differently when learning through discussion in ways which may confer benefit or disadvantage to their effectiveness in learning.

There are some very straightforward implications for training and developing facilitators of PBL and other discussion intensive forms of learning. As part of induction activities in GEMS, it may be helpful to encourage learners to enhance their approach to learning through effective speech behaviour in discussions such as PBL, by self-identification of their personality type on the big five dimensions, which can be carried out for no cost in under ten minutes (Gosling 2003). On the basis of this study this may be particularly relevant for the introvert/extravert, calm/volatile dimensions and for identifying potential to enhance their range and sophistication in conducting professional discussions, particularly where gender may have an impact.

There are clear implications for enhancing the approach of PBL facilitators and to optimise PBL group learning in the light of individual differences.

- The genders may be encouraged to self-monitor and extend their repertoire of communication strategies so that in general males are encouraged to employ more tentative and open questions and seek the contributions of others, particularly females, whilst generally, females might be encouraged to make more definitive statements of knowledge when warranted.
- Facilitators and learners may explicitly invite more frequent spoken contributions by introverted learners and from females, push for more complex, longer, topic-focussed contributions from extraverts, and seek to moderate the unproductive anxiety of volatile learners.

Abrams, D. R. (2007). "Princeton University Library Introduction to Regression." from http://dss.princeton.edu/online_help/analysis/regression_intro.htm.
Adolphs, S. (2006). Introducing electronic text analysis: A practical guide for language and literary studies, Routledge.
Adolphs, S. and D. Knight (2010). "Building a spoken corpus; what are the basics?." ADOLPHS, S. AND KNIGHT, D., 2010. Building a spoken corpus: what are the basics?. In: IN O'KEEFFE, A. AND MCCARTHY, M., ed., The Routledge Handbook of Corpus Linguistics Routledge.
Alghasham, A. A. (2012). "Effect of students' learning styles on classroom performance in problem-based learning." Medical Teacher **34**(s1): S14-S19.
Allport, G. and H. S. Odbert (1936). "Trait names: A psycho-lexical study." *Psychological Monographs* **47**: 1-117.
Anderson, H. and H. A. Goolishian (1988). "Human Systems as Linguistic Systems: Preliminary and Evolving Ideas about the Implications for Clinical Theory." Family Process **27**(4): 371-393.
Argamon, S., M. Koppel, J. Fine and A. R. Shimoni (2003). "Gender, genre, and writing style in formal written texts." TEXT-THE HAGUE THEN AMSTERDAM THEN BERLIN- **23**(3): 321-346.
Armstrong, E. and R. Parsa-Parsi (2005). "How can physicians' learning styles drive educational planning?" Academic Medicine **80**(7): 680-684.
Atenstaedt, R. (2012). "Word cloud analysis of the BJGP." The British Journal of General Practice **62**(596): 148-148.
Austin, P. C. and E. W. Steyerberg (2015). "The number of subjects per variable required in linear regression analyses." Journal of Clinical Epidemiology **68**(6): 627-636.
Ausubel, D. P., J. D. Novak and H. Hanesian (1968). "Educational psychology: A cognitive view."
Azer (2007). Navigating problem-based learning, Elsevier Australia.
Azer (2009). "Interactions between students and tutor in problem-based learning: The significance of deep learning." The Kaohsiung journal of medical sciences **25**(5): 240-249.
Baker, P. (2014). Using Corpora to analyse gender. London, Boomsbury.
Bales (1950). Interaction process analysis; a method for the study of small groups. Oxford, England, Addison-Wesley.
Bales, R. F. (1970). "Personality and interpersonal behavior."
Bandura, A. (1986). "Social foundations of thought and action." Englewood Cliffs, NJ.
Barbuto, J. E. (1997). "A CRITIQUE OF THE MYERS-BRIGGS TYPE INDICATOR AND ITS OPERATIONALIZATION OF CARL JUNG'S PSYCHOLOGICAL TYPES." Psychological Reports **80**(2): 611-625.
Barrick, M. R. and M. K. Mount (1991). "THE BIG FIVE PERSONALITY DIMENSIONS AND JOB PERFORMANCE: A META-ANALYSIS." Personnel Psychology **44**(1): 1-26.
Barrick, M. R. S., Greg L.; Neubert, Mitchell J.; Mount, Michael K. (1998). "Relating member ability and personality to work-team processes and team effectiveness." Journal of Applied Psychology **Vol 83(3**: 377-391.
Barrows, H. S. (1986). "A taxonomy of problem‐based learning methods." Medical education **20**(6): 481-486.
Barrows, H. S. (1988). The tutorial process, Southern Illinois Univ.
Barrows, H. S. and R. M. Tamblyn (1980). Problem-based learning: An approach to medical education, Springer Publishing Company.
Batchelder, A. e. a. (2012). Does allocating students to study groups according ot preferred team roles improve academic performance? ASME. Brighton.

Becker, H. S., B. Geer, E. C. Hughes and A. L. Strauss (1961). Boys in White. Chicago, IL, University of Chicago Press.
Belbin, R. M. (2000). Beyond the team, Routledge.
Belsley, D. A., E. Kuh and R. E. Welsch (2005). Regression diagnostics: Identifying influential data and sources of collinearity, John Wiley & Sons.
Benson (2012). The Big book of Psychology, Dorling Kindersley.
Benson, G., C. Noesgaard, M. Drummond-Young and E. Rideout (2001). "Facilitating small group learning." Rideout E. Transforming nursing education through problem-based learning. Hamilton: Jones and Bartlett: 75-102.
Bentler, P. M., Los Angeles: BMDP Statistical Software, In (1989). "EQS, Structural Equations, Program Manual, Program Version 3.0."
Biber, D. (2006). University language: A corpus-based study of spoken and written registers, John Benjamins Publishing.
Biber, D. and R. Reppen (1998). "Comparing native and learner perspectives on English grammar: A study of complement clauses." Learner English on computer: 145-158.
Bigsby, E., I. C. McManus, P. Sedgwick and P. McCrorie (2013). Which medical students enjoy problem-based learning?
Bishu, S. G. and M. G. Alkadry (2016). "A Systematic Review of the Gender Pay Gap and Factors That Predict It." Administration & Society: 0095399716636928.
Blanch, D. C., J. A. Hall, D. L. Roter and R. M. Frankel (2008). "Medical student gender and issues of confidence." Patient Education and Counseling **72**(3): 374-381.
Blissett, S., R. B. Cavalcanti and M. Sibbald (2012). "Should we teach using schemas? Evidence from a randomised trial." Medical Education **46**(8): 815-822.
Bloom, B. (1989). "Engelhart, MD, Furst, EJ, Hill, WH, & Krathwohl, DR (1956)." Taxonomy of educational objectives: Handbook I: Cognitive domain.
Blumer, T., & Doering, N, (2012). "Are we the same online? The expression of the five factor personality traits on the computer and the Internet." Cyberpsychology: Journal of Psychosocial Research on Cyberspace**.(3)**.
BNC. (2007). "BNC Educational Sampler." from http://corpora.lancs.ac.uk/BNCweb/.
BNC. (2007). "British National Sampler." Retrieved 13/02/17, from http://www.natcorp.ox.ac.uk/corpus/index.xml?ID=products#sampler.
Boud, D., R. Keogh and D. Walker (1985). "Reflection: Turning learning into experience." Kongan Page, London.
Bouhuijs, G., Kerkhofs (1984). The use of group data to trace the influence of individual students on group functioning. Tutorials in PBL: new directions in training for th health professions. d. V. Schmidt. Maastricht**:** 221-224.
Brescoll, V. L. (2011). "Who Takes the Floor and Why." Administrative Science Quarterly **56**(4): 622-641.
Bridge (1962). "The Language of Medicine A Quantitative Study of Medical Vocabulary." Academic Medicine **37**(3): 201-210.
Bridge, E. (1962). "The Language of Medicine A Quantitative Study of Medical Vocabulary." Academic Medicine **37**(3): 201-210.
Brookfield and Preskill (1999). Discussion as a way of teaching, Citeseer.
Brookfield, S. D. and S. Preskill (2012). Discussion as a way of teaching: Tools and techniques for democratic classrooms, John Wiley & Sons.
Bruffee, K. A. (1999). Collaborative learning: Higher education, interdependence, and the authority of knowledge, ERIC.
Burnard, L. (2007). "BNC User Reference Guide Spoken texts." from http://www.natcorp.ox.ac.uk/docs/URG/cdifsp.html#cdif22.
Cain, S. (2013). Quiet: The power of introverts in a world that can't stop talking, Random House Digital, Inc.

Camperio, Capiluppi, Veronese, Sartori, Capiluppi, Veronese and Sartori (2006). "The adaptive value of personality differences revealed by small island population dynamics." European Journal of Personality **21: 3-22**.
Capraro, R. M. and M. M. Capraro (2002). "Myers-briggs type indicator score reliability across: Studies a meta-analytic reliability generalization study." Educational and Psychological Measurement **62**(4): 590-602.
Carter, R. (1998). "Orders of reality: CANCODE, communication, and culture." ELT Journal **52**(1): 43-56.
Cashdan, A. and E. Grugeon (1972). "Language in Education: A Source Book."
Caspi, A., E. Chajut and K. Saporta (2008). "Participation in class and in online discussions: Gender differences." Computers & Education **50**(3): 718-724.
Caspi, R., Shiner (2005). "PERSONALITY DEVELOPMENT: Stability and Change." Annu. Rev. Psychol. 2005. 56:453-84.
Cattell, R. B. (1943). "The description of personality: basic traits resolved into clusters." The journal of abnormal and social psychology **38**(4): 476.
Cattell, R. B. (1957). "Personality and motivation structure and measurement."
Channel (1994). Vague Language. Oxford, Oxford University Press.
Chapman, B. P., P. R. Duberstein, S. Sörensen and J. M. Lyness (2007). "Gender Differences in Five Factor Model Personality Traits in an Elderly Cohort: Extension of Robust and Surprising Findings to an Older Generation." Personality and individual differences **43**(6): 1594-1603.
Charmaz, K. (2009). Shifting the grounds - constructivist grounded theory methods. Developing grounded theory: the second generation. California, Left Coast Press, Inc.
Clark, R., P. A. Kirschner and J. Sweller (2012). "Putting students on the path to learning: the case for fully guided instruction." Am Educa**.36**((1):): 6-11.
Cleland, J. A., S. Nicholson, N. Kelly and M. Moffat (2015). "Taking context seriously: explaining widening access policy enactments in UK medical schools." Medical Education **49**(1): 25-35.
Clouston, T. J. and S. W. Whitcombe (2005). "An emerging person centred model for problem‐based learning." Journal of further and higher education **29**(3): 265-275.
Coates, J. (2015). Women, men and language: A sociolinguistic account of gender differences in language, Routledge.
Coffield, F., D. Moseley, E. Hall and K. Ecclestone (2004). Learning styles and pedagogy in post 16 learning: a systematic and critical review, The Learning and Skills Research Centre.
Cohen, L., Manion, L., Morrison, K. (2008). Research methods in education. London/New York, Routledge.
Cook, R. D. W., Sanford. (1982). "Residuals and Influence in Regression." Retrieved 27/03/2017, from http://hdl.handle.net/11299/37076.
Costa, P. T. "McCrae. RR (1992)." Revised NEO Personality Inventory (NEO PI-R) and NEO Five Factor Inventory. Professional Manual.
Costa, P. T. and R. R. McCrae (1985). "The NEO personality inventory."
Costa, P. T. and R. R. McCrae (1992). "Four ways five factors are basic." Personality and Individual Differences **13**(6): 653-665.
Covey, S. (1989). "The seven habits of effective people." New York: Simon.
Creswell, J. W. (2007). Qualitative inquiry and research design: Choosing among the five approaches. Thousand Oaks, California, SAGE Publications Inc.
Creswell, J. W. (2009). Research design: Qualitative, quantitative and mixed methods approaches. Thousand Oaks, California, SAGE Publications Inc.
Crowdy, S. (1994). "Spoken Corpus Transcription." Literary and Linguistic Computing **9(1)**: 25-28.

Crowdy, S. (1994). "Spoken Corpus Transcription." Literary and Linguistic Computing **9(1)**: 25-28.
Curry, L. (1999). "Cognitive and learning styles in medical education." Academic Medicine **74**(4): 409-413.
Da Silva, A. L. (2014). "PhD Thesis."
Da Silva, A. L. and R. Dennick (2010). "Corpus analysis of problem-based learning transcripts: an exploratory study." Medical Education **44**(3): 280-288.
Darwin, C. (1998). The expression of the emotions in man and animals, Oxford University Press.
Davis, M. H. (1999). "AMEE Medical Education Guide No. 15: Problem-based learning: a practical guide." Medical Teacher **21**(2): 130-140.
Day, D. V. and S. B. Silverman (1989). "Personality and job performance: Evidence of incremental validity." Personnel Psychology **42**(1): 25-36.
De Grave, W. S., H. P. A. Boshuizen and H. G. Schmidt (1996). "Problem based learning: Cognitive and metacognitive processes during problem analysis." Instructional Science **24**(5): 321-341.
Dennick (2016). "Constructivism: reflections on twenty five years teaching the constructivist approach in medical education." International Journal of Medical Education **7**: 200.
Dennick, R. (2012). "Twelve tips for incorporating educational theory into teaching practices." Med Teach **34**(8): 618-624.
Department of Health (2004). Medical Schools: Delivering the Doctors of the Future. Crown Publications. T. D. o. Health.
Dickson, D., C. Saunders and M. Stringer (1993). Rewarding People: The Skill of Responding Positively (International Series on Communication Skills), Routledge.
Doherty and Nugent (2011). "Personality factors and medical training: a review of the literature." Medical Education **45**(2): 132-140.
Dolmans, D. and D. Gijbels (2013). "Research on problem‐based learning: future challenges." Medical education **47**(2): 214-218.
Dolmans, D. H. J. M., W. De Grave, I. H. A. P. Wolfhagen and C. P. M. Van Der Vleuten (2005). "Problem‐based learning: future challenges for educational practice and research." Medical education **39**(7): 732-741.
Dolmans, D. H. J. M. and H. G. Schmidt (1994). "What drives the student in problem‐based learning?" Medical Education **28**(5): 372-380.
Donaldson, L. (2006). Good Doctors, Safer Patients. Department of Health. C. M. Officer. London.
Dong, Y. and C.-Y. J. Peng (2013). "Principled missing data methods for researchers." SpringerPlus **2**: 222.
Dow, B. J. and J. T. Wood (2006). The Sage handbook of gender and communication, Sage.
Driscoll, M. P. (1994). Psychology for learning for instruction. Boston, Allyn and Bacon.
Duek, J. E. (2000). Equity of student discourse in PBL. Problem Based Learning: A research Perspective on Learning interactions.
Dweck, C. S. (1986). "Motivational processes affecting learning." American psychologist **41**(10): 1040.
Ekman, P. (2003). "Micro Expressions Training Tool and The Subtle Expressions Training Tool (METT AND SETT)." MOZGO Media, Venice, CA.
Ellis, R. and D. Whittington (1981). A guide to social skill training, Croom Helm London.
Engels, P. T. and C. de Gara (2010). "Learning styles of medical students, general surgery residents, and general surgeons: implications for surgical education." BMC medical education **10**(1): 1.

Ericsson, K. A. (2016). "Summing Up Hours of Any Type of Practice Versus Identifying Optimal Practice Activities Commentary on Macnamara, Moreau, & Hambrick (2016)." Perspectives on Psychological Science **11**(3): 351-354.
Eva, K. W. (2016). "The more things stay the same, the more they change." Medical Education **50**(1): 1-2.
Eysenck, H. J. (1977). Crime and Personality.
Eysenck, H. J., & Eysenck, S. B. G. (1975). Manual of the Eysenck Personality Qyestionnaire (Junior and Adult), Hodder & Stoughton.
Eysenck, S. B. and H. J. Eysenck (1963). "An experimental investigation of "desirability" response set in a personality questionnaire." Life Sciences **2**(5): 343-355.
Faraone and Hurtig (1985). "An examination of social skill, verbal productivity, and Gottman's model of interaction using observational methods and sequential analyses." Behavioral Assessment.
Felder, R. M. and R. Brent (2005). "Understanding student differences." Journal of engineering education **94**(1): 57-72.
Ferguson, E., D. James and L. Madeley (2002). "Factors associated with success in medical school: systematic review of the literature." Bmj **324**(7343): 952-957.
Ferguson, E., D. James, F. O'Hehir, A. Sanders and I. C. McManus (2003). "Pilot study of the roles of personality, references, and personal statements in relation to performance over the five years of a medical degreeCommentary: How to derive causes from correlations in educational studies." BMJ **326**(7386): 429-432.
Ferguson, E., D. James, F. O'Hehir and A. Saunders (2003). A pilot study of the roles of personality, references and personal statements in relation to performance over the five years of a medical degree. Brit Med J.
Festinger, L. (1957). "A Theory of Cognitive Dissonance.".
Finn, G. M., M. Carter, M. Sawdon, N. Thompson and P. Tiffin (2012). "The relationship between personality traits, self-report conscientiousness the Conscientiousness Index and academic performance in undergraduate medical students."
Finn, G. M., M. Carter, M. Sawdon, N. Thompson and P. Tiffin (2012). The relationship between personality traits, self-report conscientiousness the Conscientiousness Index and academic performance in undergraduate medical students. ASME Annual Scientific Meeting. Brighton Centre, Brighton.
Finucane, P., D. Flannery, D. McGrath and J. Saunders (2013). "Demographic attributes and knowledge acquisition among graduate-entry medical students." Medical teacher **35**(2): 134-138.
Firth (1986). "Levels and sources of stress in medical students." BMJ.
Fox, N. (1992). The Social Meaning of Surgery. Milton Keynes, Open University Press.
Franke, R. H. and J. D. Kaul (1978). "The Hawthorne experiments: First statistical interpretation." American sociological review: 623-643.
Funder, D. C. and C. D. Sneed (1993). "Behavioral manifestations of personality: an ecological approach to judgmental accuracy." Journal of personality and social psychology **64**(3): 479.
Furnham, A. (1996). "The big five versus the big four: the relationship between the Myers-Briggs Type Indicator (MBTI) and NEO-PI five factor model of personality." Personality and Individual Differences **21**(2): 303-307.
Furnham, A., C. J. Jackson and T. Miller (1999). "Personality, learning style and work performance." Personality and Individual Differences **27**(6): 1113-1122.
Fusch, P. I., & Ness, L. R. (2015). "Are We There Yet? Data Saturation in Qualitative Research." The Qualitative Report **20**(9): 1408-1416.

Gabrielatos, C. and A. Marchi (2011). "Keyness: Matching metrics to definitions." Theoretical-methodological challenges in corpus approaches to discourse studies-and some ways of addressing them.
GAMSAT. (2017). "About GAMSAT." Retrieved 13/02/2017, 2017, from http://gamsat.acer.edu.au/gamsat-uk.
Garner, I. (2000). "Problems and Inconsistencies with Kolb's Learning Styles." Educational Psychology **20**(3): 341-348.
Garrud, P. (2011). "Who applies and who gets admitted to UK graduate entry medicine?-an analysis of UK admission statistics." BMC medical education **11**(1): 71.
Garrud, P. and J. Yates (2012). "Profiling strugglers in a graduate-entry medicine course at Nottingham: a retrospective case study." BMC medical education **12**(1): 124.
Geen, R. G. and E. C. O'Neal (1969). "Activation of cue-elicited aggression by general arousal." J Pers Soc Psychol **11**(3): 289-292.
Geerligs, T. (1995). "Students' thoughts during problem-based small-group discussions." Instructional Science **22(4)**: 269-278.
Gill, D. and A. Griffin (2010). "Good Medical Practice: what are we trying to say? Textual analysis using tag clouds." Medical education **44**(3): 316-322.
Glaser, B. G. and A. L. Strauss (1967). The Discovery of Grounded Theory: Strategies for Qualitative Research. Chicago, Aldine Publishing Company.
GMC. (2016). "List of Registered Medical Practitioners - statistics." Retrieved Accessed 25 May 2016, from http://www.gmc-uk.org/doctors/register/search_stats.asp.
GMC. (2016). "Promoting excellence: standards for medical education and training." from www.gmc-uk.org/education/standards.asp.
GMC. (2016). "Promoting excellence: standards for medical education and training." Retrieved 07/02/2017, from http://www.gmc-uk.org/education/standards.asp,.
Goldberg, L. R. (1990). "An alternative" description of personality": the big-five factor structure." Journal of personality and social psychology **59**(6): 1216.
Goldstein, H. (1968). "Longitudinal studies and the measurement of change." Journal of the Royal Statistical Society. Series D (The Statistician) **18**(2): 93-117.
Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). "A Very Brief Measure of the Big Five Personality Domains." Journal of Research in Personality **37**: 504-528.
Grabowski., G. J. T. a. S. R. (2002). "Principles of Anatomy and Physiology." (Tenth Edition).
Grave, D., Dolmans and v. d. Vleuten (2002). "Student perspectives on critical incidents in the tutorial group." Advances in Health Sciences Education **7**(3): 201-209.
Green, J. and N. Thorogood (2009). Qualitative methods for health research. London, SAGE Publications, Inc.
Gurpinar, E., M. K. Alimoglu, S. Mamakli and M. Aktekin (2010). "Can learning style predict student satisfaction with different instruction methods and academic achievement in medical education?" Adv Physiol Educ **34**(4): 192-196.
Gurpinar, E., H. Bati and C. Tetik (2011). "Learning styles of medical students change in relation to time." Adv Physiol Educ **35**(3): 307-311.
Haines, C. (2010). Assessing participation in PBL using corpus analysis: a pilot study. 12th Ottawa Conference: Assessment of competence in Medicine and the Healthcare Professions, Miami.
Haines, C. (2012). A Research presentation: Assessing participation in Problem Based Learning using corpus analysis: a PhD study. 13th Ottawa Conference: Assessment of competence in Medicine and the Healthcare Professions. Kuala Lumpur.
Haines, C. (2013). Presentation: An exploration of the use of figurative language in PBL student explanations. Association for Medical Education in Europe - AMEE. Prague.

Hamburger, Y. A., & Ben-Artzi, E. (2000). "The relationship between extraversion and neuroticism and the different uses of the Internet." Computers in Human Behavior **16**: 441-449.
Hammick, M., T. Dornan and Y. Steinert (2010). "Conducting a best evidence systematic review. Part 1: From idea to data coding. BEME Guide No. 13." Medical teacher **32**(1): 3-15.
Harland, T. (2003). "Vygotsky's zone of proximal development and problem-based learning: Linking a theoretical concept with practice through action research." Teaching in higher education **8**(2): 263-272.
Hartmann, Christensen and Keelig (2011). "Training young horses to social separation: effect of a companion horse on training efficiency." Equine Vert **43**: 580 - 584.
Harvey, K. J., B. Brown, P. Crawford, A. Macfarlane and A. McPherson (2007). "'Am I normal?' Teenagers, sexual health and the internet." Social Science & Medicine **65**(4): 771-781.
Hawk, T. F. and A. J. Shah (2007). "Using learning style instruments to enhance student learning." Decision Sciences Journal of Innovative Education **5**(1): 1-19.
Hawkins, K. and C. B. Power (1999). "Gender differences in questions asked during small decision-making group discussions." Small Group Research **30**(2): 235-256.
Hayes, K., A. Feather, A. Hall, P. Sedgwick, G. Wannan, A. Wessier-Smith, T. Green and P. McCrorie (2004). "Anxiety in medical students: is preparation for full-time clinical attachments more dependent upon differences in maturity or on educational programmes for undergraduate and graduate entry students?" Medical Education **38**(11): 1154-1163.
Heffler, B. (2001). "Individual learning style and the learning style inventory." Educational studies **27**(3): 307-316.
Hendry, G. D., G. Ryan and J. Harris (2003). "Group problems in problem-based learning." Medical Teacher **25**(6): 609-616.
Heron, J. (1999). The complete facilitator's handbook, Kogan Page London.
Hickerson, A. and J. Gastil (2008). "Assessing the Difference Critique of Deliberation: Gender, Emotion, and the Jury Experience." Communication Theory **18**(2): 281-303.
Hmelo-Silver, C. E. (2004). "Problem-based learning: What and how do students learn?" Educational Psychology Review **16**(3): 235-266.
Hogan, J. and R. Hogan (1989). "How to measure employee reliability." Journal of Applied psychology **74**(2): 273.
Hogan, R. M. and W. Kintsch (1971). "Differential effects of study and test trials on long-term recognition and recall." Journal of Verbal Learning and Verbal Behavior **10**(5): 562-567.
Hojat, M., J. S. Gonnella, T. J. Nasca, S. Mangione, M. Vergare and M. Magee (2002). "Physician empathy: definition, components, measurement, and relationship to gender and specialty." American Journal of Psychiatry **159**(9): 1563-1569.
Holen, A., K. Manandhar, D. S. Pant, B. M. Karmacharya, L. M. Olson, R. Koju and D. I. Mansur (2015). "Medical students' preferences for problem-based learning in relation to culture and personality: a multicultural study." Int J Med Educ **6**: 84-92.
Honey, P. M., A. (1982). Manual of Learning Styles. London, P Honey.
Huang, H.-C., S.-T. Yang, Y.-T. Kuo and S.-C. Yang (2007). "The study of" Free-rider" in Problem-based Learning Situation."
Hunston, S. (2002). Corpora in applied linguistics, Ernst Klett Sprachen.
Iacobelli, F., A. J. Gill, S. Nowson and J. Oberlander (2011). Large Scale Personality Classification of Bloggers. Affective Computing and Intelligent Interaction: Fourth International Conference, ACII 2011, Memphis, TN, USA, October 9–12, 2011, Proceedings, Part II. S. D'Mello, A. Graesser, B. Schuller and J.-C. Martin. Berlin, Heidelberg, Springer Berlin Heidelberg**:** 568-577.

Jackson, C. and M. Lawty-Jones (1996). "Explaining the overlap between personality and learning style." Personality and Individual Differences **20**(3): 293-300.
Jadoon, Y., Raza, Shehzad, Zeshan. (2010). "Anxiety and depression among medical students: a cross-sectional study." J Pak Med Assoc. Aug;60(8):699-702.
Jamieson (2013). "Developing a socio-constructivist, community of inquiry-based conceptualisation of critical thinking and applying this to investigate critical thinking in problem-based learning tutorials." Amee abstract conference presentations Prague 2013.
Jamieson (2014). Critical Thinking in a Community of Inquiry: applying a socio-constructivist framework to prolem-based learning tutorials in a Scottish medical curriculum, to investigate critical thinking and the factors that influence this. EdD Doctorate of Education, University of Strathclyde, Schol of Education.
Jefferson, L., K. Bloor, Y. Birks, C. Hewitt and M. Bland (2013). "Effect of physicians' gender on communication and consultation length: a systematic review and meta-analysis." Journal of Health Services Research & Policy **18**(4): 242-248.
Jennings, P. (2013). PBL but not as we know it: an ethnography of the practice and facilitation of'problem-based learning'within a hybrid graduate-entry medical programme in England, University of Nottingham.
Johnson, D. (2015). "Emotional intelligence as a crucial component of medical education." International Journal of Medical Education: 179 - 215.
Johnson, D. W. and R. T. Johnson (1987). Learning together and alone: Cooperative, competitive, and individualistic learning, Prentice-Hall, Inc.
Johnson, D. W. and R. T. Johnson (2008). Social Interdependence Theory and Cooperative Learning: The Teacher's Role. The Teacher's Role in Implementing Cooperative Learning in the Classroom. R. M. Gillies, A. F. Ashman and J. Terwel. Boston, MA, Springer US**:** 9-37.
Jung, C. G. (1921). Psychology of the Unconscious: A Study of the Transformations and Symbolisms of the Libido; a Contribution to the History of the Evolution of Thought, Moffat, Yard.
Kamp, R. J. A., D. H. J. M. Dolmans, H. J. M. van Berkel and H. G. Schmidt (2012). "The relationship between students' small group activities, time spent on self-study, and achievement." Higher Education **64**(3): 385-397.
Keirsey, D. and M. M. Bates (1984). Please understand me, Prometheas Nemesis.
Kelly, S. and R. Dennick (2009). "Evidence of gender bias in True-False-Abstain medical examinations." BMC Medical Education **9**: 32-32.
Kirschner, P. A., J. Sweller and R. E. Clark (2006). "Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching." Educational Psychologist **41**(2): 75-86.
Knowles, M. S., Holton, E. & Swanson, R.A. (). (1998). The Adult Learner: The Definitive Classic in Adult Education and Human Resource Development Houston, Tex, .: Gulf Pub. Co.
Koh, G. C.-H., H. E. Khoo, M. L. Wong and D. Koh (2008). "The effects of problem-based learning during medical school on physician competency: a systematic review." Canadian Medical Association Journal **178**(1): 34-41.
Kolb, A. Y. and D. A. Kolb (2005). The Kolb Learning Style Inventory—Version 3.1 Technical Specifications, Hay Group.
Kolb, A. Y. and D. A. Kolb (2005). "Learning styles and learning spaces: Enhancing experiential learning in higher education." Academy of management learning & education **4**(2): 193-212.
Kolb, D. A. (1971). Individual learning styles and the learning process, MIT.
Kolb, D. A. (1981). "Learning styles and disciplinary differences." The modern American college: 232-255.
Kolb, D. A. (1984). Experiential learning: Experience as the source of learning and development, Prentice-Hall Englewood Cliffs, NJ.

Kondracki, N. L., N. S. Wellman and D. R. Amundson (2002). "Content analysis: review of methods and their applications in nutrition education." Journal of nutrition education and behavior **34**(4): 224-230.
Kurtz, S. M. and J. D. Silverman (1996). "The Calgary—Cambridge Referenced Observation Guides: an aid to defining the curriculum and organizing the teaching in communication training programmes." Medical Education **30**(2): 83-89.
Laerd (2016) "Simple Linear Regression."
Lakoff, G. and M. Johnson (1999). Philosophy in the flesh: The embodied mind and its challenge to western thought, Basic books.
Lave, J. (1991). "Situating learning in communities of practice." Perspectives on socially shared cognition **2**: 63-82.
Lee Ellis, S. H., Evelyn Field, Scott Wersinger, Sergio Pellis, David Geary, Craig Palmer, Katherine Hoyenga, Amir Hetsroni, Kazmer Karad (2008). Sex Differences: Summarizing More than a Century of Scientific Research. New York, Taylor & Francis Group.
Leech, G. (1992). "100 million words of English: the British National Corpus (BNC)." Language Research **28**(1): 1-13.
Levin, K. A. (2006). "Study design III: Cross-sectional studies." Evid-based Dent **7**(1): 24-25.
Lievens, F., P. Coessier, F. De Fruyt and J. Maeseneer (2002). Medical Education **36**(null): 1050.
Lievens, F., D. S. Ones and S. Dilchert (2009). Journal of Applied Psychology **94**(null): 1514.
Lipton, Huxham and Hamilton (1984). "Predictors of success in a cohort of medical students." Med Educ.
Loehlin, J. C., R. R. McCrae, P. T. Costa and O. P. John (1998). "Heritabilities of common and measure-specific components of the Big Five personality factors." Journal of Research in Personality **32**(4): 431-453.
Lukar, P. (1989). "PhD Dissertation: Academic Staff Development in Universities with Specific Reference to Small Group Teaching."
Lynch, T. G., N. N. Woelfl, D. J. Steele and C. S. Hanssen (1998). "Learning style influences student examination performance." The American Journal of Surgery **176**(1): 62-66.
MacDonald, A. P. (1971). "Internal-external locus of control: Parental antecedents." Journal of consulting and clinical psychology **37**(1): 141.
MacNell, L., A. Driscoll and A. N. Hunt (2015). "What's in a Name: Exposing Gender Bias in Student Ratings of Teaching." Innovative Higher Education **40**(4): 291-303.
Mahlberg, M. (2013). Corpus Stylistics and Dickens's Fiction. London: , Routledge.
Maier, S. F. and M. E. Seligman (1976). "Learned helplessness: Theory and evidence." Journal of experimental psychology: general **105**(1): 3.
Mammen, J. M., D. R. Fischer, A. Anderson, L. E. James, M. S. Nussbaum, R. H. Bower and T. A. Pritts (2007). "Learning styles vary among general surgery residents: analysis of 12 years of data." Journal of surgical education **64**(6): 386-389.
Manning, G. and P. Garrud (2009). "Comparative attainment of 5-year undergraduate and 4-year graduate entry medical students moving into foundation training." BMC medical education **9**(1): 76.
Marton, F. and R. Saljo (1976). On Qualitative Difference in Learning. I - Outcome and Process.
Marton, F. and R. Saljo (1976). "On qualitative differences in learning: outcome and progression." British Journal of Educational Psychology **46**(1): 4-11.
Maslow, A. H. (2013). Toward a psychology of being, Start Publishing LLC.
Mast, M. S. (2002). "Dominance as expressed and inferred through speaking time." Human Communication Research **28**(3): 420-450.
Masterson, J. G. (1965). Letters from Jenny, Harcourt.

Mathers, J., A. Sitch and J. Parry (2016). "Population-based longitudinal analyses of offer likelihood in UK medical schools: 1996–2012." Medical Education **50**(6): 612-623.
Mazzoni, D. (2016). Audacity.
McAdams, D. P. and J. L. Pals (2006). "A new Big Five: fundamental principles for an integrative science of personality." American Psychologist **61**(3): 204.
McCrae, R. R. and P. T. Costa Jr (1999). "A five-factor theory of personality." Handbook of personality: Theory and research **2**: 139-153.
McCrae, R. R. and P. T. Costa (2010). NEO inventories for the NEO personality inventory-3 (NEO-PI-3), NEO five-factor inventory-3 (NEO-FFI-3), NEO personality inventory-revised (NEO PI-R) : professional manual. Lutz, Fla, PAR.
McCrae, R. R. and A. Terracciano (Mar 2005). "Universal Features of Personality Traits From the Observer's Perspective: Data From 50 Cultures." Journal of Personality and Social Psychology **Vol 88(3), , 547-561.**
McEnerie, H. (2011). "Statistics in corpus linguistics." from http://corpora.lancs.ac.uk/clmtp/2-stat.php.
McLachlan, J. (2010). "Measuring conscientiousness and professionalism in undergraduate medical students." The Clinical Teacher **7**(1): 37-40.
McLachlan, J. C., G. M. Finn and J. Macnaughton (2009). "The Conscientiousness Index : a novel tool to explore students' professionalism." Academic medicine. **84**(5): 559-565.
McManus, I., A. Keeling and E. Paice (2004). "Stress, burnout and doctors' attitudes to work are determined by personality and learning style: A twelve year longitudinal study of UK medical graduates." BMC Medicine **2**(1): 1-12.
Medical Schools' Council. (2017). "Studying Medicine. ." Retrieved 05/02/2017, from https://www.medschools.ac.uk/studying-medicine.
Mehl, M. R. and J. W. Pennebaker (2003). "The sounds of social life: a psychometric analysis of students' daily social environments and natural conversations." Journal of personality and social psychology **84**(4): 857.
Mehmood, S. I., M. A. Khan, K. M. Walsh and J. C. C. Borleffs (2013). "Personality types and specialist choices in medical students." Medical teacher **35**(1): 63-68.
Micase (2007). "Michigan Corpus of Academic English Biochemistry Group."
Misch, D. A. (2002). "Andragogy and Medical Education: Are Medical Students Internally Motivated to Learn?" Advances in Health Sciences Education **7**(2): 153-160.
Mischel, W. (1968). Personality and Assessmen. New York: , Wiley.
Moskowitz, D. S. Z., David C. (2004). "Flux, Pulse, and Spin: Dynamic Additions to the Personality Lexicon." Journal of personality and social psychology **Vol 86(6)**.
Moust, J. H. C., F. Schmidt, de Volder, Belien and W. De Grave (1986). The effects of verbal participation in small group discussion on learning. Student Learning: Reserch in education and Cognitive Psychology. P. Richardson, Eysenck**:** 147-155.
Moustakas, C. (1994). Phenomenological research methods. Thousand Oaks, California, SAGE Publications Inc.
Moutafi, J., A. Furnham and J. Crump (2006). "What facets of openness and conscientiousness predict fluid intelligence score?" Learning and Individual Differences **16: 31-42.**
Mulac, A. and J. J. Bradac (1995). "Women's Style in Problem Solving Interaction: Powerless, or Simply." Gender, power, and communication in human relationships: 83.
Myers, I. B., M. H. McCaulley and R. Most (1985). Manual, a guide to the development and use of the Myers-Briggs type indicator, Consulting Psychologists Press.
Myers, M. (2000). "Qualtitative research and the generalisability question: standing firm with Proteus." The Qualtitative Report [on line serial] Retrieved 9/4/09, from http://www.nova.edu/ssss/QR/QR4-3/myers.html.

Nederhof, A. J. (1985). "Methods of coping with social desirability bias: A review." European Journal of Social Psychology **15**(3): 263-280.
Newman, M. L., C. J. Groom, L. D. Handelman and J. W. Pennebaker (2008). "Gender Differences in Language Use: An Analysis of 14,000 Text Samples." Discourse Processes **45**(3): 211-236.
Norman, G. R. and H. G. Schmidt (2016). "Revisiting 'Effectiveness of problem-based learning curricula: theory, practice and paper darts'." Medical Education **50**(8): 793-797.
Norman, W. T. (1967). "2800 PERSONALITY TRAIT DESCRIPTORS--NORMATIVE OPERATING CHARACTERISTICS FOR A UNIVERSITY POPULATION."
Nowson, S. (2006). "The Language of Weblogs: A study of genre and individual differences."
O'Cathain, A. O. and K. Thomas (2006). Combining qualitative and quantitative methods. Qualitative research in health care. C. Pope and N. Mays. Oxford, Blackwell Publishing Ltd.
O'Connor, D. B., M. Conner, F. Jones, B. McMillan and E. Ferguson (2009). "Exploring the benefits of conscientiousness: An investigation of the role of daily stressors and health behaviors." Annals of Behavioral Medicine **37**(2): 184-196.
O'Malley, A. J. and P. V. Marsden (2008). "The analysis of social networks." Health services and outcomes research methodology **8**(4): 222-269.
Oberlander, J. and A. J. Gill (2006). "Language With Character: A Stratified Corpus Comparison of Individual Differences in E-Mail Communication." Discourse Processes **42**(3): 239-270.
Oxford, U. o. (2017). "Medical Literature Searching Skills." Retrieved 07/02/2017, from learntech.physiol.ox.ac.uk/cochrane_tutorial/cochlibd0e187.php.
Papinczak, T., L. Young, M. Groves and M. Haynes (2007). "An analysis of peer, self, and tutor assessment in problem-based learning tutorials." Medical Teacher **29**(5): e122-e132.
Park, S., A. Griffin and D. Gill (2012). "Working with words: exploring textual analysis in medical education research." Medical Education **46**(4): 372-380.
Parsons, H. M. (1974). "What Happened at Hawthorne?" New evidence suggests the Hawthorne effect resulted from operant reinforcement contingencies **183**(4128): 922-932.
Paul, S., M. Bojanczyk and J. Lanphear (1994). "Learning preferences of medical students." Medical Education **28**(3): 180-186.
Pavitt, C. (1994). Small group discussion: A theoretical approach, Gorsuch Scarisbrick.
Peduzzi, P., J. Concato, E. Kemper, T. R. Holford and A. R. Feinstein (1996). "A simulation study of the number of events per variable in logistic regression analysis." Journal of Clinical Epidemiology **49**(12): 1373-1379.
Penland, P. R. and S. Fine (1974). Group dynamics and individual development, M. Dekker.
Pennebaker, J. W. and L. A. King (1999). "Linguistic styles: language use as an individual difference." Journal of personality and social psychology **77**(6): 1296.
Perry, W. G., Jr. (1970). "Forms of Intellectual and Ethical Development in the College Years: A Scheme .".
Physics, I. o. (2017). "Gender Balance." Retrieved 21/02/2017, 2017, from http://www.iop.org/education/teacher/support/girls_physics/page_41593.html.
Plovnick, M. S. (1975). "Primary care career choices and medical student learning styles." Academic Medicine **50**(9): 849-855.
Ponterotto, J. G. (2006). " Brief Note on the Origins, Evolution, and Meaning of the Qualitative Research Concept Thick Description." The Qualitative Report **11**(3): 538-549.
Poole, M. S. and A. B. Hollingshead (2005). Theories of small groups: Interdisciplinary perspectives, Sage.
Pope, C., N. Mays, S. Ziebland, A. le May, S. Williams, M. Coombs, A. Le May, D. Wicke, R. Coppin and P. Doorbar (2006). "Qualitative methods in health research." methods **1**: 2.
Prideaux, D., C. Roberts, K. Eva, A. Centeno, P. McCrorie, C. McManus, F. Patterson, D. Powis, A. Tekian and D. Wilkinson (2011). "Assessment for selection for the health care

professions and specialty training: Consensus statement and recommendations from the Ottawa 2010 Conference." Medical Teacher **33**(3): 215-223.
Quirk, R. (1986). Words at work: lectures on textual structure, NUS Press.
Rayson, P. (2003). Matrix: A statistical method and software tool for linguistic analysis through corpus comparison, Lancaster University.
Rayson, P. (2008). "From key words to key semantic domains." International Journal of Corpus Linguistics **13**(4): 519-549.
Rentfrow, Jost, S. D. Gosling and Potter (2009). Statewide Differences in Personality Predict Voting Patterns in 1996–2004 U.S. Presidential Elections. Social and Psychological Bases of Ideology and System Justification**:** p.314-348.
Richards, J. C. and R. W. Schmidt (2013). Longman dictionary of language teaching and applied linguistics, Routledge.
Ridley, M. (2003). Nature via nurture: Genes, experience, and what makes us human. New York, NY, US, HarperCollins Publishers.
Ringelmann, M. (1913). "Recherches sur les moteurs animés: Travail de l'homme" [Research on animate sources of power: The work of man],." Annales de l'Institut National Agronomique **2nd series, vol. 12**(pages 1-40.).
Rivard, L. P. and S. B. Straw (2000). "The effect of talk and writing on learning science: An exploratory study." Science Education **84**(5): 566-593.
Roberts, B. W., K. E. Walton and W. Viechtbauer (2006). "Patterns of mean-level change in personality traits across the life course: a meta-analysis of longitudinal studies." Psychological bulletin **132**(1): 1.
Roberts, C., Atkins, S, Hawthorne, K (2014). "Performance features in clinical skills assessment."
Robertson, S., D. Wilkie and V. Edwards "Of Quintax® Types and Traits."
Rosenfeld, B. and S. D. Penrod (2011). Research methods in forensic psychology, John Wiley & Sons.
Roter, D. L., J. A. Hall and Y. Aoki (2002). "Physician gender effects in medical communication: A meta-analytic review." JAMA **288**(6): 756-764.
Ruigrok, A. N. V., G. Salimi-Khorshidi, M.-C. Lai, S. Baron-Cohen, M. V. Lombardo, R. J. Tait and J. Suckling (2014). "A meta-analysis of sex differences in human brain structure." Neuroscience & Biobehavioral Reviews **39**: 34-50.
Sacks, H., E. A. Schegloff and G. Jefferson (1974). "A simplest systematics for the organization of turn-taking for conversation." Language: 696-735.
Sackville, A. and C. Sherratt (2006). Styles of Discussion: online facilitation factors. th International Conference on Networked Learning.
Samuel and Widiger (2008). Clinical Psychology Review **28**(null): 1326.
Samuel DB1, W. T. (2011). "Conscientiousness and obsessive-compulsive personality disorder." Personal Disord. Jul;2 **(3):161-74. doi: 10.1037/a0021216.**
Savery, J. R. and T. M. Duffy (1995). "Problem based learning: An instructional model and its constructivist framework." Educational technology **35**(5): 31-38.
Savin-Baden, M. and C. H. Major (2004). Foundations of problem-based learning, McGraw-Hill International.
Savin-Baden, M. and C. H. Major (2013). Qualitative research: The essential guide to theory and practice, Routledge.
Savin Baden, M. and K. Wilkie (2004). Challenging research in problem-based learning, McGraw-Hill Education (UK).
Schiffrin, D., D. Tannen and H. E. Hamilton (2008). The handbook of discourse analysis, John Wiley & Sons.
Schmidt and Moust (2000). Small Group Learning: A review of research.

Schmidt, F. G. W. (1990). Causal Modelling of PBL American Education Research Association. Boston.
Schmidt, H. G., J. I. Rotgans and E. H. J. Yew (2011). "The process of problem-based learning: what works and why." Medical Education **45**(8): 792-806.
Schmidt, H. G., Spaay, G. & deGrave, W.S. (1988). "Opsporen van misconcepties bij middelbare scholieren." Tijdschrift voor Onderwijsresearch: 13: 129-140.
Schmidt, H. G., H. T. Van Der Molen, W. W. R. Te Winkel and W. H. F. W. Wijnen (2009). "Constructivist, Problem-Based Learning Does Work: A Meta-Analysis of Curricular Comparisons Involving a Single Medical School." Educational Psychologist **44**(4): 227-249.
Schmit, M. J., A. M. Ryan, S. L. Stierwalt and A. B. Powell (1995). "Frame-of-Reference Effects on Personality Scale Scores and Criterion-Related Validity." Journal of Applied Psychology **Vol. 80**(No. 5): 607-620.
Schön, D. A. (1983). The reflective practitioner: How professionals think in action, Basic books.
Schroeder, H. E., R. F. Rakos and J. Moe (1983). "The social perception of assertive behavior as a function of response class and gender." Behavior Therapy **14**(4): 534-544.
Scott, M. (1997). "PC analysis of key words." System **Vol. 25**(No. 2): pp. 233 245.
Sealey, A. and S. Bates (2016). "Prime ministerial self-reported actions in Prime Minister's Questions 1979–2010: A corpus-assisted analysis." Journal of Pragmatics **104**: 18-31.
Shen, C. (1997). "Predicting medical students' academic performances by their cognitive abilities and personality characteristics." Acad Med. Sep;72(9):781-6.
Silverman, D. (2007). Doing Qualitative Research. London, Sage.
Sinclair, J. (2004). Trust the text: Language, corpus and discourse, Routledge.
Slavin, R. (2003). Educational psychology: Theory and practice. Boston:, Pearson Education.
Sowell, Thompson, Holmes, Jernigan and Toga (1999). "In vivo evidence for post-adolescent brain maturation in frontal and striatal regions." Nat Neurosci **10**: 859-861.
Stake, R. (1980). The case study inquiry in social inquiry. Towards a science of the singular. H. Simons. Norwich, CARE:UEA.
Stewart, S. M., C. Betson, T. H. Lam, I. B. Marshall, P. W. H. Lee and C. M. Wong (1997). "Predicting stress in first year medical students: a longitudinal study." Medical Education **31**(3): 163-168.
Strobel, J. and A. Van Barneveld (2009). "When is PBL more effective? A meta-synthesis of meta-analyses comparing PBL to conventional classrooms." Interdisciplinary Journal of Problem-based Learning **3**(1): 4.
Suler, J. (2004). "The online disinhibition effect." Cyberpsychology & behavior **7**(3): 321-326.
Tanguay, E. (2015). "Disciplinary Talk: A systemic functional exploration of university seminar discussions." PhD thesis**. . Aston University, Birmingham, UK.**
Tannen, D. (1993). Gender and conversational interaction, Oxford University Press.
Tannen, D. (2001). You just don't understand: Women and men in conversation, HarperCollins.
Tannen, D. (2002). "Agonism in academic discourse." Journal of pragmatics **34**(10): 1651-1669.
Tavakol, M. and J. Sandars (2014). "Quantitative and qualitative methods in medical education research: AMEE Guide No 90: Part I." Med Teach **36**(9): 746-756.
Thibaut, J. W. and H. H. Kelley (1959). "The social psychology of groups."
Thompson, P. and H. Nesi (2001). "The British Academic Spoken English (BASE) Corpus Project." Language Teaching Research **5**(3): 263-264.
Thorne, A. (1987). "The press of personality: A study of conversations between introverts and extraverts." Journal of Personality and Social Psychology **Vol 53(4)**.

Tipping, J., R. F. Freeman and A. R. Rachlis (1995). "Using faculty and student perceptions of group dynamics to develop recommendations for PBL training." Academic medicine **70**(11): 1050-1052.
Tokode, A. (2016). "Collaborative Knowledge Construction in Graduate Entre Problem-Based Learning, unpublished PhD Thesis, University of Nottingham."
Trofimova, I. (2014). "Observer Bias: An Interaction of Temperament Traits with Biases in the Semantic Perception of Lexical Material." PLOS.
Tuckman, B. W. (1965). "Developmental sequence in small groups." Psychological bulletin **63**(6): 384.
Tuckman, B. W. and M. A. C. Jensen (1977). "Stages of small-group development revisited." Group & Organization Management **2**(4): 419-427.
Tyssen, R., F. C. Dolatowski, J. O. Røvik, R. F. Thorkildsen, Ø. Ekeberg, E. Hem, T. Gude, N. T. Grønvold and P. Vaglum (2007). "Personality traits and types predict medical school stress: a six-year longitudinal and nationwide study." Medical Education **41**(8): 781-787.
Ubiqus. (2017). "Ubiqus - Who we are." from https://www.ubiqus.co.uk/about-us/.
UKCAT. (2017). "UK information." from https://www.themedicportal.com/application-guide/ukcat/.
Valle, R., L. Petra, A. Martínez-Gonzáez, J. A. Rojas-Ramirez, S. Morales-Lopez and B. Piña-Garza (1999). "Assessment of student performance in problem-based learning tutorial sessions." Medical Education **33**(11): 818-822.
Van Boxtel, C., J. Van der Linden and G. Kanselaar (2000). "Collaborative learning tasks and the elaboration of conceptual knowledge." Learning and instruction **10**(4): 311-330.
Visschers-Pleijers, A. J. S. F., D. H. J. M. Dolmans, I. H. A. P. Wolfhagen and C. P. M. Van Der Vleuten (2004). "Exploration of a method to analyze group interactions in problem-based learning." Medical Teacher **26**(5): 471-478.
Vittinghoff, E. and C. E. McCulloch (2007). "Relaxing the Rule of Ten Events per Variable in Logistic and Cox Regression." American Journal of Epidemiology **165**(6): 710-718.
von Elm, A., Egger, Pocock, Gøtzsche, Vandenbroucke; STROBE Initiative. (2008). "The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE)statement: guidelines for reporting observational studies." J Clin Epidemiol. 2008 Apr;61(4):344-9. PMID: 18313558.
Vygotskiĭ, L. S., and Michael Cole. . P, (1978). "Mind in Society: The Development of Higher Psychological Processes."
Wallis, S. and G. Nelson (2001). "Knowledge discovery in grammatically analysed corpora." Data Mining and Knowledge Discovery **5**(4): 305-335.
Waterhouse, L. (2006). "Multiple intelligences, the Mozart effect, and emotional intelligence: A critical review." Educational Psychologist **41**(4): 207-225.
Webb (1980). "Group process: the key to elarning in groups." New Directions for methodology of social and behavioral science **6**: 77-87.
Webb, L. E., C. G. van Reenen, M. B. Jensen, O. Schmitt and E. A. M. Bokkers (2015). "Does temperament affect learning in calves?" Applied Animal Behaviour Science **165**: 33-39.
Wheelwright, S., S. Baron-Cohen, N. Goldenfeld, J. Delaney, D. Fine, R. Smith, L. Weil and A. Wakabayashi (2006). "Predicting Autism Spectrum Quotient (AQ) from the Systemizing Quotient-Revised (SQ-R) and Empathy Quotient (EQ)." Brain Research **1079**(1): 47-56.
Wilkie, D., S. Robertson and Y. Allan (2006). Images of Competency: Self Report, Assessment Centre, 360 Degree and Personality Perspectives. Poster presented at BPS DOP Conference, Citeseer.
Williams, J. L. (1971). "Personal space and its relation to extraversion introversion " Canadian Journal of Behavioural Science **3**(2): 156-160.

Wilson, A. (2013). "Embracing Bayes factors for key item analysis in corpus linguistics. In: New approaches to the study of linguistic variability." Language Competence and Language Awareness in Europe: 3-11.
Wood, D. F. (2003). "Problem based learning." British medical journal **326**(7384): 328.
Wun, Y. T., E. Y. Y. Tse, T. P. Lam and C. L. K. Lam (2007). "PBL curriculum improves medical students' participation in small-group tutorials." Medical Teacher **29**(6): e198-e203.
Yarkoni, T. (2010). "Personality in 100,000 Words: A large-scale analysis of personality and word use among bloggers." Journal of Research in Personality **44**(3): 363-373.
Yeo, J.-Y. and S.-H. Ting (2014). "Personal pronouns for student engagement in arts and science lecture introductions." English for Specific Purposes **34**: 26-37.
Yerkes, R. M. D., J. D. (1908). "The Relation of Strength of Stimulus to Rapidity of Habit-Formation. ." Journal of Comparative Neurology and Psychology **18**: 459-482.
Yew, E. H. J. and H. G. Schmidt (2007). "Process study of verbal interactions in problem-based learning." ICT: Providing choices for learners and learning. Proceedings ascilite Singapore **2007**.
Yin, R. K. (1994). Case study research: design and methods. London, SAGE Publications, Inc.
Zajonc, R. B., A. Heingartner and E. M. Herman (1969). "Social enhancement and impairment of performance in the cockroach." Journal of Personality and Social Psychology **13**(2): 83.
Zoghi, M., T. Brown, B. Williams, L. Roller, S. Jaberzadeh, C. Palermo, L. McKenna, C. Wright, M. Baird and M. Schneider-Kolsky (2010). "Learning style preferences of Australian health science students." Journal of allied health **39**(2): 95-103.

6.10 Appendices

Appendix A: Ethical consent from School of Medicine

The University of Nottingham
Faculty of Medicine and Health Sciences

Medical School Research Ethics Committee Division of Therapeutics & Molecular Medicine

D Floor, South Block

Queen's Medical Centre Nottingham

NG7 2UH

7th December 2010

Dr Reg Dennick

Assistant Director of Medical Education

Medical Education Unit

B Floor, Medical School

Dear Dr Dennick

Ethics Reference No: D/9/2008 - *Please quote this number on all correspondence*

Study Title: Monitoring the development of technical language and clinical reasoning in problem based learning groups of graduate entry medical students. **Lead Investigator:** Dr Reg Dennick, Assistant Director of Medical Education **Co Investigators:** Ana Linda Da Silva, PhD Student, School of Community Health Sciences. Catherine Haines, Lecturer in Medical Education, Medical Education Unit.

Thank you for your letter dated 17th November 2010 requesting the following amendments to the above study:

- Give all participants the following questionnaires: Quintax 'Big Five' personality inventory and Kolb's Learning styles inventory - both questionnaires were enclosed for review
- Request for student demographic data to include the results of examinations up to and including ACE course final examinations (MCQs and OSCEs)

These have been reviewed and are satisfactory and the study amendments are approved.

Approval is given on the understanding that the Conditions of Approval set out below are followed.

Conditions of Approval

You must follow the protocol agreed and any changes to the protocol will require prior Ethics Committee approval.

This study is approved for the period of active recruitment requested. The Committee also provides a further 5 year approval for any necessary work to be performed on the study which may arise in the process of publication and peer review.

You promptly inform the Chairman of the Ethics Committee of

(i) Deviations from or changes to the protocol which are made to eliminate

(ii) Any changes that increase the risk to subjects and/or affect significantly the conduct of the research.

(iii) All adverse drug reactions that are both serious and unexpected,

(iv) New information that may affect adversely the safety of the subjects or the conduct of the study,

(v) The attached End of Project Progress Report is completed and returned when the study has finished,

Yours sincerely



Professor R C Spiller

Chairman, Nottingham University Medical School Research Ethics Committee

Appendix B: Ethics Addendum

The University of
Nottingham

Faculty of Medicine and Health Sciences
Medical Education Unit

Medical School

Queen's Medical Centre
Nottingham

NG7 2UH

Tel: +4.4 (0) 115 82 30028
Fax: +44 (0) 115 82 30014

17/1 1/2010

Medical School Ethics Committee

Dear Professor Spiller

Ethics reference No: D/9/2008

Monitoring the development of technical language and clinical reasoning in PBL groups of graduate entry medical students.

I wish to make an Addendum to the above Ethics request.

We wish to do the following:

Give all participants the following questionnaires:

- **Quintax 'Big Five' personality inventory (on-line but details attached)**
- **Kolb's Learning Styles inventory (attached)**

In addition we wish to extend our request for student demographic data to include the results of examinations up to and including the ACE course final examinations (MCQs and OSCEs).

Yours sincerely

Professor Reg Dennick

Professor of Medical Education

Assistant Director of Medical Education

Course Director: Masters in Medical Education

Appendix C: Sample study consent form

Observational Study Consent Form *University of Nottingham*

The individual language of PBL 2011. **Medical Education Unit**

Professor Reg Dennick Reg.dennick@nottingham.ac.uk

Catherine Haines Catherine.haines@nottingham.ac.uk

You may choose to participate in this observational study entitled, **The individual language of PBL 2011.** The purpose of the study is to look at the relation between personality factors and the quantity and quality of participation in PBL. If you decide to join the study, you will be asked to give permission for researchers to audio transcribe PBL sessions and will complete on-line personality tests, which will take about 30 minutes at most. The University Ethics Committee has given approval for this study to take place, so long as each individual is able to opt in or out of the study. Your joining is important to the success of this study.

Purpose of the Observational Study

The main purpose of this Observational Study is to learn more about the way students learn through discussion, in particular, medical students through Problem Based Learning. The University staff has determined that you are eligible to join this part of the study.

Reasons for the Observational Study

Currently students at GEMS learn through discussing clinical cases for at least 4.5 hours per week during the first 18 months of their course. If we could better understand the mechanisms for this and the connections between the amount of talk, the personal preferences, the different levels of success, we could help students and facilitators be even more effective in their learning.

What Will You Be Doing?

If you decide to take part in the study, and your fellow PBL participants and Facilitator also consent, you will be audio recorded during one full cycle of PBL; ie: three sessions. The audio recording will be professionally transcribed, the researcher will identify who is speaking, using a sample of your recorded voice and will produce individual transcription files. We may follow up on this study during the remainder of your GEMS study. During this follow-up period, you will receive feedback on the pattern of participation within your PBL group during the recorded cycle of PBL.

Benefits and Risks

There is no risk for you through taking part in this study. We hope that you will receive some benefit through seeing your relative contribution to PBL discussions and through having access to your personality profile.

Costs

The tests, procedures and visits that are a part of this study cost you only your time in completing the on-line quizzes. The personality and learning styles tests will be paid for on your behalf.

Confidentiality

All of your study records will be kept confidential to the extent required by law. Your personal identity will not be revealed in any publication or release of results. Study records will be kept indefinitely for analysis and follow-up. All primary data collected through this study will be kept securely in digital form. The data will be stored in digital form in the locked office of the researcher and an offsite copy stored. The audio transcription will be anonymised and only the researcher will know which code corresponds to which individual in which group. The anonymised transcripts may be used in future training sessions with PBL facilitators and for future study. The audio recordings will not be used for any purpose other than for study. The results of this study will be available in an anonymised form and may be presented at conferences and papers and as part of a PhD thesis in the medical education sector.

Right to Withdraw

Your decision to join in this study is voluntary. You may quit at any time, for any reason, without notice. We hope you will take part for the entire time of the study because we will use all of this information to draw correct conclusions.

Voluntary Consent

If you have any questions about any aspect of the study or your rights as a volunteer, a staff member will be on hand to answer them before you sign this consent form. Additionally, if you have any questions about your rights as a participant in this study, you may call or email either of the investigators listed at the beginning of this form.

Before you sign this form, please ask any questions on any part of this study that is unclear to you.

I **consent** to taking part in a PBL cycle which will be

audio recorded to produce a transcript from which my contribution is redacted _____(initials)

I **consent** to taking on-line personality and learning styles tests

and will share the results the researchers _____(initials)

I do **not consent** to taking part in a PBL cycle which will be

audio recorded to produce a transcript _____(initials)

I do **not consent** to taking on-line personality and learning styles tests
_____(initials)

Investigator's Statement

I have provided an explanation of the above research program. The participant was given an opportunity to discuss these procedures, including possible alternatives, and to ask any additional questions. A signed copy of the consent form has been given to the participant.

_____
Signature of Principal Investigator or Designee Date

Participant Statement

I certify that I have read, or had read to me, and that I understand the description of the Observational Study. I have asked and received answers to all my questions about the study. I voluntarily consent to join the study. I understand that I may quit the study at any time. I understand that I may ask further questions at any time and that I will receive a copy of this signed consent form for my records. I have had an opportunity to carefully review the Informed Consent form and ask questions about it.

_____
Signature of Participant Date

_____
Signature of Witness Date

Current PBL Group No. _____

Email address _____

Thank you!

Appendix D: Sample individual Quintax report

Quintax®

PROFILE CHART

Name: Seema Nadeem
Date: 08 November 2016 Norm Group: Quintax 2015 Norm: Men and Women

Scale	Score	Left pole description	Position on scale 1	2	3	4	5	6	7	8	9	10	Right pole description
Extraversion	5	INTROVERT Reflective, likes to think before acting. Prefers few interests pursued deeply. Needs time for quiet concentration. Enjoys privacy. Appears quiet and reserved to others.					●						EXTRAVERT Action oriented: do first, think later. Prefers broad range of interests, busy environments, working with others. Tells others own thoughts. Appears open and friendly.
Criticality	3	PERSONABLE Makes decisions based on feelings and values of self or group. Likes to fit in. Achieves group goals through harmony. Can show great enthusiasm.			●								LOGICAL Bases decisions on logical and reasoned analysis of facts. Values truth and sticking to principles. Takes a critical stance. Achieves group goals through efficiency.
Organisation	6	ADAPTABLE Always ready to take in extra information. Prefers to deal with things as they arise. Likes to start the new before finishing the old. Quick to respond in a crisis.						●					STRUCTURED Takes an organised approach. Prefers to schedule and plan. Can be impatient with the unforeseen. Enjoys making decisions and completing tasks.
Intellectual Focus	4	GROUNDED Likes to deal with practical matters in the here and now. Focus on facts and details. Prefers to use existing skills and likes pragmatic problem solving. Conventional.				●							THEORETICAL Prefers to consider the wider meaning of information and the ramifications of proposals. Likes theory, questioning. Focus on innovation and learning new skills.
Emotional Involvement	7	CALM Not easily upset by events. Tends not to be affected by strong emotions. Tries to maintain control over feelings. Usually resilient to day-to-day pressures and stresses.							●				VOLATILE Generally lets others know of feelings. Can react emotionally to events. Tends to show effects of pressure and stress. Shows strong emotional involvement in activities.
		A score of	1	2	3	4	5	6	7	8	9	10	
		is obtained by	2%	4%	9%	15%	19%	19%	15%	9%	4%	2%	of the norm group

Team Role	Organiser	Driver	Thinker	Evaluator	Networker	Harmoniser	Deliverer	Finisher	Expert
Score	6	6	5	5	5	7	7	8	7

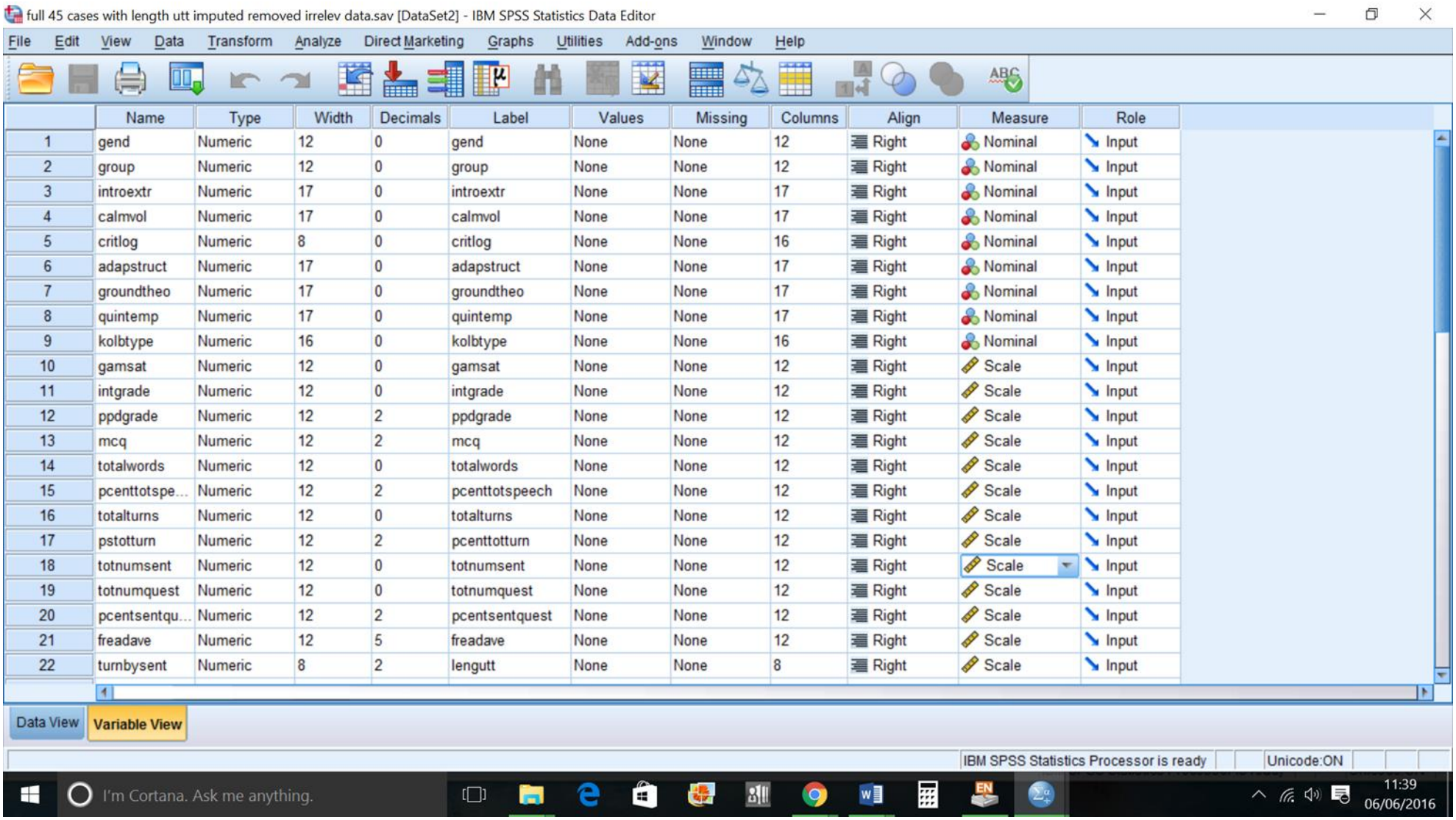
© 1998, 2015 Stuart Robertson & Associates Ltd. Other than for the purpose of using Stuart Robertson & Associates Ltd's Quintax On-line assessment service, no part of the present work may be reproduced or copied by any means or through any medium without the express written permission of the publishers. No permission is granted for the editing and/or reuse of any portion of this material. Published exclusively by Stuart Robertson & Associates Ltd, Empress Buildings, 380 Chester Road, Manchester M16 9EA, UK. Quintax is a registered trade mark of Stuart Robertson & Associates Ltd.

Appendix E Spreadsheet showing summary and averages for each PBL group's participation

PBL Gp	Personality orientation: Introvert or Extravert continuum: i	Personality neuroticism: Calm versus volatile: calm = 1 vola	Personality, Criticality: Personable versus logical: per1log2	Personality, ordering: adaptable =1 v structured = 2	grounded = 1 theoretical =2	Quintax Type 1pt, 2 lt, 3 ag, 4 sg	Kolb LSI type: assim = 1, conv =2, accom = 3, diverg= 4	Gamsat score on entry	Interview grades: 3 acceptable grades. MMI: A=1, B=2, C	Personal and Professional Development Mark %; PPD 1 Mar	Knowledge Exam Year 1 end: MCQ TOTAL /120. MCQ TOT	MCQ Score expressed as Average %. MCQ Average %	Total number of transcribed words for students: 3 sessions	total words per student	% total speech per student	total number of turns taken to speak	number of sentences	total number of ?s	proportion of sentences questions
													17792		13%	917	1420	367	26%
2	1	2	1	2	2	1	2	62	2	58.2	82	68		3047	17%	140	219	78	36%
2	1	1	2	1	2	2	1	63	2	63	67	62		2891	16%	62	178	24	13%
2	2	1	2	1	2	2	2	77	3	76.2	85	85.5		3206	18%	128	242	34	14%
2	2	1	2	1	2	2	3	68	1	76.6	69	68.5		2856	16%	140	230	33	14%
2	1	2	1	2	1	4	1	74	2	79.6	101	90.5		1121	6%	91	116	45	39%
2	1	2	1	1	2	1	1	73	3	63.8	73	71		808	5%	92	81	14	17%
2	1	2	1	1	1	3	4	76	1	64.6	87	81.5		1037	6%	64	83	19	23%
2	1	1	2	2	2	2	3	70	2	80.4	94	85.5		2826	16%	200	271	120	44%
													5442		14%	304	476	217	46%
7	1	2	2	1	1	3	3	66	1	59.8	77	74		777	14%	43	71	7	10%
7	1	2	1	1	1	3	2	70	2	87.4	93	84.5		132	2%	19	22	11	50%
7	1	1	2	1	1	3	2	62	1	71.4	82	79		403	7%	32	41	10	24%
7	1	2	1	1	1	3	2	71	2	58.8	84	77.5		244	4%	27	28	15	54%
7	2	1	2	1	2	2	4	61	1	79.8	74	75		1741	32%	44	101	18	18%
7	2	1	1	1	2	2	2	76	3	53.6	82	81.5		886	16%	52	86	24	28%
7							1	70	3	55.4	90	84.5		1259	23%	87	127	16	13%
													23399		14%	1534	1819	431	24%
8	1	1	2	1	2	2	1	81	1	79.4	101	90.5		3881	17%	218	251	54	22%
8	1	2	1	1	2	1	4	62	1	74.6	80	69.5		2864	12%	143	180	37	21%

												21995		13%	1085	1907	482	25%
10	2	1	1	2	2	1	4	63	2	77.4	93	90		3259	15%	161	293	69	24%
10	2	1	2	1	1	3	2	65	2	54.6	89	84		3235	15%	112	299	81	27%
10	2	1	2	1	2	2	4	61	2	68.4	74	69.5		3277	15%	187	298	77	26%
10	1	1	2	1	2	2	1	67	2	62	79	80		2656	12%	126	212	48	23%
10	1	1	2	1	2	2	1	61	2	61.8	87	79		2878	13%	98	165	39	24%
10	2	2	1	2	2	1	3	62	1	72.4	78	84.5		2388	11%	113	212	42	20%
10	2	2	1	1	2	1	4	72	3	60.8	78	67.5		1848	8%	113	177	35	20%
10	1	1	2	2	2	2	1	68	1	74	107	89		2454	11%	175	251	91	36%
												14371		13%	847	812	255	31%
11	1	1	2	2	2	2		66	2	61.4	81	76		1816	13%	36	29	6	21%
11	2	2	1	1	2	1	1	63	2	50.2	76	69.5		1490	10%	112	103	43	42%
11	1	2	1	2	1	4		65	1	72.2	82	76		1322	9%	76	65	23	35%
11	2	2	2	1	1	3		68	3	63.2	87	76		1819	13%	150	141	50	35%
11	1	1	2	1	1	3	2	65	2	73	74	74		1482	10%	137	126	27	21%
11	1	2	1	1	1	3		64	2	57.2	83	74		3946	27%	158	179	49	27%
11	2	2	2	2	1	4		63	1	51	81	74.5		1302	9%	87	96	24	25%
11	1	2	1	1	2	1		73	3	63	73	73		1194	8%	91	73	33	45%

Appendix F: SPSS variables



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	gend	Numeric	12	0	gend	None	None	12	Right	Nominal	Input
2	group	Numeric	12	0	group	None	None	12	Right	Nominal	Input
3	introextr	Numeric	17	0	introextr	None	None	17	Right	Nominal	Input
4	calmvol	Numeric	17	0	calmvol	None	None	17	Right	Nominal	Input
5	critlog	Numeric	8	0	critlog	None	None	16	Right	Nominal	Input
6	adapstruct	Numeric	17	0	adapstruct	None	None	17	Right	Nominal	Input
7	groundtheo	Numeric	17	0	groundtheo	None	None	17	Right	Nominal	Input
8	quintemp	Numeric	17	0	quintemp	None	None	17	Right	Nominal	Input
9	kolbtype	Numeric	16	0	kolbtype	None	None	16	Right	Nominal	Input
10	gamsat	Numeric	12	0	gamsat	None	None	12	Right	Scale	Input
11	intgrade	Numeric	12	0	intgrade	None	None	12	Right	Scale	Input
12	ppdgrade	Numeric	12	2	ppdgrade	None	None	12	Right	Scale	Input
13	mcq	Numeric	12	2	mcq	None	None	12	Right	Scale	Input
14	totalwords	Numeric	12	0	totalwords	None	None	12	Right	Scale	Input
15	pcenttotspe...	Numeric	12	2	pcenttotspeech	None	None	12	Right	Scale	Input
16	totalturns	Numeric	12	0	totalturns	None	None	12	Right	Scale	Input
17	pstotturn	Numeric	12	2	pcenttotturn	None	None	12	Right	Scale	Input
18	totnumsent	Numeric	12	0	totnumsent	None	None	12	Right	Scale	Input
19	totnumquest	Numeric	12	0	totnumquest	None	None	12	Right	Scale	Input
20	pcentsentqu...	Numeric	12	2	pcentsentquest	None	None	12	Right	Scale	Input
21	freadave	Numeric	12	5	freadave	None	None	12	Right	Scale	Input
22	turnbysent	Numeric	8	2	lengutt	None	None	8	Right	Scale	Input

Appendix G Poster pilot 2: (Haines 2010)

Assessing participation in PBL using corpus analysis

Catherine Haines and Reg Dennick, Medical Education Unit , University of Nottingham, NG7 2UH, UK
(catherine.haines@nottingham.ac.uk)

The University of Nottingham

Introduction

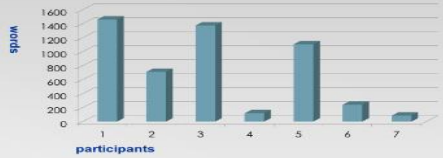
It would be useful to evaluate the oral contribution that individuals make to Problem-based learning sessions. This could be fed back to learners and facilitators and could help to improve the quality of learning generated by PBL.

Method

We transcribed an audio recording of a PBL cycle, consisting of 3 sessions, totaling approximately 5 hours, from students at the University of Nottingham, in Derby. Participant contributions were identified by examining a video of the same cycle.

Results

Session 1: The problem, is analysed and Learning Objectives are generated.



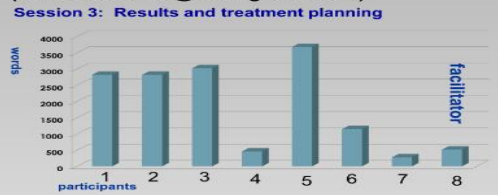
Participation in this session is characterised by 'process' talk to establish understanding of the case and what the group wishes to learn. The facilitator speaks very little to keep the group on task. Speakers 4, 7 and 6 make fewer contributions.

Session 2: Explanations are given , the results of investigations considered and relevant anatomy is discussed.



The same pattern persists during the longest, teaching orientated session.

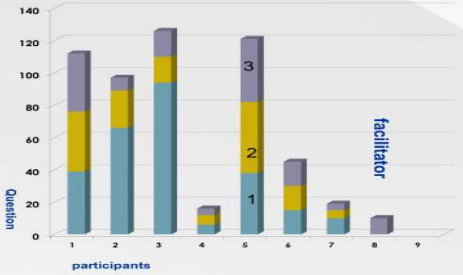
Session 3: Results and treatment planning



Participation in the final session discussing treatment management and the learning points of the cycle shows more participation by the facilitator, to ensure learning points are covered, and the highest contribution by previous lower quantity contributors.

Questions: evidence of reasoning?

The initial study focuses on simple quantity: word count and utterance. When we consider the proportion of utterances which are questions we may begin to identify an early indicator for reasoning skills, and an indicator of the quality of the participants' contributions.

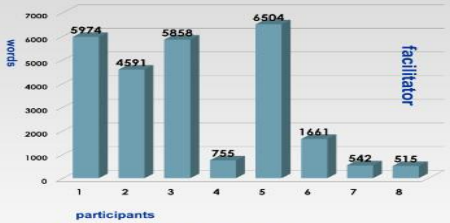


Utterances per individual, showing proportion which are questions, by PBL session, 3, 2 or 1

The big picture

The final picture shows the relative contributions of each participant for the complete cycle. This pilot study will now progress to comparing participation over ten cycles of PBL, tracking participation by individuals in different groups and over time. It will use corpus analysis to assess patterns and frequency of medical word use, frequency of questions and indications of more sophisticated clinical reasoning and learning.

Entire PBL Cycle: word count per participant



Take Home Message:

This pilot study of one PBL cycle, establishes that studying the individual contributions of participants is feasible and uncovers significant differences in participation in PBL which merits further study.

Studying PBL in such detail helps us to understand the mechanism by which participants learn through discussion. It can help groups assess their effectiveness, and can identify dys-functional groups and individuals and help their facilitators assess contributions more effectively.

References

Da Silva, A and Dennick, RG (2010) Corpus Analysis of problem-based learning transcripts: an exploratory study. Medical Education 44, 280-288

Appendix H: Blank Kolb inventory

1

Kolb 3.1 Learning-Style Inventory

The Learning-Style Inventory describes the way you learn and how you deal with ideas and day-to-day situations in your life. Below are 12 sentences with a choice of endings. Rank the endings for each sentence according to how well you think each one fits with how you would go about learning something. Try to recall some recent situations where you had to learn something new, perhaps in your job or at school. Then, using the spaces provided, rank a "4" for the sentence ending that describes how you learn *best*, down to a "1" for the sentence ending that seems least like the way you learn.
Be sure to rank all the endings for each sentence unit. Please do not make ties

	A		B		C		D	
1. When I learn:	___	I like to deal with my feelings.	___	I like to think about ideas.	___	I like to be doing things.	___	I like to watch and listen.
2. I learn best when:	___	I listen and watch carefully.	___	I rely on logical thinking.	___	I trust my hunches and feelings.	___	I work hard to get things done.
3. When I am learning:	___	I tend to reason things out.	___	I am responsible about things.	___	I am quiet and reserved.	___	I have strong feelings and reactions.
4. I learn by:	___	feeling.	___	doing.	___	watching.	___	thinking.
5. When I learn:	___	I am open to new experiences.	___	I look at all sides of issues.	___	I like to analyze things, break them down into their parts.	___	I like to try things out.
6. When I am learning:	___	I am an observing person.	___	I am an active person.	___	I am an intuitive person.	___	I am a logical person.
7. I learn best from:	___	observation.	___	personal relationships.	___	rational theories.	___	a chance to try out and practice.
8. When I learn:	___	I like to see results from my work.	___	I like ideas and theories.	___	I take my time before acting.	___	I feel personally involved in things.
9. I learn best when:	___	I rely on my observations.	___	I rely on my feelings.	___	I can try things out for myself.	___	I rely on my ideas.
10. When I am learning:	___	I am a reserved person.	___	I am an accepting person.	___	I am a responsible person.	___	I am a rational person.
11. When I learn:	___	I get involved.	___	I like to observe.	___	I evaluate things.	___	I like to be active.
12. I learn best when:	___	I analyze ideas.	___	I am receptive and open-minded.	___	I am careful.	___	I am practical.

Remember: **4 =** *most* like you **3 =** *second most* like you **2 =** *third most* like you **1 =** *least* like you

Thesis word count: 99,688