VALIDITY THEORY (VT) AND MODERN TEST THEORY (MTT) IN ASSESSMENT MEASUREMENT AND DESIGN

by

SUSAN. A. BLYDEN

A thesis submitted to the University of Birmingham, in partial fulfilment of the requirement for the degree of
DOCTOR OF EDUCATION (EdD).
Learning and Learning Contexts

October 2019
School of Education
The University of Birmingham
Edgbaston, Birmingham
B15 2TT, UK

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

Background: Healthcare education use assessment results (test scores) to make inferences about students' knowledge and ability. The NMP curriculum, therefore, assumes a standard of proficiency achievement in successive student cohorts. Therefore, these assessments are high-stakes and must be valid as they may have implications for patients' safety.

Aims: This study aimed to apply Kane's (2006) validity theory to evidence whether the assessments capture the proficiency it purports to assess and modern test theory to measure the appropriateness of the score interpretation.

Method: Theory-Driven Thematic Analysis (TDTA) was used to code and content analyse the reflections; Ryle (1949), Mezirow (1981), Benner (2001) and Bloom's (1956) theories instructed the content validity, quality and evidenced mastery of the NMP subject and theory domains. The Structured Clinical Examination (SCE) criterion was analysed using the Rasch one-parameter logistic model (1-PL) to measure relationships between the test item criteria, the student ability and test-items parameters capacity to discriminate among students with varying ability.

Results: Representation of Ryle's, Mezirow's, Benner's, and Bloom's theories in the reflections were at the lower levels of achievement. The SCEs, criterion analyses showed the test item favoured students with limited ability, indicating an easy test; overall, suggesting the assessments' validity is compromised.

Conclusion: Applying validity and measurement theory to assessed work could elucidate the meaning of assessment scores and ensure that assessments are appropriate to the examinees' level of ability.

Dedication

I give the highest honour and praise to my Lord and Saviour, Jesus, the Christ.
John 3:16-The Holy Bible

To my mother for her consistent prayers and encouragement to pursue the professional and academic milestones I aspired to achieve.

Acknowledgements

Acknowledgements to my friends, family members, and colleagues, I wish to thank you for your support during the journey toward completing this study. I am particularly grateful to the following people.

I extend thanks to Dr Ian Davison in the School of Education for his supervision, particularly his constructively critical comments on this thesis's draft works, which was appreciated. I advance my appreciation to Dr Donna Dawkins for her remarks on my work which helped improve the thesis's quality and presentation.

I thank my colleague, Alison, the course director for the Non-Medical Prescribing (NMP) program for allowing the use of the NMP curriculum as a vehicle for this study, and the NMP students for their consent to use their assessments as data. To my external colleague Lindsey, thank you for generously giving valuable time to intra-rating the coding used in this study for the content analysis. To Jean Astley-Cooper and Caroline Green, who for a short period over the thesis's duration organised those much-needed breaks away in Europe.

I am grateful to my close friends Penny and Davin for their excellent hospitality and for listening to my musings about this thesis's subject, which helped formulate my thoughts and ideas. To Steph and Andrew, for your kindness, support and for kindling my creativity with a continued provision of 'chocolates' during my journey toward completing this Doctorate. Finally, a heartfelt thank you to my sister in Christ and dear friend Marcia "Manty" Mantack, who has been my biggest cheerleader for many years in my professional and academic endeavours.

OUTLINE OF THE CHAPTERS

Chapter 1: Introduction

Chapter 1 begins with a historical overview of non-medical prescribing, including the instructional design used for the curriculum content and delivery. The Research Problem, Statement and Question are given, and a discussion of the conceptual and theoretical frameworks pertinent to this research. Several theoretical frameworks relevant to the test measurements are presented, clarifying the thinking and planning behind the data collection and analysis, and explaining the findings in answer to the research questions. This chapter closes with a summary of the chapter contents.

Chapter 2: Literature Review

The review focuses on the literature in healthcare education and education, which contributed to understanding the measurement strategies used for students' assessed work. This study's challenge was identifying the available measurement methods and their ability to make the connection between student performance and the test scores to provide appropriate evidence to support their inference and use. The reviewed studies highlighted that any scheme or framework used to measure students' assessed work is likely to be imperfect.

Chapter 3: Methodology

Chapter 3 contains the justification for the research design, methodology and methods, the ethical considerations, qualitative and quantitative data collection and analyses. Outlining the process for code development using Theory-Driven Thematic Analysis (TDTA) mapped against the NMC (2006, 2018) standards for proficiency for prescribing, together with the frameworks of Ryle (1949), Mezirow (1981), Benner (2001), and Bloom (1956) in the TDTA analysis. Its use for the coding and text mining of the content of sampled N=28 written reflections (WRs) is discussed. Also discussed is the purpose and application of Item Response Theory (IRT) and Rasch analysis and the implications of Modern Test Theory (MTT) for Structured Clinical Examination (SCE) data analysis. The data analysis software package in R for text mining R-QDA is highlighted and includes R-itm's illustrative examples of the command programming language used for the Rasch analysis calculations.

Chapter 4: Written Reflection Data Analysis

This chapter presents the results of the content analysis of the written reflections using TDTA. Samples of the students' text illustrate their application of the theoretical frameworks of Ryle, Mezirow, Benner and Bloom to the N=28 sampled assessed work. These theories are utilised in examining the content validity of reflections and the student's ability to address the NMP subject domains.

Chapter 5: Analysis of the Structured Clinical Examination (SCE) Data

Chapter 5 offers the analysis of the SCEs criterion measurement results, specifically, Item Analysis (IA), applying the Rasch 1PL model for the measurement of the SCE's item difficulty parameters and capacity to discriminate between students' with high and low ability (traits). The results provided invaluable insight into the SCE item criteria' performance and its scope to capture and measure the desired NMP criteria.

Chapter 6: Conclusions, Limitations and Recommendations

This final chapter opens with a review of the results, followed by an overview of the chapter and the interpretive validity argument as applied in this study. This chapter discusses the relevance and representativeness of the assessments in capturing student performance and level of ability in the assessment activity in the context of the patient's safety. This chapter ends with the conclusion, recommendations, and limitations of the study, closing with comments on the contribution to education and future research suggestions.

Table of Contents

	Page no
Abstract	ii
Dedication	iii
Acknowledgements	iv
The Chapters Content	v
Table of Contents	vi
Glossary of Terms	xii
CHAPTER 1: INDEPENDENT NON-MEDICAL PRESCRIBING (INMP) EDUCATION	1
1.0 Introduction	1
1.1 Background to Non-Medical Prescribing (NMP)	1
1.1.2 Prescribing Policy and Clinical Governance	2
1.1.3 Independent NMP Education, Curriculum Design and Content	4
1.1.3.1 Continuing Professional Development (CPD).	4
1.1.4 Competence and the Competency Model: The Structure for NMP Teaching and Learning	6
1.1.4.1 Competency and Proficiency: Defining the Constructs	7
1.1.4.2 The Relationship Between Competence and Proficiency	9
1.1.4.3 The Limits of Competence	10
1.2 The Research Problem, Statement, and Question	11
1.2.1 The Significance of the Problem	11
1.2.2 The Problem Statement and Research Question(s)	13
1.2.3 The Researcher's Orientation	14
1.3. The Theoretical Frameworks Underpinning the NMP Assessment	15
1.3.1 Theory, Practice and Pedagogy	15
1.3.1.1 Teaching and Learning Theories	16
1.3.1.2 Classical Test Theory (CTT) in Assessment Measurement	16
1.3.2 Validity Theory (VT)	19
1.3.3 The Conceptual Framework: Kane's Interpretive Validity Argument	20
1.3.4 Modern Test Theory (MTT), for Substantiating Kane's IVA	23
1.4 Knowledge, Theory and Ways of Knowing: The Concept of Mastery in Education	28
1.4.1 Ryle and Knowledge Theory	30
1.4.2 Theories of Reflection: An Overview	33
1.4.3 Mezirow's (1981) Seven Concepts of Reflection	34
1.4.4 Benner's Novice to Expert Model	36
1.4.5 Bloom's Taxonomy of Educational Objectives: Cognitive Domain	38
1.4.5.1 Chapter Summary	40
1.4.5.2 Conclusion	40
CHAPTER 2: THE LITERATURE REVIEW	42
2.0 Chapter Overview	42
2.1 Introduction	42
2.1.1 The Review Aims	43
2.1.2 The Review Methodology, Design and Methods	43
2.1.2.1 The Search Strategy	44
2.1.2.2 Search Terms	45
2.1.2.3 The Inclusion-Exclusion Criteria	45
2.1.2.4 Search Results	47
2.1.3 The Review Design	49

2.2 The Literature Review 52
2.2.1 Background: Educational Measurement 52
2.2.1.1 Studies Examining Written Reflection (WR) 53
2.2.1.2 Studies Examining Structured Clinical Examination (SCE) 58
2.2.3 The Discussion 65
2.2.4 What these Studies Highlighted 66
2.2.5 How the Review Helped to Refine the Methodology and Methods 67

CHAPTER 3: THE RESEARCH METHODOLOGY, DESIGN AND METHODS 69

3.0 Chapter Overview 69
3.1 Reflexivity 69
3.1.1 Research Design 69
3.1.2 Establishing Rigour in the Case Study Design 71
3.1.2.1 The Research Methodology and Methods 72
3.1.2.2 Sampling 72
3.1.3 Ethics 72
3.1.4 The Recruitment Process 73
3.1.5 Data Collection 74
3.2 Data Analyses Methodology 75
3.2.1 Developing the Codes for Content Analysis 75
3.2.2 Cohen's Kappa Statistic for Inter-rater Reliability 76
3.2.3 Content Analysis of the Written Reflections (WRs) 77
2.2.4 Text Mining using R-QDA 78
3.3. The Structured Clinical Examination (SCE) 79
3.3.1 The Rasch Model for Analysis: The Purpose and Function 79
3.1.1.1 Interpreting the Test Item Characteristic Curves (ICC) 81
3.3.1.2 The Item Difficulty (DIF) Values 82
3.3.1.3 The Item Information Function (IIF) Values 84
3.3.1.4 The Test Information Function (TIF) Values 85
3.3.1.5 The Implication of Rasch Modelling in Modern Test Theory 87
3.3.1.6 Quantitative Data Analysis using R-itm 87
3.3.1.7 The Goodness-of-fit Command for Rasch 89
3.4 The Structure of the Interpretive Validity Argument (IVA) 90
3.4.1 The Theories as Evidence Measure in the IVA 91

CHAPTER 4: THE QUALITATIVE DATA ANALYSIS FOR THE SUMMATIVE WRITTEN REFLECTIONS. 93

4.0 Introduction 93
4.1.1 The Results and Findings 93
4.1.1 The Non-Medical Prescribing (NMP) Subject Domains: Theme 01- The Consultation. 95
4.2 Ryle's (1949) Knowledge Theory: Declarative (Know-that) Knowledge 96
4.2.1 TDTA code 05 The Consultation 96
4.2.2 TDTA Code 01 Allergies 97
4.2.3 NMP Subject Domain -Theme 02: Establishing Options (Patient Safety) 98
4.2.3.1 Knowledge Theory: Procedural (Know-how) Knowledge. 99
4.2.3.1.1 TDTA code 25: The Physical Examination 99
4.2.3.1.2 TDTA code 34: Requesting and Interpreting Diagnostic Tests 101
4.3 NMP Subject Domain- Theme 03: Communicating with Patients 103
4.3.1 Knowledge Theory: Strategic (Know-Why) Knowledge 104

4.3.1.1 TDTA code 42: Deals Sensitively with Patient 105
4.3.1.2 TDTA code 50: Non-Drug Treatment Options 106
4.4 Mezirow's (1981)Seven Concepts of Reflection 107
4.4.1 Mezirow's 1st Concept: Reflectivity 110
4.4.1.1 Mezirow's 2nd Concept: Affective Reflectivity 111
4.4.1.2 Mezirow's 3rd Concept: Discriminant Reflectivity 113
4.4.1.3 Mezirow's 4th Concept: Judgemental Reflectivity 114
4.4.1.4 Mezirow's 5th Concept: Conceptual Reflectivity 115
4.4.1.5 Mezirow's 6th Concept: Psychic Reflectivity 116
4.4.1.6 Mezirow's 7th Concept: Theoretical Reflectivity 116
4.5 Benner's (2001) Novice to Expert Model 117
4.5.1. Novice: Level 1 118
4.5.1.1 Advanced Beginner: Level 2 119
4.5.1.2 Competence: level 3 121
4.5.1.3 Proficient: Level 4 124
4.5.1.4 Expert: Level 5 125
4.6 Bloom's (1956)Taxonomy of Educational Objectives: The Cognitive Domains 125
4.6.1 Bloom's Taxonomy Level 1, Knowledge 126
4.6.1.1 Bloom's Taxonomy Level 2, Comprehension 128
4.6.1.2 Bloom's Taxonomy Level 3, Application 131
4.6.1.3 Bloom's Taxonomy Level 4, Analysis 132
4.6.1.4 Bloom's Taxonomy Level 5, Synthesis 134
4.6.1.5 Bloom's Taxonomy Level 6, Evaluation 134

CHAPTER 5: THE QUANTITATIVE DATA ANALYSIS OF THE SUMMATIVE STRUCTURED CLINICAL EXAMINATION 135

5.0 Chapter Overview 135
5.1 Introduction 135
5.1.1 The Sampled SCE 136
5.1.2 The Goodness-of-Fit Results 137
5.2 The Data Analyses Results: The Mandatory SCE Test Item Criteria 137
5.3 Legal Prescription Writing 141
5.4 Generic SCE Test Items: Decision Making and Rationale 146
5.5 Generic SCE Test Items: Consultation and History Taking 149
5.6 Generic SCE Test Items Continued: Consultation and History Taking 152
5.7 Item Discrimination (ID) Parameter 156

CHAPTER 6: THE DISCUSSION, CONCLUSION AND RECOMMENDATIONS 158

6.0 Introduction 158
6.1 Summary of the WR Results 158
6.1.1 Summary of the SCE Results 159
6.1.2 Re-stating the Research Problem 159
6.1.2.1 The NMP Assessments Validation Process 161
6.1.3 The Four Theories as Evidence Measures in Kane's (2006) IVA 162
6.1.4 Chapter Overview 163
6.2 DISCUSSION PART 1 163
6.2.1 Professional Education and Accountability in the Profession 163
6.2.1.1 The Role of Authentic Assessment in Professional Education 164
6.2.2 Knowledge, Professionalism and Professional Practice 165

6.2.3 The Rhetoric of Reflection 169
6.2.4 The Standard Practice and Formal Academic Standards 172
6.2.4.1 The Practice Level for Advanced Practice 172
6.2.4.2 The Academic level for Advanced Practice 176
6.3 DISCUSSION PART 2 178
6.3.1 Introduction 178
6.3.2 The SCE Quality and Construct 178
6.3.3 Asking the Right Questions in the Right Way 182
6.3.4 Psychometric Item Analysis (IA): Evidencing the Test Scores 188
6.4 CONCLUSION 190
6.4.1 Recommendations 191
6.4.1.1 Recommendation 1 192
6.4.1.2 Recommendation 2 192
6.4.1.3 Recommendation 3 193
6.4.2 Study Limitations 193
6.4.2.1 An Alternative to the Case Study Design 193
6.4.2.2 Secondary (Post-Hoc) Data Use 194
6.4.2.3 Bracketing- Refocusing on the Data 195
6.4.2.3.1 Contextualisation 195
6.2.3.4 Transferability 196
6.5 Contribution to Nurse Education and Suggestions for Further Research 196
6.5.1 Closing Remarks 197

LIST OF FIGURES

Figure 1.3.3 A Visual Representation of the application of Kane's Interpretive Validity Argument (IVA) 22
Figure 1.3.3.1 An Overview of the Validity Theories 23
Figure 1.3.4 A Dialectical (inferential/theoretical) Link Between MTT and VT for Test Validity 25
Figure 1.3.4.1 The Research Question[s]: Validity Theory and Data Analyses in Summary 27
Figure 2.1.2.4. A Flow Diagram of the Search Process for the WRs 47
Figure 2.1.2.4.1. A Flow Diagram of the Search Process for the SCEs 48
Figure 3.3.1.1 The ICC Exemplar for the Single Test Item 81
Figure 3.3.1.2 The Example ICC for a Single Test Item, Illustrating the Test Item Difficulty (DIF) 83
Figure 3.3.1.3 The Example ICC for a Single Test Item, Illustrating the Test Item Information Function (IIF) 85
Figure 3.3.1.4 The Example ICC for a Single Test Item, Illustrating the Test Item Information (TIF) 86
Figure 3.4 The Application and Structure of Kane's (2006) IVA: The Teaching and Learning Theories used as Evidence 87
Figure 4.4 Mezirow's (1981) Seven Concepts of Reflectivity, Results in Summary 108
Figure 4.4.1 Mezirow's Concept 1: Results in Summary 109
Figure 4.4.1.2 Mezirow's Concept 2: Results in Summary 110
Figure 4.4.1.3 Mezirow's Concept 3: Results in Summary 114
Figure 4.4.1.4 Mezirow's Concept 5: Results in Summary 115
Figure 4.5 Benner's (2001) Novice to Expert Model, Results in Summary 117
Figure 4.5.1 Benner's Criterion, Level 1: Results in Summary 118
Figure 4.5.1.1 Benner's Criterion, Level 2: Results in Summary 120
Figure 4.5.1.2 Benner's Criterion, Level 3: Results in Summary 122
Figure 4.5.1.3 Benner's Criterion, Level 4: Results in Summary 124
Figure 4.5.1.4 Benner's Criterion, Level 5: Results in Summary 124
Figure 4.6 Bloom's (1956) Taxonomy of Educational Objectives: The Cognitive Domain, Results in Summary 126
Figure 4.6.1 Bloom's Taxonomy level 1: Knowledge, Results in Summary 127
Figure 4.6.1.1 Bloom's Taxonomy level 2: Comprehension, Results in Summary 129
Figure 4.6.1.2 Bloom's Taxonomy level 3: Application, Results in Summary 131
Figure 4.6.1.3 Bloom's Taxonomy level 4: Analysis, Results in Summary 133
Figure 5.2 The Item Characteristic Curve (ICC) Decision Making and Rationale for the N=28 Sampled 139
Figure 5.2.1 The Item Information Function (IIF) for Decision Making and Rationale 140
Figure 5.2.1.1 The Test Information Function (TIF) for Decision Making and Rationale 141
Figure 5.3.1 The ICC for Mandatory Prescription Writing for the N=28 sampled SCEs 144
Figure 5.3.1.1 The IIF for Mandatory Prescription Writing for the N=28 sampled SCEs 145
Figure 5.3.1.2 The TIF for Mandatory Prescription Writing for the N=28 sampled SCEs 146
Figure 5.4 The ICC for Decision Making and Rationale for the N=28 sampled SCEs 148
Figure 5.4.1 The IIF for Decision Making and Rationale for the N=28 sampled SCEs 148
Figure 5.4.1.1 The TIF for Decision Making and Rationale for the N=28 sampled SCEs 149
Figure 5.5 The ICC for the Generic Test Items for the N=28 sampled SCEs 150
Figure 5.5.1 The IIF for the Generic Test Items for the N=28 sampled SCEs 151
Figure 5.5.1.1 The TIF for the Generic Test Items: Consultation and History Taking for the N=28 sampled SCEs 152

Figure 5.6. The ICC for the Generic Test items: Consultation and History Taking Continued for the N=28 sampled SCEs 154
Figure 5.6.1 The IIF for the Generic Test items: Consultation and History Taking Continued -in the N=28 sampled SCEs 155
Figure 5.6.1.1 The TIF for the Generic Test items: Consultation and History Taking Continued -in the N=28 sampled SCEs 156

LIST OF TABLES

Table 1.3.3 Illustration of Kane's of Kane's (2006) Interpretive Validity Argument (IVA) 21
Table 1.3.4 Illustration of Representational Theory Measurement (RTM) 26
Table 1.4.1 A Summary (Ryle's (1949) of Knowledge Theory 31
Table 1.4.3 Mezirow's (1981) Seven Concepts of Reflection 35
Table 1.4.4 Benner's (2001) Novice to Expert Model 37
Table 1.4.5 Bloom's (1959) Taxonomy of Educational Objectives: The Cognitive Domain 39
Table 2.1.2 Formulating the Research Question 43
Table 2.1.2.2 Search Terms used for the Peer-reviewed Studies 45
Table 2.1.2.3 Inclusion-Exclusion criteria for the WRs 46
Table 2.1.2.3.1 Inclusion-Exclusion criteria for the SCEs 46
Table 2.1.3 Summary of Studies Examining the Written Reflections (WRs) 50
Table 2.1.2.3.1 Summary of Studies Examining the Structured Clinical Examination (SCE) 51
Table 3.2.2 Cohen's Kappa Values and Interpretation 77
Table 3.3.1.6 The Goodness-of-Fit Command for Rasch Analysis in R-*itm* 88
Table 4.1 Cohen's Kappa Results for 2 Independent Raters 94
Table 4.1.1 TDTA Code Key for Theme 01: NMP Subject Domain-The Consultation 95
Table 4.2.1 Code 5: The consultation, Results in Summary 96
Table 4.2.2 Code 1: Allergies, Results in Summary 97
Table 4.2.3 Code Key for TDTA Theme 02: NMP Subject Domain Establishing Options (Patient Safety) 99
Table 4.2.3.1.1 Code 25: Physical Examination (PE), Results in Summary 100
Table 4.2.3.1.2 Code 34: Requesting and Interpreting Tests 102
Table 4.3 TDTA Code Key for Theme 03: NMP Subject Domain- Communicating with Patient's 104
Table 4.3.1.1 Code 42: Deals Sensitively with Patient, Results in Summary 105
Table 4.3.1.2 Code 50: Non-Drug Treatment Option 106
Table 5.1.2 The Goodness-of-Fit Results for the Non-Mandatory SCE Questions for the N=28 Sampled 136
Table 5.1.2.1 The Goodness-of-Fit Results for the Mandatory SCE Questions for the N=28 Sampled 136
Table 5.1.2.2 The Goodness-of-Fit Results for All SCE Question Items from the N=28 Sampled 137
Table 5.2 The Mandatory SCE Test Items- Decision-making and Rationale for the N=28 Sampled 138
Table 5.3. The Mandatory SCE Test Items-Prescription Writing for the N=28 Sampled 142
Table 5.3.1. The Mandatory SCE Test Items- Prescription Writing Continued for the N=28 Sampled 143
Table 5.4. The Generic SCE Test Items: Decision Making and Rationale for the N=28 Sampled 147
Table 5.5 The Generic SCE Test Items: Consultation and History Taking for the N=28 sampled 150
Table 5.6 The Generic SCE Test Items: Consultation and History Taking continued, for the N=28 sampled 153
Table 6.3.2 A Sample of the SCE Target Domains for Consultation 179
Table 6.3.2.1 The Proposed SCE Scorning: Benner's Model illustrated 180
Table 6.3.3 The Current SCE Item Checklist for Legal Precision Writing 184
Table 6.3.3.1 The Proposed SCE using a Global Rating Scale for Legal Prescription Writing 187

ENDNOTES 198

APPENDICES 202

Appendix 1 Ethics Approval 203
Appendix 2 Consent Form 206
Appendix 3 Study Information Sheet 207
Appendix 4 The Written Reflection (WR), Assignment Brief 210
Appendix 5 The Structured Clinical Assessment (SCE) 212
Appendix 5a The SCE Assessed Work Sample 215
Appendix 6 The TDTA Codes, NMP Themes and Categories 218
Appendix 6a Ryle's TDTA Codes 221
Appendix 6b Mezirow's TDTA Codes 222
Appendix 6c Benner's TDTA Codes 223
Appendix 6d Bloom's TDTA Codes 224
Appendix 6e The Coded WR, Assessed Work Sample 225
Appendix 7 The R-itm command codes for the Rasch Analysis 230
Appendix 8 The Confidence Intervals and Cohen's Kappa Results 240

LIST OF REFERENCES 251

Glossary of Terms

AA	Authentic Assessment
AB	Advanced Beginner
ADE	Adverse Drug Event
ADR	Adverse Drug Reaction
AEI	Approved Education Institution
AKI	Acute Kidney Injury
ALERT	Acute Life-Threatening Event Recognition and Treatment
ANOVA	Analysis of Variance
AP	Advanced Practice
APP	Assessments of Physiotherapy Practice
APs	Advanced Practitioners
AS	Authentic Assessment
BIC	Bayesian Information Criterion
BNF	British National Formulary
BMD	Bone Mineral Density
BP	Blood Pressure
CA	Content Analysis
CAT	Critical Appraisal Tool
CBT	Cognitive Behavioural Therapy
CC	Critical Consciousness
CCU	Critical Care Unit
CFA	Confirmatory Factor-Analysis
CK	Content knowledge
CKP	Critical Knowledge for Practice
CM	Classical Measurement
CI	Confidence Intervals
CID	Code Identifier
CPD	Continuing Professional Development
CPR	Cardiopulmonary Resuscitation
CT	Critical Theory
CTT	Classical Test Theory
CHI	Commission for Health Improvement
CMP	Clinical Management Plan
CST	Communication Skills Training
CRD	Centre for Reviews and Dissemination
CRT	Criterion Reference Test
CV	Content Validity
DA	Distractor Analysis
DH	Department of Health
DI	Drug Interaction
DIF	Differential Item Functioning
DIUS	Department for Innovation, Universities and Skill
DSA	Disease Severity Assessment
DXA	Dual-energy X-ray Absorptiometry
EBP	Evidence-Based Practice
ECG	Electrocardiogram
EFA	Exploratory Factor Analysis
EHEA	European Higher Education Agenda
ELT	Experiential Learning Theory
ENB	English National Board

EQF	European Qualifications Framework
EU	European Union
FA	Factor Analysis
FACP	Framework for Advanced Clinical Practice
FHLS	Faculty of Health and Life Sciences
FtR	Failure to Rescue
GP	General Practitioner
GMC	General Medical Council
GRS	Global Rating Scores
GT	Grounded Theory
GUI	Graphics User Interface
HC	Human Capital
HE	Higher Education
HEE	Health Education England
HEI	Higher Education Institute
HF	Heart Failure
HCP	Health Care Professionals
HCPC	Health Care Professionals Council
HOT	Higher-Order Thinking
HRT	Hormone Replacement Therapy
HSCA	Health and Social Care Act
IA	Item Analysis
IC	Informed Consent
ICC	Item Characteristic Curve
ID	Item Discrimination
IIF	Item Information Function
IP	Independent Prescribing
IPs	Independent Prescribers
INMP	Independent Non-Medical Prescriber
IoP	Index of Performance
IRA	Interrater Agreement
IRC	Interrater Correlation
IR	Item Response
IRR	Inter-rater Reliability
IRT	Item Response Theory
IVA	Interpretive Validity Argument
IAU	Interpretive Argument and Use
K (𝜅)	Cohen's Kappa
KDIGO	Kidney Disease: Improving Global Outcomes
KM	Knowledge Management
KMAS	Knowledge, Motor, Attitude, Structure
KSF	Knowledge and Skills Framework
KSP	Knowledge Skills and Proficiency
KWL	**K**- what we know; **W**- what we want to find out; **L**- what we learned and still need to learn
LCJR	Lasater Clinical Judgement Rubric
LO and LOs	Learning Outcome(s)
LTT	Latent Trait Theory
LT	Latent Trait
LTA	Latent Trait Analysis

MDT	Multi-Disciplinary Team
MFRM	Multi-Faceted Rasch Model
MHRA	Medicine and Health Products Regulatory Agency
MH	Medications History
ML	Master's Level
MM	Matrix Method
MOD	Multiple Organ Dysfunction
MSA	Measurement of Sample Adequacy
MnSq	Mean Square
MT	Measurement Theory
MTT	Modern Test Theory
NCEPOD	National Confidential Enquiry into Patient Outcomes and Death
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NIP	Nurse Independent Prescriber
NMC	Nursing and Midwifery Council
NMP	Non-Medical Prescribing
NMPs	Non-Medical Prescriber[s]
NPSA	National Patient Safety Agency
OECD	Organisation for Economic Co-operation and Development
OR	Odds Ratio
OSCE	Objective, Structured Clinical Examination
OTC	Over the Counter Medication
PAC	Principle Axial Factoring
PCC	Patient-Centred Care
PCK	Pedagogical Content Knowledge
PE	Physical Examination
PEO	Population Exposure Outcome
PICO	Population, Intervention, Comparators, Outcomes
PISA	Programme for International Student Assessment
PLD	Performance Level Descriptors
P1PN	Procollagen type 1, N-propeptide
PPP	Patients Pharmacological Profile
PQRI	Portfolio Quality Rating Inventory
PRHO	Pre-Registration House Officer
PS	Patient Safety
PT	Psychometric Theory
PTT	Prothrombin Time
P-VALUE	Probability value
PW	Prescription Writing
QA	Quality Assurance
QAA	Quality Assurance Agency
QR	Quantitative Reasoning
RCN	Royal College of Nursing
RJ	Reflective Judgement
RFT	Respiratory Function Test
RM	Rasch Measurement
RPS	Royal Pharmaceutical Society
RQDA	R- Qualitative Data Analysis
RTM	Representational Theory Measurement

SAE	Serious Adverse Effects
SCE	Structured Clinical Exams
SHA	Strategic Health Authority
SId	Student Identifier
SMK	Subject Matter Knowledge
SPSS	Statistical Package for Social Sciences
St	Student
TA	Thematic Analysis
TC	Teacher Candidates
TDTA	Theory-Driven Thematic Analysis
TIF	Test Information Function
ToC	Tests of Competence
TV	Test Validity
VT	Validity Theory
VE	Validity Evidence
WHO	World Health Organisation
WR	Written Reflection

CHAPTER 1: INDEPENDENT NON-MEDICAL PRESCRIBING (NMP) EDUCATION

1.0. Introduction

The Independent Non-Medical Prescribing (INMP), hereafter (NMP) assessment, derives inferences from the test scores about students' abilities. The inferences concern students' academic performance and the standard of professional achievement presumed as achieved, including accountability, proficiency levels, and higher-order thinking (HoT) skills (Fry et al., 2009). However, these conclusions may not be accurate. The inferences arise from a chain of reasoning, beginning with assessment score data accepted as representative of student performance, professional achievement standards, and proficiency. The chain of reasoning ends with decisions about students' readiness for entry into NMP practice. As such, these assessments are high-stakes tests, as the pass scores infer that students are NMP proficient and subsequently ready for certification/ licensing for practice (Norcini, 1994; Sackett et al., 2008; Francis, 2013). Therefore, these tests require valid evidence of what is measured and achieved by the students in theory and practice (Linn et al., 1991; Norcini, 2004; Sackett et al., 2008; Sadler et al., 2009a). This study aims to provide supporting evidence regarding the validity of the test scores and the proficiency inferences related to the NMP assessments to determine the assessment processes' quality and intrinsic value. The NMP assessments' validity research evidence will inform the course team concerning the current NMP assessments' performance, used for redesign or utilised in future NMP coursework development.

1.1.0 Background to Non-Medical Prescribing (NMP)

The Cumberlege Report (1986) recommended extending prescribing rights to healthcare disciplines outside of the medical profession. The Crown Report (1998) and the subsequent Crown Review (1999) advocated for professions other than doctors should initiate and prescribe drugs. The Department of Health (DH, 2005) National Health Service (NHS) Modernisation Agenda initiative embraced the general concept of NMP. In 2005, nurses acquired limited, supplementary prescribing authority, and in 2006, Independent Prescribing (IP) authority. Further legislation

extended NMP to other Health Care Professionals (HCPs), citing a workforce crisis, specifically a shortage of doctors in many NHS Trusts (DH, 2006a). The National Health Service (NHS) workforce crisis spearheaded the Advanced Practice (AP) evolution and AP role inception. AP is not limited to nurses but includes a range of HCPs such as pharmacists, chiropodists, physiotherapists and radiographers (Sibbald et al., 2006; DH, 2008a; 2009; 2010). APs are now firmly embedded in the healthcare workforce, directing nurses and the HCPs' career progression towards the uptake of AP responsibilities, including NMP (DH, 2010).

1.1.2 Prescribing Policy and Clinical Governance

The legislation covering the legal and jurisdictive basis for NMP education and practice are:- The Nursing and Midwifery Council (NMC), which legislate the professional standards for safe, ethical and quality care (2006, 2018). The Medicines Act (1968; 1998), section 58; the Medicines Products: Prescriptions for Nurses Act (1992); and Medicines for Human Use (Marketing Authorisations, Etc.) Regulations of Health and Social Care Act (HSCA, 2001), section 63. There were incremental policy changes to the HSCA (2003; 2011) and the Medicines for Human Use Act (MHUA 2004, amended in 2006). These changes increased the range of medicines and medical conditions where nurses, midwives, health visitors (HVs), and pharmacists could prescribe. Further amendments, MHUA (2009:2), allowed designated NMPs to prescribe unlicensed medicines in all clinical areas (Payne, 2010). The prescribing governance framework defines two critical types of prescribers: 'supplementary' and 'independent.' The DH (2005a) and the Medicine and Healthcare Products Regulatory Agency (MHRA, 2006) provided the following descriptors: The MHRA defines independent NMPs as

> "A practitioner [is] responsible and accountable for the assessments of patients with undiagnosed or diagnosed conditions and for decisions about the clinical management required, including prescribing.," (MHRA, 2006: 5).

The DH defines supplementary NMPs prescribers as

> "A prescriber… [Is] responsible for the continuing care of patients who have been clinically assessed by an independent (medical or dental) prescriber. They may prescribe for the full range of medical conditions … under the terms of a patient-specific Clinical Management Plan (CMP) and within their sphere of competence",
>
> (DH, 2005a: 2[8]).

Part of the process of achieving independent prescribing was the inception of the NMP program. There is a requirement for a legislated standard of proficiency (Nursing and Midwifery Council, 2006; 2018). The outcome claim of the independent NMP education program is that the students are proficient at the point of completion. Barton et al., (2012) pointed out that proficiency assumes progression beyond the level of competence. Well-developed knowledge and nursing skills combined with newer but equally developed NMP skills are applied to deliver autonomous, proficient independent patient care management. At the point of the nurse's initial qualification and registration, competence is implied. Proficiency, however, is the next progression point, characterised by self-sufficiency, synonymous with an advance in practice, HoT skills and knowledge required for shaping and delivering complex healthcare and prescribing activities (Rolf, 2014). The expectation is that NMPs practice proficiently at the course endpoint to provide clinical judgment beyond skilled performance to speed decision-making in time-critical events (Benner, 2001; Hoffman and Militello, 2008; NMC, 2018). NMPs responsibility includes anticipating prescribing demands, re-planning and acting prudently by logic, reasoning and judgment, rather than by the rule governance of competence (DuHamel et al., 2011; Gallagher et al., 2011; The Kings Fund, 2017).

1.1.3 Independent NMP Education, Curriculum Design and Content

The professional standards of proficiency (NMC, 2006) for prescribing informed the NMP curriculum content at the data collection time. The six central components of the NMP model are (1) the consultation, including medical health assessment, (2) taking a medications history, (3) knowledge of pharmacology and numeracy, (4) effects of drug therapies on other medications, (5) medications' side effects, and (6) working within legal and professional systems. The 21 standards include prescribing and administration (standard 9) and prescribing and dispensing (10). The 2006

standards are congruent with the new NMC (2018) standards of proficiency and the updated Royal Pharmaceutical Society (RPS, 2016) competency framework, which the NMC agreed to adopt. However, where the RPS framework uses the term competency related to effective performance, the NMC (2018) retain proficiency as the standard requirement. Proficiency is central to the development of independence by nurse prescribers in practice to ensure NMPs preparation for the responsibilities they will face as practitioners (McGaghie, 2015). The NMC and RPS standards are used collectively in multiple summative assessments to verify that nursing students' meet the criteria for prescribing (NMC, 2018). Aside from the 2007 course requirement to achieve a pass mark of 80% in pharmacology and 100% in numeracy, there have been no other amendments to the other discrete assessments. For example, the written reflections (WR) and the practical Structured Clinical Examination (SCE). The SCE is a single station assessment, unlike the Objective structured Clinical Examination (OSCE), a multiple station design (Harden et al., 2016).

At the Higher Education Institute (HEI), where the research was conducted, the independent NMP student must undertake an Advancing Practice (AP) education program before or concurrent with their prescribing course (RCN, 2012; Rolf, 2014). The AP pathway in which Independent NMP sits is accredited at master's degree level consisting of core theoretical elements such as human anatomy, pathophysiology, pharmacology and drug calculations. NMP also includes patient clinical and health assessments, and NMP learners monitor patients during the consultation using multiple physiological parameters. The parameters include health assessment, taking medications history, and performing disease severity assessments to support prescribing practice decisions (Dillon, 2007; Rhoads and Wiggins-Petersen, 2014).

1.1.3.1 Continuing Professional Development (CPD)

In nursing education, Master's level study is an accepted part of a process of continuing professional development (CPD). The European Higher Education Arena (EHEA) identified the knowledge and learning required using the competency model to guide content development. The competency model frames the learning requirement at each level of a degree qualification according to the award's educational and professional interests (EHEA, 2010; Lokhoff et al., 2010). The Tuning

model is a structure for best practice concerning the master's degree design education and nursing students' preparation to practice internationally. The work of Gobbi and Kaunonen on the Tuning Project provides a transparent and free to use stratagem for the content design and delivery of degree programs in nurse education (2018, p.9). In conjunction with the EHEA (2010) standards for higher education, the Tuning Project is influential in the nurse education curriculum. The master's courses in Tuning's core components are: First, the learning objectives for the subject domains learned. Second, the units of competence for the learning objectives and finally, the predetermined level criteria, i.e., competency for learners (Tuning, 2018, p12-17; European Qualifications Framework, (EQF, 2018).

The EHEA (2010) competency specifies performance and outcomes standards. A critical foundation of the Tuning system (Wagenaar, 2014). The competency model is supported because it promotes comparable degrees and degree portability between EU countries. See, for example, the Organisation for Economic Co-operation and Development (OECD, 2015); and the Programme for International Student Assessment (PISA, 2019). The term competency is counterintuitive to mastery, and the historiographic definition of mastery in the literature and the mastery index describes a strikingly different and higher level of cognitive skill and practice dynamics to learning than competency (Guskey, 2001, 2012; McGaghie, 2015). Mastery requires that students demonstrate deep learning and an adaptable understanding of a given subject (Guskey, 2010, 2013; Unwin and Fuller, 2010; Guskey and Anderman, 2012). There is an expectation of a range of knowledge for students at the master's academic learning level. A further expectation at level is several years of clinical experience and practice. It is applied to reinterpret the new learning experience transforming it to guide future action (Mezirow, 1991; Taylor, 2012). Therefore, students' independent NMP expectation is prior knowledge and complex material use at advanced levels to inform theory and practice (Oermann and Gaberson, 2014; Rolf, 2014).

1.1.4 Competence and the Competency Model: The Structure for NMP Teaching and Learning

In healthcare education, assessment and testing are ubiquitous. The assessment has an enormous impact on the practice. Both policymakers and the lay public recognise its importance to the professions (Nitko and Brookhart, 2011; Oermann and Gaberson, 2014). Competence concerning training and working practices is closely associated with the idea of a professional qualification (Cohen-Scalia, 2012). In nurse education in the UK, the competency model is aligned with the vocational nurse's concept. However, interrelated developments in nursing and healthcare education and practice may have disrupted the competency model's legitimacy as a framework for curriculum development in nurse education (Brown, 2011; McGaghie, 2012, 2015). The legitimacy of the competency model is challenged. Firstly, the NMC (2006, 2018) proficiency requirement for a degree qualification as the first level for nurse registration, thereby providing nursing students with the disciplinary knowledge they need to continue their studies at a higher level within their field (Risjord, 2009; Young and Muller, 2014). Secondly, the changing healthcare environment and the AP role's growth, complexity, and difficulty (DH, 2010; NHS Future Forum, 2012; HEE, 2017). The AP practitioner now needs a progression point beyond the traditional endpoint of competency (McGaghie, 2012, 2015).

The competency model underplays both the workplace's complexity, the knowledge-theory context in which practice actions occur, and the complexity of learning (Young and Muller, 2014; McGaghie, 2014, 2015; Dyson, 2018). The constructive alignment needs to be tied to higher domain knowledge levels to represent better the increasingly complex relationship between the nurse's role, nurse education and teaching and learning (Thompson and Smith, 2010). The ideal is to ensure the positioning of the NMP students to apply their knowledge to practice and include the critical thinking requirement for clinical problem-solving (Megginson and Whitaker, 2017).

1.1.4.1 Competency and Proficiency: Defining the Constructs

The purpose of competency is to define the task and task achievement deduced by performance (Wheelahan, 2010). Its premise in the NMP context is to harmonise the common aspects of responsibility within prescribing, designed to be used by any prescriber at any point in their career (RSP, 2016). The RSP competency framework supports the evolution of the scope of practice for multiple healthcare professionals. However, the use of the competency structure promulgates NMP performance and workforce capability (Cohen- Scali, 2012). Firstly, the educational philosophy of competence is designed to impart instructional information and structural knowledge for practice. Secondly, transactional and the skills developed are frequently treated as a single clinical procedure in a silo. These skills are often unrelated to any past or future events and focused on the practical technical skills required for job-related performance (Moore and Hoffman, 2011; Smith et al., 2014). On a micro level, 'competence gives rise to know-how action'. The affective component of performance and the meso state is role identity and behaviours (Cohen-Scali, 2012 p.31). On a macro level, competence intersects with education; however, knowledge and understanding lose importance to the skills and performance process Wheelahan, 2010; Halpern, 2014). As Harden et al., (2016) suggest, competency characteristics are at a benchmark level. The knowledge, skills, and behaviour differ from that required for the higher-level proficient job performance.

In contrast, the NMC (2018) standards of proficiency for registered nurses are driven toward an advanced level of knowledge and skills more robust than competently good and significantly better than (or above) the competent average (Oxford Dictionary, 2016; Roget, 2019). Proficient represents a level of expectation for student learning, a progressive change in the learner, which is higher for all students (Hoffman et al., 2014). Beyond competent defines the expected standard for education and practice (NMC, 2018). The pedagogic model for proficiency is recognised as a complex, multi-dimensional process involving interpreting and organising knowledge differently (Hoffman et al., 2014; Hardern, 2016). The NMC document uses both competency and proficiency interchangeably; however, there are conceptual and contextual differences.

Conceptually, proficiency speaks of complexity, accuracy, and flexibility, terminology that the NMC standards cite as requisites for today's healthcare practitioner in the contemporary healthcare environment (NMC, 2018). Complexity is characterised as cognition and metacognition, referring to the students' ability. To accurately use a wide and varied range of theory applied and synthesised within sophisticated and multi-dimensional knowledge structures for complex, clinical problem-solving (Schraw and Robinson, 2011; Dinsmore and Alexander, 2016; Dinsmore, 2017). Accuracy is the aptitude to deliver complex care and reduce errors resulting from cumulative learning (Dosselle et al., 2010; Campbell et al., 2014). Flexibility is the capability to identify and grasp patient-centred clinical problems, including assessing patients and intervening quickly, responding to them appropriately and mobilising help as necessary (Benner, 2009; Young and Muller, 2014).

Therefore, proficiency focuses less on performance behaviours and more on the cognitive processes and their influence on practice behaviours (Hoffman et al., 2014). Where the focus of achievement is above the beginner-competent levels of learning, indicating meaningful learning, intentional application of the material required by n the assignment's specification and the educational purpose of the assignment (Bloom, 1959; Alexander, 2003; Adamson and Bailie, 2012; Halpern, 2014). On a meso level, the achieved professional standards build upon prior capabilities and clinical experiences to demonstrate advanced practice achievement across the theory-practice domains (Rolf, 2014). On a macro level, proficiency leads to a conceptual 'transformative' shift allowing intricate, related knowledge to be better organised and synthesised, making the knowledge use open, discriminating and the frames of reference for application flexible (Bellman, 2010; Dix, 2016). Proficiency opens the nurse to a broader knowledge base and vision of practice, enabling understanding a range of possible alternative solutions to clinical problems (Dencev and Collister, 2010; Brent et al., 2011; Seidel et al., 2013). Proficiency does not change the scope of the NMP practice but optimises it. Therefore, when applied, the expectation is not more of the (competent) same (Rolf, 2014; Dinsmore, 2017; Dyson, 2018).

1.1.4.2 The Relationship Between Competence and Proficiency

The relationship between competency and proficiency is in their application to the assessment. Both are concerned with measurement. Both describe and measure aptitude and capability against performance level descriptors (PLD) in a given context (Astin et al., 2012). Competence and proficiency are theoretically grounded, and quality secured concerning the theoretical concepts and constructs assessed and measured (Hoffman et al., 2014). However, distinctions need to be drawn between the two concerning the student capabilities associated with a particular performance situation and measurement of the learning demonstrated and achieved.

It is established earlier in section 1.1.4 of this work that competence is what students know and can do (outcome) at the summative assessment point. The product 'competency' used in outcome assessment criteria and scores are applied as measurement and summarise student achievement of the academic goals. Test scores create a link between the curriculum, assessment and assessed work (Astin et al., 2012; Guskey, 2013a). The intent is to measure students' learning towards competency and the educational aim and benchmark within the competency framework (Jung and Guskey, 2010; Cohen-Scali, 2012; Mulcahy et al., 2015). As an outcome-based pedagogy for assessment, the competency model may not provide an accurate and comprehensive picture of what students accomplish relative to the explicit NMC proficiency demand for prescribing (Miller et al., 2010; Rolf, 2014).

Conversely, proficiency emphasises process and progress criteria favoured by educators. Proficiency, unlike competency, is not represented by product criteria where the complete picture of student learning is, therefore, incomplete (Brookhart et al., 2016). Process criteria are about how the students arrived at the level achieved. In NMP, the progress criteria are concerned with how far students have advanced in their learning. Scoring is related to assessing whether students reach or surpass the (proficient or mastery) achievement level (Jung and Guskey, 2010). Mastery (advanced from proficiency) is formulated and developed on the understanding that learning is sequential. Prior (competent) knowledge and experience are the foundation of subsequent proficiency or achievement (Guskey, 2010, 2012, 2012a; Jung and Guskey, 2010).

In the assessment (and measurement) of progress, the master's degree criteria include which theory-practice skills are contextually developed. What is also measured is the 'cognitive' process of developing proficiency, mastery or expertise in the theory-practice context (Guskey, 2010; McGaghie, 2015; Dinsmore and Alexander, 2016). The OECD has enshrined the notion of cognitive skills, and the Programme for International Assessment of Adult Competencies (PAICC) assess proficiency in several subject domains, including cognition (2016). However, the test scores are criticised as arbitrary as the individual students passing 'cut' score may not accurately represent the intended level descriptors or the achievement level for that standard concerning NMC expectations concerning progression (Guskey, 2013a; Brookhart et al., 2016). For discussion on cut-scores, see Classical Test Theory (CTT) in section 1.3.1.1. Given the complexity of patient's demographics and the healthcare environment, assurance is needed that the NMP learners are proficient, acquiring and transforming their knowledge and skills to make them flexible. As a course outcome, independent in the NMP practice situations.

1.1.4.3 The Limits of Competence

Since Project 2000's large-scale changes to nursing education and the NHS improvements around 2008, competence, i.e., the technicality and specialisation of skilled performance in nursing roles, is increasingly recognised. Therefore, competence rather than proficiency is reinforced as a standard of practice (RCN, 2010, 2012; NMC, 2018). In healthcare education, this appears as an internal post-Project 2000 education reform struggle between the competency model, its application to assessment and the mismatch to the clinical practice and sustainable knowledge requirement (Drummond and Standish, 2009; McGaghie and Kristopaitis, (2015). Competency-based education is counter to the knowledge concept in the traditional academic sense (Wheelahan, 2010; Halpern, 2014; Miller, 2018). Theoretical knowledge proposes interrelated knowledge; of facts, disciplinary knowledge, congruence with teaching and learning theory, and cognitive skills to continue acquiring and applying (new) knowledge (Halpern, 2014). What counts as competence may be counteractive to contemporary nursing and the nursing professions regulatory demand for proficiency (McGaghie, 2015). For the NMP assessment, this presents a dilemma because traditional forms of competency as a

level of achievement are potentially incompatible with the profession's legislation for proficiency (Lineberry et al., 2015).

The commodification of knowledge and the logic of performativity (Lyotard, 1984) result in routinised competence, distributed and marketed as a skill. Independent of the (knowledge) context (Drummond and Standish, 2007; Guzman and Trivelato, 2011). The terms in which education takes place in HE has shifted in balance to policy agendas and skills. The labour market demand has led to educational reforms of knowledge acquisition displaced by competent performativity. The result is the high value placed on doing rather than thinking about learning and professional practice (Hockings, 2010). Performativity can damage learning as academic education may no longer emphasise knowledge and assess students' cognitive understanding. The outcome is cognitively restrictive knowledge acquisition and impoverished learning experiences (Wheelahan, 2010; Zepke, 2013; Halpern, 2014; Naidoo and Williams, 2015).

1.2. The Research Problem, Statement, and Question

1.2.1 The Significance of the Research Problem

The assessment of students' performance and achievement in independent NMP education focuses on two main accountability goals. One is to provide evidence regarding the students' fitness for healthcare practice. The second is ensuring the course assignments reflect the course or program's objectives. Clinical practice and education in nursing are legislated by professional nursing bodies to protect the public, and achievement is parallel with patient safety (NMC, 2018, 2018a). The literature disseminates a compelling argument concerning adverse events, e.g., patient harm. Adverse events are reduced in hospitals and wards where practitioners are highly qualified, capable of critical thinking and problem solving (Kendall-Gallagher et al., 2011; Acquaviva et al., 2012; Aiken et al., 2010, 2014; Audet et al., 2018. Other characteristics are cognitive ability, hypothesised to influence patient surveillance effectiveness (Levitt-Jones, 2014; Cho et al., 2015; Lau et al., 2016). Therefore, the independent NMP (summative) assessment must show the students'

capability and establish their preparedness toward mastery. Moreover, proficiency to enter the community of practice of that NMP profession is a principal factor in NMP education and assessment (Benner, 2009; Duff, 2012; NMC, 2008, 2018a).

In the independent NMP programme, the emphasis is placed on the assessment's reliability, established from the assessment rubrics influenced and developed by experts (Hardern et al., 2016). The feasibility of the assessment tool remains a critical criterion on which to judge the assessment. A second and more important goal is related to the assessment's accountability, assessment quality, and evaluation (Fry et al., 2009; Quality Assurance Agency (QAA, 2010). Accountability works in the following ways: (a) the assessments are evidence of the theory-practice development of students (b) they do not distort the interpretation of the assessment purposes, i.e., licencing/certification, and (c) the assessed results are representative of the learning achievement (Hutchinson and Young, 2011; Crooks, 2011). These points have consequential validity (beneficial or harmful) and are vital concerning patient safety (Brewer, 2014; Aiken et al., 2014, 2017). Consequential validity is relevant and connected to the NMP assessment's validity concerning the proficiency claim (Messick,1980; Sackett et al., 2008; Kane, 2006, 2011).

The goal of validity is to evaluate (through measurement) the inferences and assumptions inherent in the proposed interpretation and use of the NMP assessment (Kane, 2006, p.118). In response to the professions' regulators, the validity ensures the assertion that the assessments meet the proficiency requirements and succeeds in preparing students to thrive independently as NMPs in practice (Messick, 1995; Kane, 1994a, 2013). Multiple methods of assessment are employed in the NMP programme to measure students' essential skills acquisition. Including simulation or virtual patient exams referred to as in this study as Structured Clinical Examinations (SCE), the written reflection (WRs) and others demonstrate students' achievement and thereby safeguard patients.

In many professional disciplines and academic programs, assessed scores indicate a student's preparedness for their chosen profession and their achievement of the academic standards (Dante et al., 2011; Wajizigha and Sireci, 2011; Kumar et al., 2017). The Quality Assurance Agency (QAA) sets the standards and core practices

that providers of higher education (HE) must meet (QAA, 2014, 2015). The qualifications framework for setting and maintaining standards complements and contextualise the distinctive features of a qualification at a specific level, e.g., mastery. The expectation is that NMP graduates will have " in-depth and advanced knowledge and understanding of their subject and profession" (QAA 2014, p3). For example, the QAA ensures standardisation and collaborating agencies. For example, the European Higher Education Area (EHEA) provides external, international degree modernisation and quality systems. However, the QA standards' maintenance requires measurement for quality assurance at the course assessment level to ensure that the assessment (validly) functions as designed (Haladyna and Downing, 2006).

A passing grade indicates PLD achievement on the SCE and WR assessments. The inference is that they are proficient and capable of independent NMP practice. However, based on an individual's assessed activities, the reasoning, i.e., theoretical and empirical evidence to support the scores' intended use, needs establishing (Cizek et al., 2010). Kane's (2006, 2013) validity theory is the conceptual framework applied in this study and used for justifying the assessed score inferences. The different types of validity and the use of multiple evidence in constructing a validity argument are discussed in sections 1.3.2 and 1.3.3, respectively.

1.2.2 The Problem Statement and Research Question(s)

The evidence of students' proficiency and overall course outcome, i.e., passing grade, supports certification and licensing to NMP practice (NMC, 2006; 2018; RCN, 2012). This study seeks to address the following overarching question:

How well do the Assessment Processes used in the Independent Non-Medical Prescribing (INMP) course Validate Certification/ Licensing of Students to Prescribe in Practice?'

The research questions are as follows:

Sub-question [1]. Are the content of the NMP Summative Written Reflections (WRs) a Valid Constructive Indicator of NMP Students' Proficiency ?

Sub-question [2]: To What Degree are the NMP Structured Clinical Examinations (SCEs) a Valid, Accurate Representation of the NMP Criteria they are Designed to Measure?

1.2.3 The Researcher's Orientation

The researcher's professional background is nursing in critical/ intensive care. Recording and interpreting bio-metric, pathophysiological measurements are paramount to safe patient care, delivery, and management. Knowledge acquisition (epistemology) and how knowledge is known to be 'true' knowledge is historically grounded in positivism. As positivistic scientific inquiry, true knowledge has a method that claims evidentiary support, usually empirical, reinforced by logic, objectivism, and measurement (Crotty, 2009). Positivism embraces the ontological view that the nature of the world is structured by logic and objectivity. Epistemologically, the ontological claims about 'true knowledge' are reduced through measurement and understood in its mathematical form (Crotty, 2009, p.27). Positivist theory's fundamental goal is closely aligned with the researcher's epistemology. That being to minimise the mystery of true knowledge through logic, objectivity, and accepted measurement (Goertz and Mahoney, 2012).

This study's nature involves objective reality, a realist inquiry to address the validity and measurement question and consider what works, why, under what circumstances, and how (Petticrew and Roberts, 2006). This study is significant to the next area of interest, patient safety regarding the NMP proficiency claims and the potential for prescribing errors in practice due to the necessary components of NMP theory and clinical activity not being accurately assessed (Messick, 1995). The important epistemological matter here is that the students demonstrate the legislated NMC (2018) proficiency achievement for theory, practice, and NMP knowledge development. In this study, epistemology addresses how well knowledge and other learning theories are validly captured and measured by the assessment and the students' assessed work (Kane, 1994a).

Positivism is the logical choice for measuring the assessments' capacity to accurately capture and measure student assessed achievement and performance. However, this study acknowledges that the association between assessment scores and students' actual knowledge is not straightforward. Therefore, post-positivist theory, such as critical realism, is included in my orientation, influencing how the data is analysed and presented. This study aims to apply validity theory and measurement theory to the assessed content and criteria itself. A move away from presumptions about student achievement based on test scores, establishing what is definitively measured and achieved in assessments and students assessed work (Shultz et al., 2005; Kane 2006; Markus and Borsboom, 2013).

1.3 The Theoretical Frameworks Underpinning NMP Assessment

1.3.1 Theory, Practice and Pedagogy

Assessments test the NMP students capability across a range of theories and contexts. Consequently, nurse education pedagogy is concerned with optimising conditions for learning and what nurses need to know. The curriculum design is, therefore, heavily informed by the teaching and learning theories. In NMP practice, the theories are structurally aligned to the curriculum, the assessment design, and that exhibited in the assessed work (Vitale and Exley, 2016). The function of pedagogy in this sense is to take the NMP learner into account. This study's pedagogic model considers what learners know using knowledge theory (Ryle, 1949). How learners practise (Benner, 2001), what learners reflectively think (Mezirow, 1990), including their cognitive engagement (Boom, 1956). Therefore, pedagogy and learning should be concerned with the interaction between theory and practice to transform practice (Waring and Evans, 2015; Dyson, 2018). In this study, teaching and learning theory identifies and measures whether the NMP assessment method is effective. Aligning these theories can help ensure that the students demonstrate the right skills and knowledge and that these are assessed using the appropriate techniques (Haladyna and Downing, 2006; Oermann and Gaberson, 2014).

1.3.1.1 Teaching and Learning Theories in NMP

The theoretical frameworks have relevance in determining what nurses do in practice. The theory represents cognition, interpreting how practitioners view their roles and how knowledge is applied to give meaning to practice and clinical actions. Theories provide a relational ontology, a holistic framework for analysing how learners understand the theory and practice relationship in the occupational NMP field (Anfara and Mertz, 2015).In this study, the theories of Ryle (1949), Mezirow (1981), Benner (2001) and Bloom (1956) are a framework for the analysis of proficiency and proficiency scaling. The research aims to understand the relationship between the examinees' actual assessment scores, students' ability, and unity with the cited theories. The lexicon of the four theories in Tables 1.4.2 to 1.4.4 represents the classification categories. When mapped against the assessed work, the theory specific classifications can analyse the WR content and examine the SCE test criterion. In this way, they help reduce subjectivity and measurement error (Loughran and McDonald, 2011). The power of analyses increases as the theories lessen the associated subjective bias (Anfara and Mertz, 2015).

The theories of Ryle (1949), Mezirow (1981), Benner (2001) and Bloom (1959), presented in more detail in section 1.4.0, resonate with the EHEA's strategy for academic quality (Curaj et al., 2018) and explains and explores the research question by defining the concepts discussed in the preceding paragraphs. In their narrative form, these theories provide a multi-method model for measurement, on which the research pivots.

1.3.1.2 Classical Test Theory (CTT) in Assessment Measurement

Classical Test Theory (CTT) is the most frequently used theory in assessment and testing. A student's test score in CTT estimates that students' achievement level is either a simple sum of the test items' scores or weighting. Weightings are used to reflect the differing importance assigned to the test's content (Oerman and Gaberson, 2014; Raykov and Marcoulides, 2016). CTT makes assumptions about obtained (or observed) scores *[X]*, is composed of the 'true' score *[T],* and error component *[E]*, symbolically written as follows:

$$\text{Obtained score } [X_0] = \text{True Score } [X_T] \pm \text{Error } [X_E]$$
$$\text{Or}$$
$$X_0 = X_T \pm X_E$$

X_E is assumed to be uncorrelated with X_T; there is no systematic relationship between a person's X_T and whether that person has positive or negative errors [X_E], making CTT imprecise (Shultz et al., 2005: p 58; DeVellis, 2006; Mango, 2009). The X_E in CTT is easy to control by standardising the performance criteria and testing conditions. Test standardisation can minimise extraneous factors, which influence performance and, consequently, the score. For example, examinee tiredness, motivation, and mood. X_E's primary use in CTT indicates how large X_E is likely to be and, therefore, how reliable the estimation of X_0 and X_T is (Shultz et al., 2005). An essential factor in CTT is that when measuring an NMP attribute of interest, measurement of characteristics that are not of interest also occurs because X_T is isolated and measured. However, as component X_E, which is a composite of other factors, is also measured. The result is that X_0 can be biased in a positive or negative direction, influencing the X_T. Other factors that can influence or contribute to the X_E variance are assessment instrument clarity, format, content, or test item criterion (Shultz et al., 2005; Haladyna and Downing, 2006). In the NMP assessments, if any or all these factors are vague, poorly constructed or understood. The X_0 and X_T scores can be positively or negatively affected, thereby influencing score integrity.

Score integrity may have implications concerning the extent to which the obtained score [X_0] and inferred score [X_T] reflect the actual NMP skills, subject domain. Also, the knowledge targeted by the assessment instrument. Patient safety implications are many if a student's X_0 scores are biased positively, resulting in a passing grade (Mango, 2009:1-2). A more reliable measurement of X_0 and a more precisely inferred X_T are a must for NMP assessment validity purposes. CTT is the most frequently used model for the analysis of test data. The limitations of CTT are its norm-referenced basis, wherein derived test scores' meaning is achieved by comparing their position in the norm group (Shultz et al., 2005). The attraction of norm-referencing is that it provides a straightforward, though very crude method of opening entry gates or a once-and-for-all grade used as an academic achievement sorting mechanism.

Testing under CTT uses norm-referencing to provide information about student cohorts distinguishing between the top and second and lower percentile scores. However, it does not tell us what the students can do. Norm-referencing's abstract form takes no account of the quality of the NMP student's performance or ability in any absolute sense because it is concerned only with students' rank ordering (Anastasi and Urbina, 1997). Besides, suppose the errors of measurement [X_E] are substantial. In that case, they can distort the interpretation of scores (that is, student proficiency) and undermine the effectiveness (validity) of the decisions based on unreliable pass marks (Shultz et al., 2005; Sackett et al., 2008; Kane, 2006, 2013). Test takers with limited ability can potentially pass the assessment, indicating proficiency or mastery, albeit at the lowest percentile of a pass grade point. Assessors could reward candidates inappropriately for their exam performance (Borsboom, 2005; Kubinger and Gottschall, 2007). CTT represents a reasonable approximation argument for X_T providing a basis for score-based interpretations, which does not serve well for this study in testing NMP assessment criterion, content, and construct. Nor does CTT perform well in measuring test item function, i.e., the test takers' true knowledge or test interpretation or use.

CTT is flawed because its theory begins with and hinges on proxy inferences of observed performance. Not on test (item) function and variables of interest such as test-takers proficiency (Borsboom, 2007). Advances on CTT under the universal title of Modern Test Theory (MTT) relate well to measurement, the relationship between X_T and the variables of interest, that is, KSP. Therefore, to establish validity in this study, MTT discussed further in section 1.3.4 is proposed for measuring the NMP assessments. The foundation of MTT is measurement. Measurement of the level of performance and ability of the examinees' captured by the SCEs. Also, of the attainment of the constructs, the WRs assessments are designed to measure.

1.3.2 Validity Theory (VT)

Assessment in nurse education and the NMP curriculum is central, and validation of these assessments are vital to their use. In this study, the purpose of validity theory (VT) is to provide evidence that the claims about what the assessments capture,

represented by the score, correlate to student proficiency. There is an increased reliance on assessment as indicators of student performance and care quality in the current healthcare education climate (McGaghie and Kristopaitis 2015; Vitale and Exley, 2016). However, in the light of recent unfavourable care quality reports, concerning the practitioner's safety, the discussions continue (Francis, 2013; The Kings Fund, 2017). The ready acceptance of assessed scores as valid evidence of this claim about the practitioner's ability regarding theory and the academic qualities has gained momentum (Redfern et al., 2010; Mucklow et al., 2011; Vitale and Exley, 2016). The validity focuses on facilitating a defence about the NMP assessment, the assumptions and inferences that link the assessed score with the standard of proficiency claim (NMC, 2018)

The view of content validity in educational measurement presented by several influential authorities such as Messick has been in place for decades (1989). Messick's specified technical definition of content validity provides the basis for score interpretation in testing. Content validity is related to the test creation and how well it reflects the measurement of the ideal characteristic, for example, knowledge, as envisaged. However, content validity presumes a detailed description of the content domain, i.e., knowledge proficiency, which may not be the case. Guion (1977, p.6) provided minimal conditions for acceptance of content validity as a rigorous measurement in testing, proposing that (1) the content includes concerns for relevance. (2) clarity of meaning and (3) construct or external criteria as other types of evidence. Linn et al., (1991) made further distinctions about content representativeness, essential in testing where the tasks are incumbent on students' proficiency, and the cognitive behaviours proficiency requires.

There are concerns that norm-referencing under CTT does not define the individual student's achievement in this context. In comparison, criterion validity testing has grown in the wake of the continued competency-based training and clinical simulation testing. Criterion validity aims to justify the test-item specifications, reducing the inferential leap from test performance to job performance against which the test criterion is aligned. In this study, criterion validity compares how a measurement instrument, i.e., the SCE assessment, captures the NMP practice and performance criteria. Particularly the NMP subject domains the assessments

espouse to measure. Moreover, in CTT, often lost in the assessed score measurement is the student's ability level, referred to as latent traits in Item Response Theory (IRT). In this study, IRT is a psychometric measurement applied to the NMP assessment to additional evidence for the validity argument. For discussion, see chapters 3 and 5.

1.3.3 The Conceptual Framework: Kane's (2006) Interpretive Validity Argument (IVA)

Kane's (2006) Interpretive Validity Argument (IVA) is an approach to validity rather than a validity type. It provides a practical conceptual framework for generating clear evidence about the students assessed performance to support the test score inferences and use. The IVA justifies the score's interpretation and use by providing evidence obtained from analysing the relationship between the content. The subject, domain-knowledge, and construct reflectivity demonstrated by the student are the assessment's intended measure (Miller et al., 2009). The validity argument is built through a critical analysis of the plausibility that assessed content and test criteria support the score interpretation and use. Kane's IVA draws on 'The Uses of Argument' (Toulmin (1958). A theory in which logic and argument satisfy the burden of proof that the assessed work content or criteria are a fit for the score interpretation and use. Toulmin's theory defines the argument's structure about claims whose legitimacy proposes establishing the argument (2008. p.91). The IVA structure is as follows:- The claim (C) represents the destination of the argument (i.e., proficiency), the data (D) is the information (test score) on which the claim is built. The warrants are the general assumptions, statements or inferences made for which evidence (E) is provided (p.94). Table 1.3.3 illustrates the structure of the validity argument and its application to the NMP assessments.

Table 1.3.3 An illustration of Kane's (2006) Interpretive Validity Argument (IVA)

C= the Claim	D= the Data	E= the Evidence
Students are proficient	NMP Assessments: the assessed work Pass score	Warrant= Students demonstrate/achieve theory Evidence= Assessed data analysis using theories to support the claim

Table 1.3.3 summarises the validity argument as the conceptual framework for the study: How the claim, data and evidence situated in the IVA argument are for addressing the research questions and aim.

The evidence legitimises the warrant, and the theories of Ryle (1949), Mezirow (1990), Benner (2001) and Bloom (1956) provide the evidence. Their concepts specify the content and criterion for analysis, which determine the plausibility of the proficiency claim. Under Kane's IVA, the framework allows an interpretation of where the students demonstrated learning achievement lay (e.g., proficient) when analysed against the cited theories. The theories used in the analyses will either support or refute the assessment assumptions and proficiency claims.

The NMP makes the following general assumptions (warrants).

(1) That the knowledge and cognition required to achieve NMP proficiency are demonstrated successfully in the WRs

(2) The student has demonstrated the application of the underpinning teaching and learning theory in the WRs in keeping with the academic expectations and is consistent with the assessed scores

(3) The SCE test items are at the appropriate level, are related to the theory and are construct representative of the NMP criteria for Proficiency, as required by the NMC

(4) The SCE test items' quality provides stable estimates of test-takers cognitive ability and performance

In the context of summative NMP assessment, the above inferences and assumptions are accepted readily as being reasonable. However, the IVA attends whether the above points (1) to (4) are plausible or even probable. The IVA evidence can either support or invalidate the assessment claims by showing that at least one

or more of the above inferences in the interpretive argument works. Fig 1.3.3 gives an illustration of how Kane's IVA is applied. Benner's theory provides the backing (evidence) for the warrants, i.e., the assumptions and claims made about student's achievement based on the test score. Moreover, evaluating the SCE assumptions listed as points (3) to (4) requires a different evidence source. Therefore, the evidence is substantiated through measurement in the application of Modern Test Theory (MTT), specifically IRT, as discussed in 1.3.4 and the results presented in chapters four and five, respectively.

Fig 1.3.3 A Visual Representation of the Application of Kane's IVA to Benner's Model
one of the multiple sources of supporting evidence to the NMP validity claims

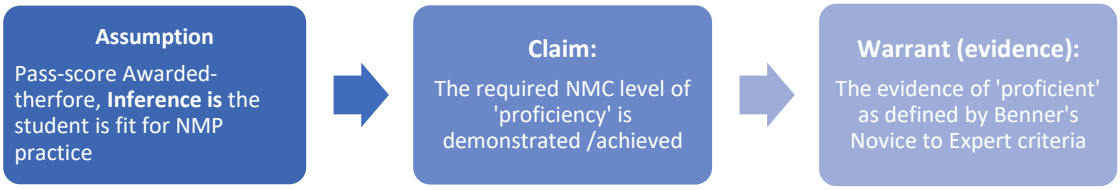


Fig 1.3.3 An Illustration of the application of Kane's (2006) IVA,. The diagram outlines IVA process concerning the assumption and inference about a student's ability, based on the passing test score. Followed the claim (of proficiency). Backed by the warrant (evidence) of achievement using Benner's (2001) model criteria for measurement of WR content, to support or refute the proficiency claim.

Critics of Kane's IVA postulate that it is overly complicated and devoid of a single test or procedure to determine validity evidence, including when and how to obtain validity evidence (Sireci, 2007; Chapelle et al., 2008). Also, Kane's IVA lacks real examples, is elusive, and its use is too cumbersome for practical use (Borsboom and Mellenbergh, 2004; Lissitz and Samuelsen, 2007; Zumbo, 2009). However, an important note is that Kane's IVA was developed to support validity investigations and remains consistent with the traditional validity theory (Borsboom and Mellenbergh, 2004, p136). Kane's IVA's task is to facilitate collating the evidence required to judge the validity of assessed work and testing in general.

Fig 1.3.3.1: Overview of the Validity Theories used in this Study

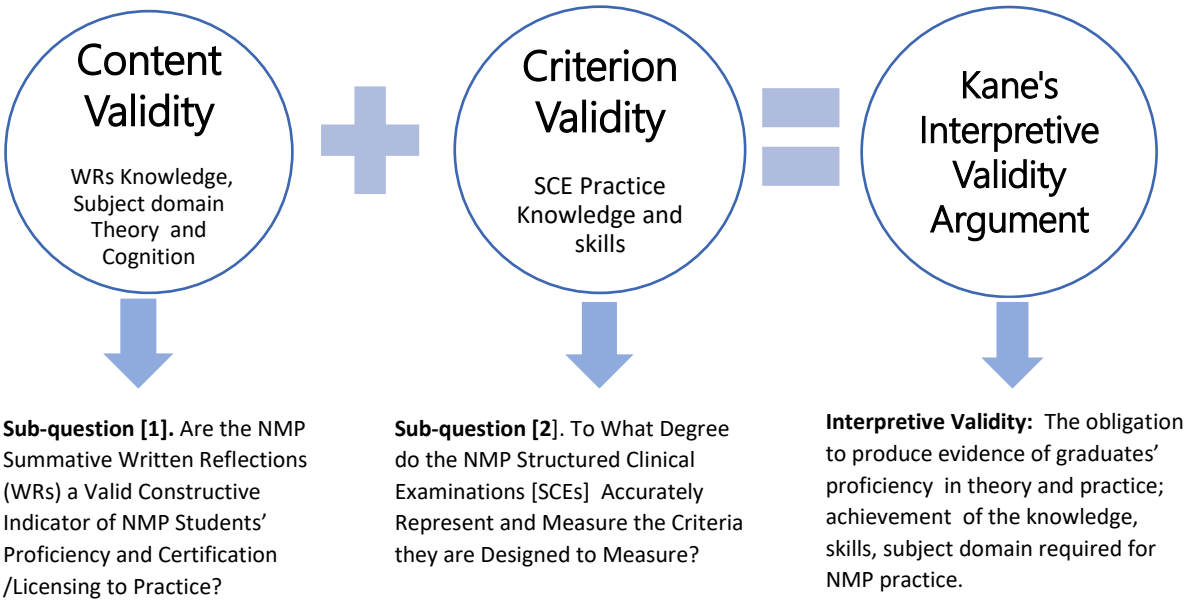


Fig 1.3.3.1: Conceptual illustration of Classic VT as applied to the research question: How well do the assessment processes used in the Non-Medical Prescribing (NMP) course validate certification/ licensing of students to prescribe in practice.

As a framework for validity judgement, the IVA encourages more robust endeavours to support the assumptions and their inferential links to student's ability through explicitly stated, identified and measured sources (Markus and Borsboom, 2013). The application of Kane's IVA is discussed further in chapter 6. In this research, Fig. 1.3.3.1 contextualises the role of classic validity theory (VT), still helpful in Kane's IVA in providing evidence of validity regarding students' scored performance. The score interpretation and IVA evidence are centred on theory descriptors of knowledge, competence and other abilities such as reflectivity, cognition and others, as presented in section 1.3.

1.3.4 Modern Test Theory (MTT) for Substantiating Kane's (2006) IVA

Validity in educational tests depends upon measurement. For decades, the standard practice has been to use statistics to legitimise the measurement method used.

However, for this study, the measurement must include the extent of the inferences from the student's response behaviour related to the test's theory or technical-based property. This study employs Modern Test Theory (MTT) as the paradigm for analyses and test measurement, as documented in chapter 2. A detailed discussion about MTT is beyond this study's scope. For this, see Markus and Borsboom (2013). However, that MTT takes care of the CTT limitations. A key factor is measurement, offering a suite of strategies and models to match the substantive theory or technical-based assessment structure more favourably.

The MTT philosophy is that measurement for validity requires methods and methodologies that recognise the relationship between the student's ability, the student's response, the domain/theory and the test scores. MTT is an attractive feature for this study. It represents a flexible framework for maximising test validity measurement, interpreting students' performance (i.e. the scores), and equating their response with the assessment tasks. Fig 1.3.4 is a design adapted from Bachman and Palmer (2010). The reworded text content Illustrates the conceptual relationship between MT and VT, highlighting how MTT is relevant to validity in this study. MTT can account for the measurement of the content and criterion's representation of the theory; in response to the evidence about the warrants and assumptions required in Kane's IVA.

1.3.4. A Dialectic Illustration of the Inferential and Theoretical Link between MTT and VT for Test validity, Test performance, Test Interpretation and Use.

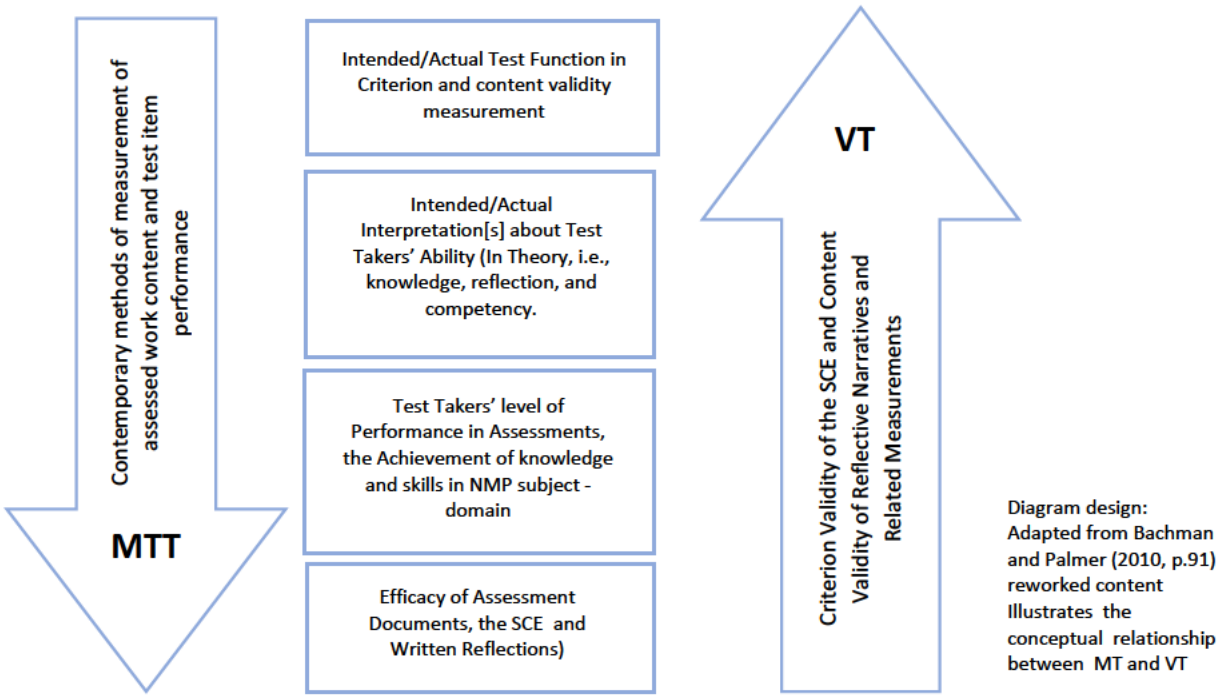


Fig 1.3.4. Illustrates the conceptual relationship between MTT and assessment measurement and its inferential link to validity theory. The downward arrow represents the statistical models used for SCE test measurement, while the upward arrow represents the types of measurements required to establish validity in the assessed written reflection and the SCE criteria.

MTT, when combined with VT and operated concurrently, there is accumulated evidence. The evidence confirms that test scores' inference and interpretation are extended beyond the quantitative CTT to provide an apparent and qualitatively meaningful test validation indication (Cizek, 2012). MTT also includes representational theory measurement (RTM), defined as an empirical theory for measurement, using numerical representations of relationships between objects (Markus and Borsboom, 2013 p.27-43). Characteristically in RTM, the numbers assigned provide a map to show how the corresponding object's property is related. Heilmann's (2015, p.5) newer interpretation of RTM does not view RTM as an empirical theory but as a more general qualitative relational structure for providing insight into the discussion of abstract concepts in qualitative data. Heilmann denotes abstract concepts as objects (O), which require clear definition, qualitative relations (R), and meaningful comparisons. Table 1.3.4 presents an illustration of RTM is as

applied in this study. RTM is complimentary to content and criterion analysis by facilitating the relationship between 'O' and 'R'.

Table 1.3.4 Qualitative Representational Theory Measurement (RTM), an Exemplar using Mezirow's (1981) Concept of Reflection

O- Object	Mezirow's Seven Concepts of Reflectivity **+**
R - The Abstract	The students' WR content mapped against Mezirows' Seven concepts of reflection **=**
0=R	Measurement of the relationship between 'O' and 'R' = IVA Evidence

Table 1.3.4 Uses Mezirow's (1981) framework of reflection illustrates RTM's application for content analysis in this study. The object of measurement (O) is Mezirow's (1981) Concept. The abstract (R) is the level of reflectivity as represented in the content of the assessed WRs. The results of 0+R provide Kane's (2006) IVA evidence to support or refute the proficiency claim.

Heilmann's RTM, therefore, provides an additional method alongside content (analysis) mapping by facilitating measurement of the correlation between the abstract (O) and the student's response (R) for generating IVA evidence. Further, Heilmann's RTM enhances the standard content validity measurements used in WRs. Providing additional IVA evidence on how far the comparisons (measure) between the theory and students' responses are exact representations of the underlying teaching and learning theories presented in 1.4.0. The traditional definition and application of RTM cited by Markus and Borsboom's (2013) are not excluded, appearing in the quantitative analysis. The psychometric statistics applied to the SCE for quantitative analysis supply information concerning the test items capacity to measure the test's power. To discriminate between the examinee's ability (trait) and the test item difficulty.

In the psychometric statistics, there is the quantitative measurement to satisfy the traditional RMT, i.e. the relationship between the student ability and the SCE test item difficulty. The psychometric analysis is part of the suite of test measurement devices available in MTT. Moreover, psychometrics is a valuable measurement for

test development or test redesign (Anastasi and Urbina, 1997; Bovaird and Embretson, 2008; Markus and Borsboom, 2013). Fig 1.3.4.1 represents the collaborative application of criterion and content VT as applied to MTT methods for measurement, which is key to providing the evidence for Kane's IVA in the NMP assessment. Fig 1.3.4.1 includes the decisions concerning measurement in this study, the evidence for the IVA, and their use in addressing the research question.

Fig 1.3.4.1 Summarises the Validity Measurement Decisions for Addressing the Research Questions

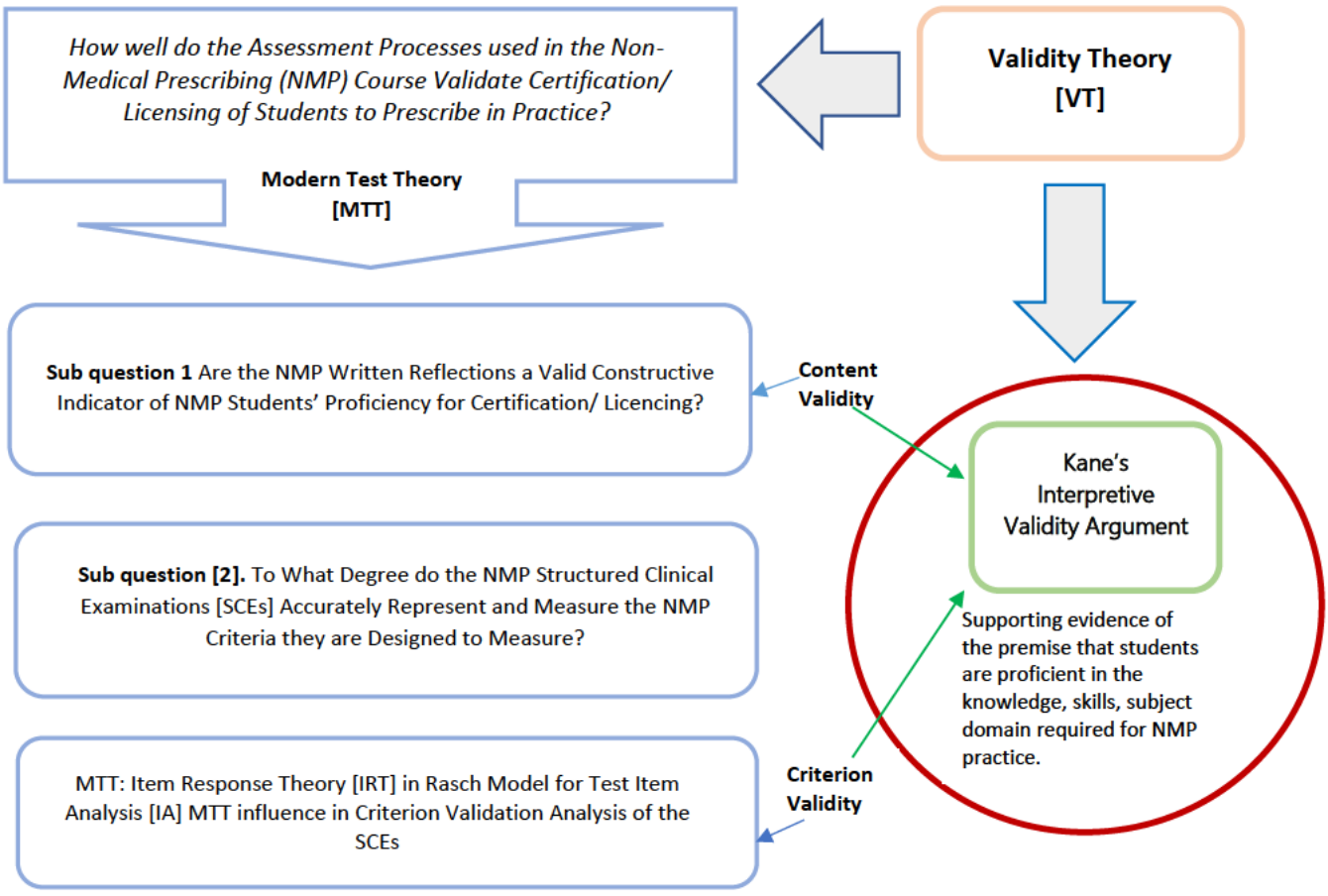


Fig 1.3.4.1 demonstrates the influential relationship between VT and MTT. CTT's limitations are its sample-based descriptive statistics (see preceding paragraphs). Both VT and MTT provide theoretical and justifiable measurement principles to establish NMP assessment validity.

Measurement in education and assessment remains a troublesome problem. Of additional interest in using MTT in this study is measurement as it relates to fallacy.

The fallacy occurs when there is a projection of understanding about the test responses (Eemeren et al., 2009). Markus and Borsboom (2013) suggest that any field of education testing of students' knowledge, cognition and other variables of interest is subject to fallacy because assessors can arrive at different conclusions about the test-takers responses. Even where information is the same, the assessor's perspectives, understanding and interpretation of the underlying theory or technical-based knowledge reflect accordingly in the scores awarded. A defined method reduces the potential for fallacy. The theories cited earlier and presented in section 1.4 serve to illuminate the assessed score's interpretation. The student responses are measured according to the cited theories and assessed performance (Putnam, 1975; Eemeren et al., 2009). MTT offers a sizeable set of measurement options that guide efficient and effective test validation measurements. Kane's IVA is consistent with content and criterion validity principles. MTT provides a more precise answer to questions about a tests' content, function, and capacity to measure test takers actual achievement. The proposed MTT application is presented in Chapter 3.

1.4 Knowledge, Theory and Ways Knowing: The Concept of Mastery in Higher Education

Continuing professional development (CPD) is a strong motivator for healthcare practitioners embarking on M-Level studies and caring about career opportunities (Risjord, 2009; Young and Muller, 2014). NMP as a CPD pathway for learning is about meeting formal academic requirements, practising professional self-conduct and improving clinical practice (Lee, 2011; Sargeant et al., 2018). Moreover, adopting a CPD mindset, self-conduct responsibilities and knowledge acquisition are definitions of professionalism (NMC, 2018a). As a result of the Bologna Process, UK universities and faculties in healthcare education have embraced Bachelor's and Master's programs and qualifications (Crosier and Parveve, 2013). Bologna's acceptance indicates the increasing demand for profession-specific education and a workforce development orientation (OECD, 2014; EHEA, 2015; NHS, 2019).

The literature gives considerable attention to professional education and the concept of mastery. It indicates much higher performance levels and achievement of essential knowledge and skills (McGaghie, 2014). Educational policy documents issued by

agencies such as the OECD (2019a) and the European Qualifications Framework (EQF, 2018) perpetuated mastery. In these documents, the cognitive features of learning, i.e., managing, updating, building new knowledge, logical, creative, and critical thinking, are promulgated as essential. However, the idea of mastery in learning and education is not new. The mastery concept was defined and illustrated in Bloom's seminal work describing the degree of cognitive participation expected from the learner (1956). Other authors continue in the quest to conceptualise and clarify the particular attributes, capabilities and attitudes required for 'mastery' and of those students enrolled in post-graduation Master's (M-Level) education (CoRe2, 2011; Gobbi et al., 2018).

M-level learning as a CPD activity aims to provide practitioners with learning skills, applying theory to support their occupation. Including the ability to work through complex clinical problems by integrating and adapting various knowledge forms (Litzinger et al., 2011). The idea of mastery in NMP education helps learners consistently reach a more significant achievement level than competence, i.e., proficiency or mastery (McGaghie, 2015; NMC, 2018, 2018a). The traditional features of mastery achievement are cognitive skills. The capability to organise intricate knowledge and deliver the knowledge informed action needed to effectively manage complex clinical situations (Schraw and Robinson, 2011; Young and Muller, 2014). Cognitive skills development endorses knowledge use; to improve the skills and strategies used in problem-solving, decision-making, clinical reasoning, and judgment (Alfaro- LeFevre, 2013). A central part of NMP education involves cognisance; this is also true for prescribing performance and the concept of knowledge transfer (Nuttall and Rutt-Howard, 2011; Miller, 2018). In this study, cognisance is considered in the validity of assessed work to account for Mastery.

Nursing is a healthcare profession where learning is valued for driving quality care (Unwin and Fuller, 2010; McGaghie, 2015). A body of evidence shows that the clinical staff's theory-practice skills are a source of substandard care (Aiken et al., 2014, 2017). Health educators and professional legislating bodies need to expect more from healthcare students, and mastery is a reliable baseline for improving knowledge and clinical attributes (Astin and Lising, 2012; McGaghie, 2015). Enforcing the higher standards of healthcare education in the professions should be a consistent quality

improvement to ensure patient safety (McGaghie, 2015; Alfaro-LeFevre, 2016). This study's essential source of validity is teaching and learning theory to help define the WR content and SCE accomplishment of standards to ensure the NMP assessment's validity.

1.4.1 Ryle and Knowledge Theory

Healthcare and NMP practitioners face a world of clinical complexity, which finds its way into the pedagogies of teaching, learning and assessment used in degree programmes. Little attention is paid to the concept of knowledge and the confusion regarding the structure of 'knowing' as a critical aspect of cognitive (proficiency) achievement (Ryle, 1949; Eraut, 2004; Dewey, 2004; Schraw and Robinson, 2011). Critically debating and unpacking the existing theory and breadth of deliberations about knowledge is beyond this work's scope. Young and Muller (2014) note that there is exclusive stress on skills at the expense of knowledge. Other authors suggest skills have become a disposition to a set of (competent) behaviours held in higher regard than the traditional academic knowledge concept (Vitale and Exley, 2016). Ryle's knowledge theory is concerned with knowledge use. In the context of the NMP assessments, how well learners use the newly acquired NMP information to (re)construct their knowledge structures (Wheelahan, 2010; Halpern, 2014). Mastery learning means that learners acquire essential knowledge as they move through the educational NMP material (Wheelahan, 2010). Table 1.4.1 contains a summary of authors whose work on knowledge theory delineates the knowledge types pertinent to this study.

The second pillar of critical thinking is prior knowledge. Thinking that is purposeful, reasoned and goal-directed (McKenna and Slavin, 2008; Schraw and Robinson, 2011; Halpern, 2014). The examination of NMP knowledge is significant to the validity argument because students need to access and apply knowledge at increasingly sophisticated prescribing practices (Nuttall and Rutt-Howard, 2011). It is imperative to know if NMP students can select and apply knowledge to arrive at a more elegant or efficient solution to an NMP problem (Oermann and Gaberson, 2014).

Table 1.4.1 A Summary of Knowledge Theory: A Framework for Measuring NMP Student Achievement and Engaged (Mastery) Learning Approach.

Author	Knowledge Type	Knowledge Acquisition, Interpretation, and Outcome
Ryle (1949), **Miller (1990).**	I. Declarative *Know-that* (or conceptual knowledge) information. I. Procedural - a compilation of declarative knowledge into units incorporating domain-specific strategies. I. Conditional - understanding where and when to use/access facts and employ procedures for changing situations.	I. The cognitive process concerning memory, factors that influence performance. I. Helps the individual in making the transition from one problem to another; contains sequence actions/performance that are domain-specific. I. The individual uses explicit and defensible (know-that) knowledge to provide actioned sequenced solutions (know-how) to solve problems.
Greeno (1978). **Rittle-Johnson and Alibali (1999).**	Know-that (conceptual) knowledge; static knowledge about facts and principles that apply in a specific domain. The interrelations between domain knowledge.	Additional information that problem-solvers add and use (know-how) to provide know-why solutions.
Gagne (1985); Merrill (1983, 1987).	Task knowledge - the psychological requirements of learning tasks; retrieval of relevant rules, and information application.	Learning a variety of procedural tasks (know-how) in external conditions of learning; demanding, realistic settings. Domain-specific (know-that) internal learning conditions for problem-solving.
McKenna and Slevin (2008).	Prior knowledge	May include a conceptual framework for social, occupational interaction. Includes tacit and explicit (know-that) knowledge.
deJong and Fergusson-Hessler (1996).	Strategic knowledge: Knowing why, when, where, and how knowledge applies.	Organised problem-solving, organising, interpreting, and analysing information. A sequence of solution activities plans of action laid down. Applicable to a wide variety of problems within a specific domain.
Posner and McLeaod (1982). **Mooney and Nowacki (2011).**	Know-why (strategic) - when declarative and procedural knowledge is successfully applied to understand causal relationships and problem-solving.	Critical thinking (conditional, know-why) knowledge, which is about demonstrating understanding through appropriate use. Application and transferring of declarative (know-that) and procedural (know-how) knowledge from one situation to another over a range of clinical endeavours. Can include hypothesis testing and generating new knowledge.
Dickoff and James (1968).	Know-why presupposes the existence of knowledge at lower levels I-III. *Know-that* I. Factor isolating or naming theory *Know-how* II. Factor-relating theory III. Situational relating theory a. Predictive theory b. Promoting or inhibiting theories *Know-why* IV. Situation producing theory	I. To have such a theory is to have the terminology without the need to labour the importance of proper and known terminology. II and III Not seeing the theory in isolation, but about situations dependent upon factors [I] already identified. a) the prior existence of a more elementary type of theory b) cause and effect relationship; a prerequisite for situation producing theory IV Significant for practical activity in the light of theory, assessment, construction and refinement of theory. Research that explores the validity of theory or research geared to stimulate thinking toward theory.

Table 1.4.1 Provides a knowledge theory Framework for interpreting how knowledge is represented and used by students. An abridge compilation of published works, the framework is used to identify the level at which knowledge is engaged with and displayed in the NMP students assessed work. Also, to operationally define knowledge concepts for measurement and for Kane's (2006) IVA.

(Kramer, 2016). In contrast, strategic knowledge consists of a range of intricate knowledge use, characterised by clear, precise, purposeful thinking: Combined with interlinked concepts and contexts defining strategic knowledge as an interconnected knowledge structure complete with the rules and strategies for application (Mooney and Nowacki, 2011). Strategic knowledge is essential to NMP. It includes if-then category relationships to cause-effect, evidence of mature cognitive reasoning (Schraw and Robinson, 2011). Used to interpret the patient's observations, the symptoms, developed through a network of arguments and supporting theories to why the symptoms react/respond (or not) to the NMP interventions.

Knowledge theory is valuable in analysing the assessed work. Determining the student's ability to acquire knowledge, taught or developed by studying subject material and contextually applying knowledge 'strategically' to inform NMP theory and practice. In this study, the measurement of strategic knowledge highlights whether the assessment allows learners to understand how theories can support each other. Or how knowledge in use provides explanatory coherence, mainly when strongly supported by knowledge and theory in different subject domains (Halpern, 2014). For Kane's (2006) interpretive validity argument, knowledge theory provides an immediately applicable framework for interpreting NMP students' knowledge and achievement level through Ryle and other knowledge categories. In this context, the role of knowledge (plus other identified) theory is essential as it articulates whether the assessment design can facilitate student response to how knowledge influences professional practice (Wheelahan, 2010; Halpern, 2014).

Knowledge theory provides crucial evidence (warrants) for the IVA to support the NMC (2006, 2018) proficiency claims that the NMP assessments make about students' achievement. As a content validity measurement, the significant point of knowledge is that the NMP assessment emphasises contextual, conceptual understanding and knowledge application. The IVA explores the students understanding of the interrelationship between the learning theories or constructs that enable NMP students to function at the desired strategic (knowledge) level (Halpern, 2014).

In summary, Ryle's (1949) et al., knowledge theory provides the framework to determine whether NMPs transfer knowledge from theory to practice, a long and deliberated subject in nurse education (Rolf, 1996; Scully, 2011). In the healthcare literature, a distinguishing feature of advanced practice is whether there are discernible signs of progress and expansion of existing knowledge (Mantzoukas and Watkinson, 2007; Hickerson et al., 2016). Strategic knowledge represents knowledge expansion, developed from (mastery of) theory and broadened to include critical thinking (Halpern, 2014; Hoffmann et al., 2014). The knowledge theory delineated in Table 1.1.4 in this study aims to determine whether the NMP assessed work is consistent with the demand for higher-level knowledge theory), practice and clinical performance (Imison et al., 2016; HEE, 2018).

1.4.2 Theories of Reflection: An Overview

Over the past three decades, reflection has become a central pedagogical approach to life-long learning and CPD in nurse education. Models of reflection are utilised in nursing and healthcare education to foster deep learning (Young, 2018), associated with critical thinking. Deep reflection is an essential academic quality exhibited in students assessed work content (Bradbury et al., 2010). The literature cites the following authors who offer theories and models of reflection, which involve practitioners looking back on something that occurred to understand what they know and do. As a developmental process, reflection is about students' knowledge of practice by reconsidering what they learned in practice (van Woerkom, 2010).

Schön's (1983) influential work 'The Reflective Practitioner' conceptualised reflection-in-action and reflection-on-action centralised what practitioners do.
Critiques of Schön's model argue it is broad, open to misinterpretation, and tricky to make sense of or achieve (Bradbury et al., 2010). Kolb (1984) and Boud et al., (2002) present the Experiential Learning Theory (ELT), which purports reflection as synonymous with higher-order cognitive processes. However, ELT's criticisms lack clarity on how reflection is achieved (Farrell, 2012). Alternatively, Gibbs' (1998) and John's (1995, 2004) models conceptualise the methods of reflection in the module's stepwise application to structure and facilitate students' application to the learning experience (Bradbury et al., 2010).

John's (2006) Model provides a step-by-step approach to reflection; however, this model soon becomes complicated with a loss of clarity for what reflection is and how to achieve it (Fook et al., 2006; Bradbury et al., 2010). These models, popular in the nursing literature and learning culture as a technical process for reflection. Mezirow's (1981,1991) philosophy of reflection, based on critical theory, offers a theoretical framework to facilitate understanding and interpreting students' engagement with reflection. Central to learning in the profession is improving patient care. Reflection as a learning technique is crucial for developing critical questioning of practices (Moon, 2006; Bradbury et al., 2010). Critical inquiry stimulates learners to relate the new NMP material of learning to previous knowledge and, thereby, learn meaningfully (Hughes and Quinn, 3013). Mezirow (1990;1990a) stressed critical questioning as necessary to transformative learning, an upgrading of learning, a feature of adult learning.

1.4.3. Mezirow's (1981) Seven Concepts of Reflection

Mezirow's concepts are preferred to the others cited. It provides a tangible and unambiguous guide on determining the students' access and facilitation of reflection. Mezirow's work represents a comprehensive means of assessing the level of reflection in the NMP assessed work. Mezirow's (1990) philosophy is a critical theory that argues that students can develop an amiable scepticism, engage in critical questioning concerning whether NMP practice is acceptable. A sense of knowing when and what to question (Rolf, 2014). As listed in Table 1.4.3, Mezirow's seven concepts, through their categorised and sequential stages, provide an illustrative framework for understanding and interpreting reflective (critical) thinking, characteristic of critical reflective discourse (Taylor and Cranton, 2012). Table 1.4.1 tells where learners stand on measures of reflection and their progress in critical reflection.

Table 1.4.3 Mezirow's (1981) Seven Components of Reflection. Adapted from 'A Critical Theory of Adult Learning (1981, p 11-13).

Levels of Reflectivity	Reflectivity Characteristics
1) **Reflectivity**	ordinary consciousness, described as recognising our habits, how we think or act, or specific behaviour.
2) **Affective Reflectivity**	It is defined as being aware of how we feel about the way we are thinking or acting and about our habits for doing so. An ability to assess the efficacy of thought, actions, habits, and ways of doing things. The ability to identify causes for thinking and actions, and what our relationship is in the situation.
3) **Discriminant Reflectivity**	Assessing the efficacy of our perceptions, thoughts, actions, and habits; of ways doing things and functioning; identifying relationships in the situation.
4) **Conceptual Reflectivity**	Self-reflection that may or may not lead to critical questioning and intellectual argument/debate.
5) **Judgemental Reflectivity**	Making value judgements about our thoughts, actions, and habits in terms of them being positive or negative, liked or disliked, beautiful or ugly. Here, Mezirow also ascertains other forms of reflectivity, such as 'conceptual' where the ability to critique and question the constructs (political, cultural, social, and economic) is employed.
6) **Psychic Reflection**	The ability to recognise in oneself the habit of making judgments based on limited information.
7) **Theoretical Reflection**	An awareness of the reason for a habit—if making judgements or assumptions—or the reason for conceptual inadequacies, meaning 'taken for granted' assumptions (central to transformative learning).

Table 1.4.3 Illustrates Mezirow's (1981) framework which provides meaning to what constitutes the lowest levels of reflectivity (non-reflective). The framework identifies the levels of reflective thinking regarded as knowledge specialised and contextual to purpose.

Critical reflection and reflective discourse are the two processes in transformative learning. These processes assume a certain level of critical thinking (cognitive) development (Mezirow, 1990). Mezirow's framework contains essential components that show the process's recursive nature and the criteria for determining students' reflective thinking. For example, (a) do NMP students reflect on practice and (b) at what level of reflection according to Mezirow's concepts. For assessed work analysis, the reflection quality based on his ideas facilitates measuring whether students have attained the associated cognitive qualities/characteristics with learning by reflection (Denon, 2011).

The NMP assessment assumes the student's capability to engage in critical reflection; therefore, the first pedagogic concern is the NMP student's engagement with the critical reflection aspect. The reflection themes such as Critical Theory (CT) subject-domain knowledge and scholarship has an essential bearing on learning and references the aspirations associated with mastery (Khem, 2010; CoRe2, 2011). The

students' cognitive activity becomes visible, only to the extent that reflective theory is engaged with and discussed, through the confrontation of clinical problems, threats, or challenges and how this impacts activities or practice. The second pedagogic concern is the reflection level. For example, the individual learner's requirement to exhibit the transformative learning progression expected in reflective practice.

The healthcare and education literature cite reflection as beneficial to skills development and individual learning needs (Norrie et al., 2012; Ryan and Ryan, 2013). In this context, a student's failure to achieve high levels of reflection can adversely affect clinical decision-making and potentially patient safety (Padden, 2013). This study aims to provide substantive data as evidence of critical reflection in NMP assessed work. In addressing the research sub-question 2, exploring the methodological means for capturing its presence is necessary. In accumulating evidence for Kane's IVA, it is essential to verify whether the NMP students reflect at the desired levels and demonstrate the critical thinking required for mastery in their assessed WRs.

1.4.4 Benner's (2001) Novice to Expert Model

A contrasting pedagogic concern is the students' level of practice (Benner, 2001). Whether students can proficiently and purposefully direct action to meet prescribing demands (Waring and Evans, 2015; Dyson, 2018). Benner's work serves as a formula for understanding how to interpret professional practice, showing knowledge and skill acquisition development from novice to expert and mastery accomplishment levels. As shown in Table 1.4.4, Benner's model expands practice theory, enabling students to demonstrate and articulate the prior nursing knowledge and skills acquired through experience.

Table 1.4.4 Benner's (2001), Novice to Expert Characteristics: Abridged from 'Novice to Expert: Excellence and Power in Clinical Practice (p.20-38)'.

Levels of Proficiency	Proficiency Descriptors
1. Novice Applies template of rules, minimal knowledge, limited or no information processing.	Learns to recognise various objective facts and features relevant to the clinical skill. Dependent on rules to determine clinical action based on those facts, sometimes reorganised and applied context-free, with minimal reference to the overall situation in which they occur.
2. Advanced Beginner Good working knowledge of rules; limited information processing and understanding of subject and area of practice.	Sometimes recognises and uses various objective facts and features relevant to the clinical practice. Primarily context-free, with some deliberate, conscious application of rules; accountable only to those rules and not the results.
3. Competent Applies rules with some limited breadth and depth of information processing; moderate understanding of discipline and area of practice.	Context-driven application of rules; has a sense that what is important is missing. Sees complex situations as a set of facts, draws a conclusion, and solves problems based on set rules. Accountable to those rules, choice of action based on those rules, and recognition of the importance of the results without conscious thought.
4. Proficient Information processing and knowledge application based on rules. Know-how is situational, recall of similar situations.	Applies rules for decision-making; choice and deliberation are contextual. Detached decision-making derived from rule-like calculations; anticipates and solves problems based on previous similar events.
5. Expert Authoritative knowledge of rules and discipline, deep tacit understanding of the subject, autonomous across most areas of practice.	Internalises rules, applied holistically to view and grasp complex situations; context-dependent and moves between intuitive and analytical approaches with ease. Critically reflective and takes responsibility for going beyond existing rules/standards; interpretive evidence-based knowledge application; derives alternative approaches to complex practice.
6. Master Autonomous across areas of practice. can instantaneously process knowledge; consistent in producing proper perspectives and associated planned action.	No rules or formulae applied; recognises a potentially new perspective that better explains past events and better dictates future actions. Abstract reasoning, practice, and concrete experience. Subject to deliberation, senses opportunities beyond those that can be seen.

Table 1.4.4 Features Benner's (2001) Novice to Expert criteria used to monitor and measure the NMP students assessed level of performance. The model is applied to determine the way decisions are made, alternative decisions generated and how students are evaluated.

A cardinal skill in nursing is making high-level clinical decisions and judgments requiring rational thinking and access to experience, domain and subject knowledge (Eraut, 2004; Levitt-Jones et al., 2010). Benner's (2001) model describes activities consistent with knowledge theory (Ryle, 1959), Mezirow's (1990) framework and Blooms (1956) Taxonomy, which articulate context-specific (critical) thinking required for prescribing procedures. Benner's model includes high-level (mastery) components, skills needed for improved patient care and, more significantly, clinical and diagnostic accuracy (Salzman et al., 2011; McGaghie, 2015). As shown in Table 1.4.4, Benner's model plots the skill acquisition stages asserting knowledge embedded in expertise and guidelines for recognising proficient behaviour. The model also identifies transitional progress in the hierarchical movement from novice to expert (2001, p.13), which again lends itself to Mezirow's (1981) concepts of

reflection and Ryle (1949) and others knowledge theory. The disparity between academic performance and practice capability is a long-debated patient safety issue. Safe, independent NMP practice assumes that NMP learners are proficient at the point of qualification (NMC,2006, 2018).

Validity measurement is essential for ensuring optimum course outcomes (Hughes, 2013). The value of Benner's work to the IVA in this study is in making sense of the NMP students' capacity to perform clinically, purposely at an appropriate level (i.e., proficiently / toward mastery) cumulative to their pre-registration nursing registration and clinical experience (EHEA, 2010; Imison et al., 2016). As a method for measuring assessed work, Benner's framework is suitable for objectively determining academic expectations across the novice, expert to mastery stages of NMP professional development and providing additional contextual validity evidence for the assessment score interpretation and use. An important consideration is that Benner's theory is operationally and conceptually well-defined for content analysis (See chapter 4). Benner's model creates a performance measurement system for validating the assessed content based on the level of experience and proficiency demonstrated, providing a framework fit for the IVA.

1.4.5 Bloom's (1956) Taxonomy of Educational Objectives: Cognitive Domain

Bloom's Taxonomy (BT) of Educational Objectives is widely known in the education community. Higher education BT concerns students' learning, teachers' action in assessment development, and curriculum design (Seaman, 2011; QAA, 2014). Bloom's taxonomy has its antecedents in cognitive psychology and encapsulates mastery learning. It defines how the learner manages complex information and incorporates past learned experience with new knowledge creation (Dwyer et al., 2014). The BT cognitive domain taxonomy classifies students' cognitive behaviour, representing the learning outcomes and educational process (Bloom, 1959). The taxonomy's six cognitive categories, listed in table 1.4.5, range from lower-order skills, i.e., the less cognitive processing, to higher-order thinking (HoT). BT's lends itself as a

framework for measurement in the NMP assessment context, the educational objective and educational standard.

Table 1.4.5. Summarised from Bloom's (1956) Taxonomy of Educational Objectives, the Cognitive Domain

Classification	Taxonomy Characteristics
1. **Knowledge**	Knowledge; information recall of learned material; ways and means of dealing with specific facts.
2. **Comprehension**	Ability to translate, interpret, and explore.
3. **Application**	Applying principles, theories, or other abstracts to new situations; relating principles and drawing conclusions.
4. **Analysis**	Deconstructing material into its constituent parts and detecting how the elements are organised, relate to one another, and link to the overall structure or purpose.
5. **Synthesis**	Reconstruction of previous knowledge and experience; derivation of ideal, new material and experience into an integrated theoretical-methodological/framework or whole.
6. **Evaluation**	Make judgements about the value of ideas, based on a theoretical/ methodological framework, solutions to clinical problems; use specific internal/external criteria and standards for appraising quality, effect, and accuracy.

Table 1.4.5 Blooms (1956) Taxonomy represents another distinction when considering knowledge, in contrast to the Knowledge Theory. Bloom defines the kind of knowledge that forms the substance of what is taught and assessed in the master's curriculum; an important validity measurement of the NMP assessments.

BT as a framework affords a procedure for precision measurement of student capability regarding thinking. BT includes the principal 'critical' attributes of Mezirow, Ryle's 'strategic' and Benner's 'expert' theories in NMP assessed work (Schraw and Robinson, 2011). BT's six category alignment makes visible the NMP students' attainment of higher levels of thinking, e.g., synthesis or evaluation, where mastery exists (OECD, 2012;2017). Mastery learning is essential in healthcare, congruent with care excellence and the requirement to prepare practitioners for the professions' future (McGaghie and Kristopaitis, 2015).

The BT model offers versatility in its application, lending itself to measurement across several learning theories. Students are advanced when they demonstrate achievement in the preceding lower levels of BTs hierarchy (Bloom, 1956). In BT,

essential cognitive distinctions regarding students' NMP achievement are made across the assessed compositions. Conceptualising and explaining whether students' NMP behaviour is intelligent or (in)dependent depends on the assessed work's content and context (Bertucio, 2017). The six BT categories can provide additional evidentiary information for the IVA concerning the content and criteria in NMP assessments and students assessed work. Further, information about the quality of the NMP assessment, course content alignment and the intellective level of achievement as a course outcome with BT are also attainable (Hughes and Quinn, 2013).

1.4.5.1 Chapter Summary

This chapter explored and discussed the educational theories of Ryle (1959), Mezirow (1981), Benner (2001) and Bloom (1956). Contending their relevance to the NMP assessment and the part they play in representing the students' incremental development through the cited theories hierarchies and their integrated use to measure the assessed work for validity. The validity types are discussed and presented in section 1.3.2, along with the frameworks used as a point of reference to validate test score interpretation in a specific way. Section 1.3.3 of this chapter presented validity and illustrated Kane's (2006) IVA as a conceptual approach to interpreting student response to learning theory in the assessed tasks. In this context, section 1.4 presents Ryle (1949), Mezirow (1981), Benner (2001) and Bloom (1956) the theoretical frameworks appropriate for measuring the assessment data. Allowing data modelling at the individual's level provides a meaningful IVA of the individual's performance against the cited teaching and learning theories.

1.4.5.2 Conclusion

As discussed in section 1.3.4, the measurement of the NMP assessed work requires a modern approach. This chapter presented specific teaching and learning theories as available and conceivable measurement methods applicable to NMP assessed work. Chapter 2 presents a focused literature review drawing on those studies with significant findings concerning educational measurement. The measurement

methods will act as a barometer for the tools and conditions to make measurement possible in this study.

CHAPTER 2: THE LITERATURE REVIEW

2.0. Chapter Overview

The introductory Chapter 1 presented an overview of this study, identifying the research background and the research questions. Chapter 1 also discussed the influential theoretical frameworks in teaching and learning in healthcare, plus modern test theory (MTT). MTT is pertinent to this study to measure students' ability and achievement and establish the test validity of the NMP assessments. Chapter 2 reviews the literature in healthcare education and education in general, contributing to the understanding measurement in students' assessed work.

2.1 Introduction

The review's focus is the operational processes of measurement: Firstly, the measurement from a functional perspective, the thing to be measured, i.e., the WR and SCE assessments and secondly, the means of measurement: Identifying and applying the techniques in determining NMP students' achievement of educational standards. These first and second points are a significant factor in addressing the research question

How well do the Assessment Processes used in the Non-Medical Prescribing (NMP) Course Validate Certification/ Licensing of Students to Prescribe in Practice?

The literature review aims to answer the research question concerning what is measured and validated by researchers regarding content and performance standards. Content standards specify 'what' (i.e., theory), whereas performance determines 'how well' and adds the specification of the level of understanding (i.e., proficiency or mastery (Miller et al., 2013). In contrast, measurement obtains a description of the degree to which a student possesses the desired characteristics. The literature review question is concerned with MTT and its use in assessing summative work, specifically, methods or strategies for content in written reflections (WR) and the criterion used for structured clinical examinations (SCEs). The objective

of the review is to indicate the directions for measurement in this study. The rationale for the review design to address the research question is discussed in section 2.1.3.

2.1.1 The Review Aims

In exploring the trends of measurement used in assessment in the healthcare literature and the aim of the review are as follows: -

> To identify the methodological process for measurement and refine these for validating the NMP assessments

2.1.2 The Review Design and Methods

For guiding the search process, the PICOS (Population, Interventions, Comparators/Control, Outcomes and Study Design) tool for searching and selecting articles were applied (Centre for Reviews and Dissemination (CRD) 2009; Biggs, 2018). PICOS is used widely in nursing, social and healthcare research and several Variations exist. This study adopted the derivative **P**opulation **E**xposure and **O**utcome (PEO) as it was more suited to this non-experimental research design (Polit and Tatano-Beck, 2009). The PEO acronym listed in Table 2.1.2 underlines its application, frequently used in research to minimise bias in identifying, selecting, filtering, and removing articles found in the search strategy.

Table 2.1.2. Formulating the research question; the PEO consideration when searching for the literature.

P - Population	The Students Assessed Coursework
E - Exposure	The Assessment Method, i.e., WR, OSCE-based Assessment
O -Outcome	Measurement techniques used in assessed work content and assessments structure

2.1.2.1 The Search Strategy

The literature search involved the following databases: Education Resources Information Centre (ERIC) to access education research, Education Research Complete (EBSCO) for health, education and testing publications. For studies related to nursing and health education, in the Allied and Complementary Medicine (AMED); the Cumulated Index to Nursing and Allied Health Literature, (CINAHL) PubMed, Medline and ProQuest databases. SCOPUS was used to find articles cited from those retrieved. Studies relevant to professional and academic literature related to measurement and statistics; JUSTOR, Pub Psych, Web of Science. PsycINFO, specifically, for literature pertinent to psychometric testing and test item criterion measurement.

The researcher applied customised import filters for several databases and interfaces, including EBSCO, CINAHL, and Scopus, to export the searches as files into the EndNote reference management database. Each database differed in the format of the journal titles, author and other fields. Therefore, the structure, e.g., the abbreviated journal names, were corrected in EndNote to streamline the search (Bremer et al., 2016). Where duplications were found, these were deleted. The automated de-duplication steps are summarised as follows,

1. Download the references of the articles that were retrieved:
2. Go to 'All References' and click on one random reference.
3. In the menu, select References > Find Duplicates.
4. In the next window, the Find Duplicates selection window, click on [Cancel].
5. Press delete to remove all highlighted duplicates.
6. Repeat the process again from step 1 onward for other searches

(Bremmer et al., 2016).

2.1.2.2 The Search Terms

Table 2.1.2.2 lists the researcher's keywords in the two-tiered literature search to find publications relevant to the NMP assessments. The Boolean logic and truncation built around the listed search terms for search one, the WRs and search two, the SCE. The addendum filters are denoted by $ between and * after the keywords identified specific and alternative terms (Fink, 2010).

Table 2.1.2.2 Search Terms used for Peer-reviewed Studies

Search 1: The Written Refection's (WRs)	**Search 2**: The Structured Clinical Examinations (SCE)
Critical Reflection/Reflective Practice	Clinical Examination
Reflective learning//knowledge	OSCE/ Objective Structured Clinical Exam
Education/Reflective Writing	Practice Skills Development
Levels of Reflection/Evaluation	Educational Measurement/ Criterion Analysis
Reflective Framework	Psychometric Tests/Latent Trait
Reflection/ Content Analysis	OSCE/ Rasch Model/Modelling
Measurement	Objective Assessment/Measurement
Reflection/Assessment/Assignment	Item Response theory
Critical Reflection/Reflective Practice	Item analysis
Reflective learning/Reflective skills	Simulation

Table 2.1.2.2 Illustrates the main ideas of the research topic, reduced to the keywords used in the database search.

2.1.2.3 The Inclusion-Exclusion Criteria

The study's inclusion-exclusion criteria in Tables 2.1.2.3 and 2.1.2.3.1 contain date restrictions (2009-2018) and considered the assessments' design and representativeness to those within the NMP program (Hulley, 2013. The design and representativeness specifications further refined the scope to WR, and SCE assessed work to identify cohorts from different disciplines. The repeated search included 2019 publications using the academic search complete database for good education coverage. Although the search found some recent articles, the inclusion-exclusion criteria eliminated these.

Table 2.1.2.3 The Inclusion-Exclusion Criteria for Written Reflections (WRs)

Inclusion Criteria	Exclusion Criteria
Written Reflection	Patient's reflections, mothers, caregivers
Education Assessment writing	Focus Group, self-monitoring /awareness
Journal Entries	Experiences/insights of using reflective
Assessing Reflection	Portfolios, diaries, interviews
Reflective theory	Factors influencing reflections
- reflection, levels	Group reflection dialogues, blogs, book chats
- Critical - Content Analysis	Implementing, promoting, enhancing reflection
- Prescribing	Value of, lessons learned from reflection
Peer Reviewed	Discussion, opinion papers
Publications between 2009-2019	Publications outside the 2009-2019 parameter

Table 2.1.2.3 Illustrates the terminology used to include/exclude published material concerning WR assessment.

Table 2.1.2.3.1 The Inclusion-Exclusion Criteria for the Structured Clinical Examination (SCE)

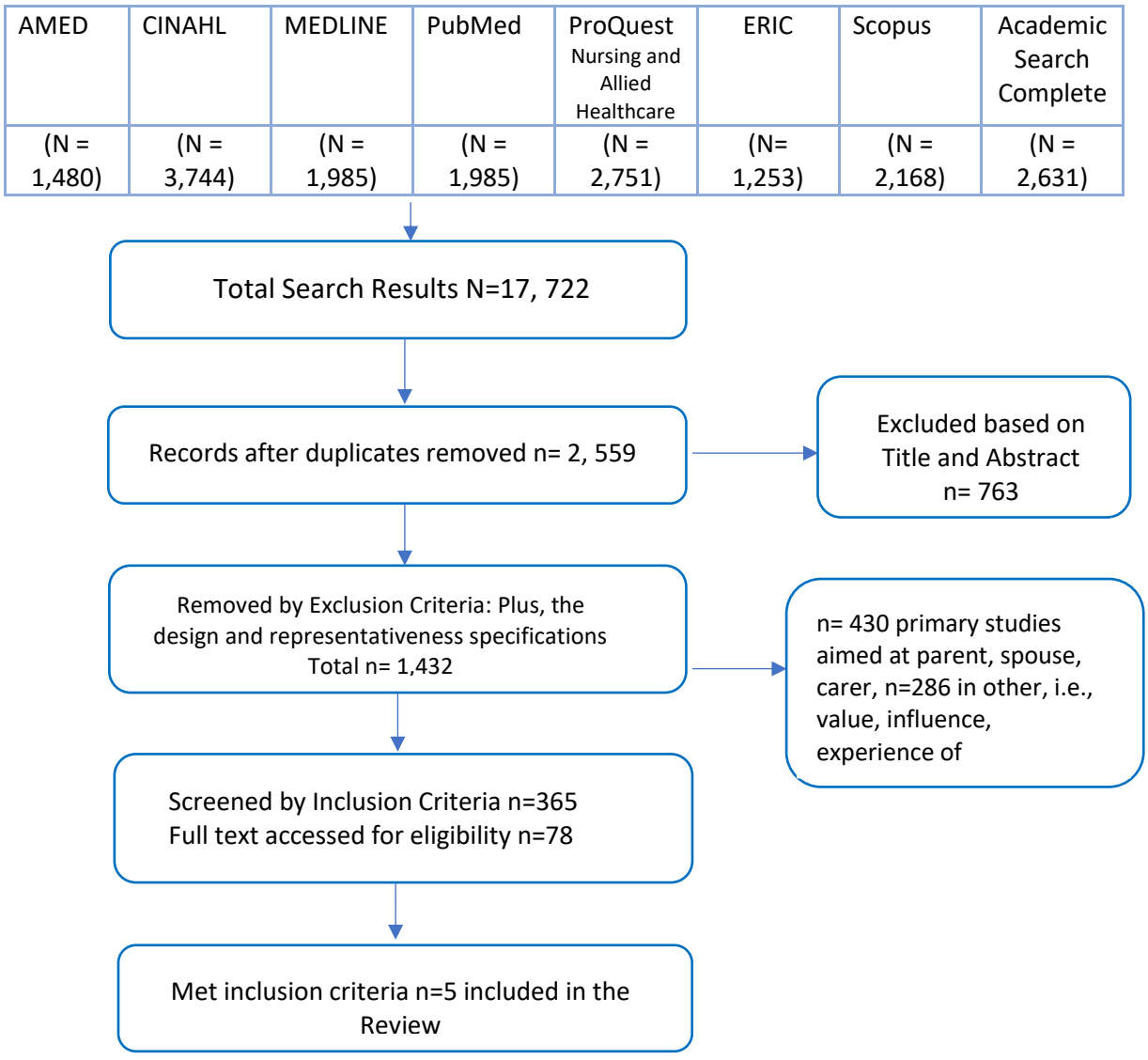
Inclusion Criteria	Exclusion Criteria
OSCE - Assessment	Measurement of ability scales;
- Validity	Sequence rotations influence performance
- Analysis/Analyses; Psychometric statistics, Rasch	Multiple-Mini Interviews as a scoring method
- Measurement, item response theory - Item Analysis -Criterion Analysis	Perceptions of, opinions of, questionnaires on, debriefing in.
Clinical Examination	Rating of assessors, station level analysis
	Web-based assessment, student satisfaction,
	Team achievement
Peer-Reviewed	Discussion, opinion papers
Publications between 2009-2018	Publications outside the 2009-2018 parameter

Table 2.1.2.3.1 Illustrates the terminology used to include/exclude published material concerning the SCE assessment.

2.1.2.4 The Search Process

In total, 14 papers were identified with n=3 for the WR search initially challenging to access. Two, because they were not full publications on the study as implied by the abstract and citation. The third paper was an erroneous citation. When found, the content reported on different outcomes to the focus of this study's review. The review included 11 articles. The flow diagram in Figure 2.1.2.4 shows the WR selection process's search results. Figure 2.1.2.4.1 shows the search, selection and results for the SCE.

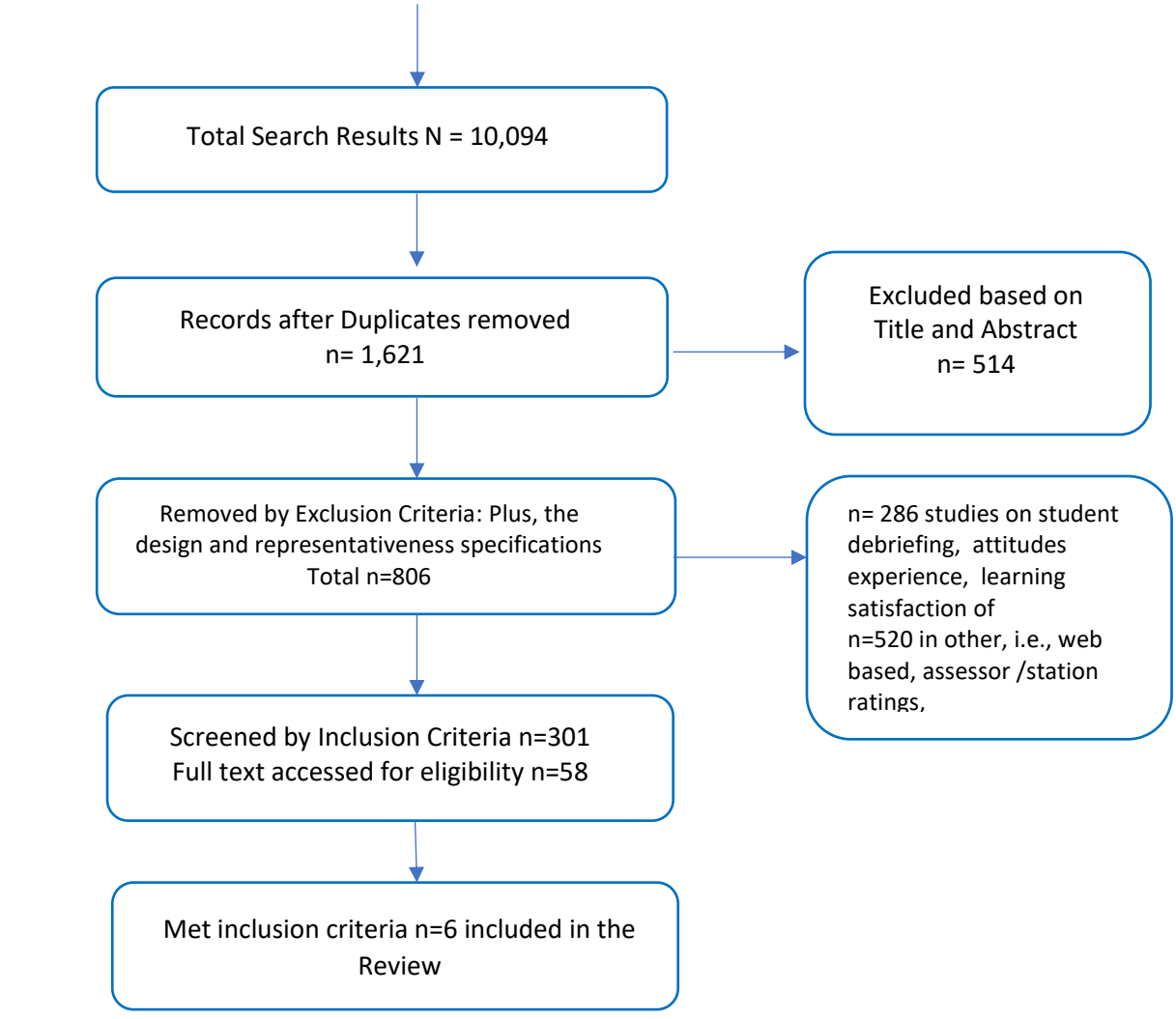
Figure 2.1.2.4. A Flow Diagram of the Search Process and Results for the Written Reflection



The n=5 WR studies originated in the USA (n = 3), Australia (n = 1) and New Zealand (n =1), reflecting a range of disciplines and the growing interest in the measurement of assessed WR content. The disciplinary fields and the focus of domains measured within the review's studies varied and included healthcare, education, design, and business. Table 2.1.3 summarises the final n=5 papers included in the WR review.

Figure 2.1.2.4.1 A Flow Diagram of the Search Process and Results for the Structured Clinical Examination

Psych Info	Science Direct	JUSTOR	PubMed	Pub Psych	ERIC	Web of Science	Academic Search Complete
(N = 1,170)	(N = 693)	(N = 1,144)	(N = 2,270)	(N = 1,673)	N=80	(N = 1,200)	(N= 1,847)



The search results shown in Fig 2.1.2.4.1 identified n=6 SCE papers for inclusion in the review; the studies in healthcare education were from Europe (n=2), Australia (n=1), USA (n=2) and Canada (n=1). These researchers measured a range of theory/domain skills with statistics, including test item analysis. Table 2.1.3.1 summarises the SCE papers reviewed.

2.1.3 The Review Design

Matrix Method (MM) is the design of choice. It allows the researcher to quickly extract exact information to discuss and synthesise methods and techniques suitable for the study (Gerrard, 2017, p147). The MM structure enables a critical review and synthesis of essential information specific to the research problem or question to assist in decision making in the methodology, design and methods. For details, see Garrard (2017). In this study, the matrix application is twofold:- (1) to help structure the critical analysis and synthesis of the selected papers, and (2) to focus on aspects of the articles related to finding suitable measurement methods and methodologies.

The research theory literature recommends using a critical appraisal tool (CAT) to assess the selected articles' methodological quality rigour and instil confidence in the review's studies (Aromataris et al., 2015). The chosen CAT is the Critical Appraisal Skills Program (CASP) checklist for systematic reviews, developed to appraise primary research designs. The generic format provided a structure for screening research reports methodological quality and their potential for application to practice (Lo-Biondo-Wood and Haber, 2017). The CASP tool also gives meaning to the information extracted while permitting the reviewer to specify items concerning the research question (Tappen, 2016; Briggs, 2017). The CASP facilitated the organisation of the material from several studies addressing the same or similar research question. The appraisal contributed to the decisions about the assessed work measurement methods (Newell and Burnard, 2011). The study's measurement problems and the research's validity question guided the information extracted from the papers. The information extraction was concerned with (a) assessment mode, (b) measurement method and (c) measurement outcome as listed in Tables 2.1.3 and 2.1.3.1.

Table 2.1.3: Studies Examining Written Reflection (WR) Content from a Multidisciplinary Educational Setting/ Perspective

Author, Date and Title	Research Design and Purpose	Setting and Participants	(a) Assessment Method	Focus Theory /Domain	(b) Measurement Method /Data Analyses	Findings/Results	c) Outcome
Cathro et al., (2017) [New Zealand] Assessing Reflection, Understanding Skill Development through Reflective journaling.	**Design**: Qualitative **Purpose**: to discover how students used reflective Journals and how these are assessed	**Business Education** International Communication Course N=31 written reflections Sampling method not explicit.	Summative Reflective Journals	Focused on Communication Competence-quality of communication Reflection, levels and depth	Kolb s (1984) Learning Cycle Content /Thematic Analysis (TA) Statistical Inter-rater reliability for coding the N=31 Axial coding in SPSS	L1. Concrete Experience-n=23/N=31 L2. Reflective Observation-n=20/N=31 L3. Abstract Conceptualisation-n=17/N=31 L4. Active Experimentation-n=9/N=31	Majority of students at L1 to L2 Students struggle to achieve L3 across communication themes. Marking Rubric base on Kolb Cycle developed
Barnes and Caprio (2016) [USA] Analysing service-learning reflections through Finks taxonomy	**Design**: Qualitative **Purpose**: to determine how students used reflective journals and how educators assessed these	**Teacher Education** Service learner course for secondary English teachers N=8 students N= 32 written Reflections Sampling method not explicitly disclosed	Written reflections, a primary objective of the study	Critical thinking and analysis of course content /material and experience.	Finks (2013) Taxonomy: six categories Content Analysis	1.Foundational knowledge n=3/N=8 2.Application n=3/ N=8 3. integration, n=1/ N=8 4. human dimensions n=4/ N=8 5.Caring n=4/ N=8 6. learning how to learn n=4/ N=8	Students academic content knowledge and critical understanding of their service work not demonstrated. Authors recommend explicit teaching of reflective mode.
Castleberry et al., (2016) [USA] Quantitative Analysis of Written Reflections during a Teacher Certification Program	**Design**: Mixed Method **Purpose**: To evaluate the success of a teaching certificate program by qualitatively evaluating the content and extent of the participants reflections	**Pharmacists Education** N=66 programs n=132 essays Sampling method not explicitly disclosed	Summative – Written Reflection Essay	Focused on Prompted Pharmacists teaching related activities Reflection, levels and depth	Kimber et al., (2008) 4 category Framework. Qualitative Content Analysis. Statistical Inter-rater reliability for coding =Y Cohens Kappa Mid-point -Difference in the extent of reflection Wilcoxon s signed-rank test	L1. Habitual action non-Reflection- L2. Understanding- L3. Reflection- L4. Critical Reflection- Cohens Kappa 0.74 for IRR, moderate	Not apparent how the Kimber et al., (2008) 4 category Framework was applied. The authors argue that the program fostered critical reflection but measurement unrelated to teaching and learning subject/content of reflections
McKay and Dunn (2015) [Australia] Student reflections in first-year, public health and health promotion unit	**Design**: Qualitative **Purpose**: To provide an overview of the use of reflection in an assessment task	Education and Clinical Practice -**Public Health** (PH) Students in PH, Year 1 Voluntary participation - N=15	Summative -Reflective Journal Writing	Focused on Engagement in Reflection in Public Health course, levels and depth Critical Thinking Critical Analysis	Kimber et al., (2008) 4 category Framework Content / Thematic Analysis (TA) Statistical Inter-rater reliability for coding =Y	L1. Habitual action non-Reflection- n=3/N=15 L2. Understanding-n=8/N=15 L3. Reflection-n=3/N=15 L4. Critical Reflection-n=0/N15	Students did not engage in Deep Reflections Some understanding of Public Health-measurement unrelated to course content, i.e., Public Health.
Gulwaldi. (2009). [USA] Using Reflective Journals in Sustainable Design	**Design**: Qualitative **Purpose**: To encourage reflection on the complexity of sustainability and sustainable design	Higher Education-Students from Students **Interior Design Program** Convenience Sample-N=30 n=226 journal entries	Summative, On-Line Reflective Journal Writing	Focused on Interior Design course Engagement in Reflection, levels and depth Critical Thinking Critical Analysis	Hatton and Smith (1995) Framework -four types of reflection Qualitative Content Analysis Inter-rater reliability for coding =N	L1. Non-Reflection- n=88 / n=226 L2. Descriptive Reflection – n=0/n=226 L3. Dialogical – n=118/n=226 L4. Critical Reflection n=15 /n=226	Assimilation of procedural and practice knowledge, at varying levels. Measurement of student reflection, not related course content/linked to the complexity of design solutions

Table 2.1.3.1: Studies Examining the Structured Clinical Examinations: A Multidisciplinary Teaching and Learning Perspective

Author, Date and Title	Research Design and Purpose	Setting and Participants	(a) Assessment Method	Focus Theory /Domain	(b) Measurement Method /Data Analyses	Findings/Results	c) Outcome
Shinnick and Woo (2018) [USA] Validation of a Time-to-Tasks Performance Assessment in Simulation: A Comparative Study Design	**Design**: Prospective comparative research **Purpose:** Assessment of Clinical Competence	**Nurse Education and Clinical practice** Year 3 Students n= 13 Qualified nurses of 5 years n=15	Simulation- subjective pass/fail and objective time to task.	**Clinical Judgement;** Lasater (2005) Model Levels of clinical proficiency, effective in the following 1. Noticing 2. Recognising deviations 3. Interpreting 4. Responding	- Descriptive statistics - Non-parametric analyses, Medians test and Chi-square - Sensitivity/specificity analysis for novice/expert differentiation - Statistical significance - standard Deviation	Correct distinction between novice and expert nurses Statistically significant differences Good correlation between reviewer in subjective pass/fail and objective time to task	The subjective measurement lacked ability to measure competency in specific aspects of theory and professionalism Subjective measurement should be avoided.
Radziej et al., (2017) [Germany] How to assess communication skills? Development of the Rating Scale ComOn Check.	**Design**: Rating Scale Development **Purpose**: to evaluate the validity of the CoMeD- OSCE communication tool for a wide range of topics.	**Medical Education** Sample N=69 Year 1 students. Randomised to into 2 groups	Simulation- patient /practitioner consultation. Video-taped Observation	**Communication** 1. behaviour 2. Conversational competence, 3. and Specific communication skills	- Statistical analysis - Pre and post assessment of communication tasks using a 5-point Likert scale. - Inter-class correlations coefficient — ICC for inter-rater consistency. - Item-level analysis, correlation coefficients	Statistical quality criteria met. Rating scale provided objectivity Precise behavioural anchor point descriptions for each OSCE Item.	Communication competence, is shown by a high correlation between the sum score and global assessment of the encounter. Results for the five-point Likert scale which can assess a more sensitive change is unreported
Yeates et al., (2017) [UK] Hawks, Doves and Rasch decisions: Understanding the Influence of Different Cycles of an OSCE on Students Scores using Many-Facet Rasch Modelling	**Design**: Multi-Site cohort study **Purpose**: Estimate the examiner's influence on students OSCE score	**Medical Education** Sample N= 16 Examiners 16 OSCE Stations	Summative Undergraduate OSCE	**Clinical Skills** 1. Communication skills 2. Procedural skills 3. Physical Examination 4. History Taking 5. Data interpretation 6. Managing Patients questions	- Many-Faceted Rash Model (MFRM) Student ability - Examiner severity/leniency - Station difficulty	Good Rasch model fit The OSCE station difficulty ranged from 0.88 logits to 0.48	Influence form different stations were greater than the influence of examiner cohort. Content from some Stations, more difficult than others. Students clinical ability unreported.
Judd et al., (2016) [Australia] The Validity of a Professional Competence Tool of Physiotherapy Students in Simulation-based Clinical Education: A Rasch Analysis	**Design**: Multi-Site Cross-Sectional Study **Purpose**: To Evaluate the validity of the Assessment of Physiotherapy Practice (APP) for student assessment in simulation settings	**Physiotherapy Education** Sample N= 1260 APPs	The Assessment of Physiotherapy tool (APP) used to measure professional competence in clinical practice.	**7 domains of practice** 1. Professional Behaviour 2. Communication 3. Patient assessment 4. Analysis and planning of interventions 5. Performing interventions 6. Evidence-based practice 7. Risk Management	- Rasch Analysis of the 20-item APP using a 5-point Likert scale. - Test item difficulty /student ability	Good Rasch model fit The APP provided evidence of construct validity for the examination of professional competence.	An overall good match of item difficulty to student ability. Only one test item demonstrated high difficulty bias. Achievement in domains of professional behaviour unreported
Wallenstein and Abner (2014) [USA] Objective Structured Clinical Examination Provide Valid Clinical Skills Assessment in Emergency Medicine Education	**Design**: Prospective cohort study **Purpose**: To assess the validity of novel management focused OSCE and evaluation of the instrument in emergency Medicine education	**Medical Education** Medical students, year-4 N=278 Prospective Cohort Study	Five Station OSCE- on Patient clerkship in Emergency Education	**Knowledge of:-** 1. Patient Medical History 2. Physical Assessment 3. Diagnostics/therapeutics 4. Communication (giving bad news) 5. Professionalism	-Downing (2003) 5 sources of validity evidence. Test Item Analysis:- - Item Difficulty - Item Discrimination - Correlation Coefficients - Standard Deviation (SD)/ OSCE and Clinical Score	Increased accuracy of skills assessment when compared to global rating score use	Test items anchored to specific clinical task promoted accuracy in the assessed competency parameters Students skills achievement unreported
Walsh et al., (2010) [Canada] The Novice Objective Structured Clinical Evaluation tool: Psychometric Testing	**Design**: Instrument design **Purpose**: to establish the validity and reliability of the OSCE assessment tool	**Nurse Education and Clinical Practice** Convenience sample N=565 year 1-Student Nurses Data collected over four sites	OSCE to assess clinical competence	**Clinical Competence** in areas of knowledge in action /safety in clinical practice 1. Safety, 2. Asepsis, 3. Knowledge 4. Organisation 5. Caring	- Likert scale 0-4, competence level - Test item Exploratory Factor Analysis using Principal Axis Factoring (PAF) - Measures of Sampling adequacy (MSA) - Kaiser-Meyer-Olkin (KMO) statistics. - Bartlett s -Cronbach s Alpha for Inter-item correlations	High Construct Validity Factor 1 – 18 items related to complex skills safety received the highest factor loading 0.75 to 0.802 Factor 2 – 4 items Considers the lowest factor loading 0.473 and below	Identified novice- expect practice in only one of the simulated events. The tool did not demonstrate enough novice-expert specificity in the remaining 3 OSCE events

2.2 The Literature Review

2.2.1 Background: Educational Measurement

Educational Measurement is a discipline concerned with measuring learners' achievement (e.g., NMP proficiency) in education (Fragkos, 2016). The philosophy of educational measurement is the interrogation of assessment on a detailed level. The purpose is to arrive at a defensible conclusion about student ability and achievement based on assessed work samples (Cizek et al., 2015). In education, measurement refers to applying scales and tools according to practices accepted by those conversant in this area (Miller et al., 2013). Therefore, this literature review focuses on identifying a precise and objective approach to inform NMP educational measurement.

Several authors from various disciplines discuss the nature of educational measurement, the rationale and the associated inherent problems. Wollack's (2011, 2015) work identifies practical issues such as test tampering as a rationale for educational measurement engagement. Publications by Lane (2003, 2015) advocate educators' responsibility for providing evidence that assessment and test measure the standards against which student achievement is assessed. The literature in English language and linguistics presents several educational measurement papers concerning educators' accountability for ensuring alignment between assessment and learning (Deville, 2011; Challoub-Deville, 2016; 2018). In comparison, others attend to measurement as a method for improving course quality and content. See, for example, Manzano-Vázquez (2014); Dearden, (2015).

The work of Stone and Xu (2011) Stone et al., (2012, 2013) offers a mixed and complex view of psychometrics in educational measurement, ranging from their benefit in place of summated test scores. Including the significance of psychometrics to cognition and intelligence (ability) testing in individual performances. Stone et al., also address psychometrics' limitations concerning the data classification requirement. Their work includes the importance of the model fit statistic with the competing psychometric models. The constraints discussed by these authors incorporate the problems measuring the psychometric dimensionality, invariance and the impact on the results' validity and reliability.

Other authors give psychometrics great attention in addressing assessment problems. For example, the interface between data, i.e., test score and the attributes of interest (Zumbo and Forer 2011). Targets for assessment measurement or improvement (Hamilton et al., 2012; Reise, 2012: Koretz, 2015; Reise and Rodriguez, 2016). Shepard's (1997, 2000, 2017) represents decades of contributory work in educational measurement concerning student achievement called for more complex proficiency demonstration, disciplinary practice, and content. Citing Pellegrino et al., (2001) influential work Shepard (2016, 2017) advocates for assessment systems that provide actionable insights into what the students know and can do. This summary highlights several educational measurement subjects, indicating that the topics are varied, and this review focuses on articles that attend to the practicalities of measurement for application to this study's research question.

2.2.1.1 Studies Examining Written Reflection (WR)

This study's difficulty pertains to the WR content, the results obtained with measurement, and the techniques derived from applying reflective theory models. The subject reflection has produced many theoretical framework options for coursework design and assessment (Oermann and Gaberson, 2014). Each framework has its academic standards, for which agreement does not necessarily exist, making measurement more complicated (Nitko and Brookhart, 2011). Therefore, examples are drawn from the literature concerning the frameworks used for WR assessment measurement. In this section 2.2.1.1, the review aims to ascertain how the reflective framework criteria are described and applied to the content measurement (see 2.1.3); to address the validity question in this study.

Reflective writing as an assessment is widespread across the healthcare curriculum and disciplines (Fragkos, 2016). In research, nurse educators have yet to mine the written compositions to examine the content and the students' reflective application and ability in the writing processes across the reflective theory and genres (Ng et al., 2015). A limited number of studies for review were found which discuss content measurements of the written compositions, as illustrated in Fig 2.1.2.4. The depth and level of students' reflections are absent and therefore is problematic. Yet measurement of the level

of reflection is closely related to the quality of written work (see Fook et al., 2006; Moon, 2006; Dyment and O'Connell, 2011).

Gulwaldi et al., (2009) convenience sampled thirty students and inspected 226 WR journal entries submitted in the final year sustainable design studio course. The sampling method is consistent with the qualitative design and the data analysis focused on the Hatton and Smith (1995) four levels of (1) non-reflection (descriptive), (2) descriptive (limited), (3) dialogical and (4) critical reflection. Two persons independently coded these aspects of reflection for a quality control purpose. However, the coding's reliability is questionable. The most-reported quality control method for the coding cited in the research literature is interrater reliability (IRR). The IRR calculations are an accepted indicator of accuracy, precision and reliability (Salkind, 2017). The limitations in ethical rigour are that ethical approval for the study is unreported. The credibility of the coding in the Gulwaldi et al., study is also a limitation. These missing research principles may affect the analysis, confirmability and consequently, the results (Parahoo, 2006; Polit and Tatano-Beck, 2009).

In the Gulwaldi et al., (2009) study, the WR content analysis and measurement of the reflection levels mapped against the Hatton and Smith (1995) framework consisted of three paragraphs. Therefore, the Gulwaldi et al., study's strength is its consideration of reflection within a reflective structure and measurement of the degree to which it occurred in the students' work. Also of significance to this study is clarity in the Hatton and Smith model applied to the WR text and data analysis; there is an element of credibility (Parahoo, 2006). The reported results showed that journal entries demonstrated all four levels of reflection. Many reflective entries [n=123/N=226] fell within the lowest reflective forms, i.e., the descriptive and dialogic categories. Gulwaldi et al., (2009) result illustrated that the reflective model/ framework application on theoretical grounds has merit for WR content measurement.

In their study, McKay and Dunn (2015) comprised a convenience sample of fifteen reflective journal entries of 1000 words from students' in their first-year public health course. The sampling and data collection method was consistent with their qualitative research design (Polit and Tatano-Beck, 2009). These authors applied the Kember et al., (2008) model in the following category levels: (1) Habitual action, (2) understanding, (3) reflection and (4) critical reflection to the WRs to determine the extent of students'

engagement in reflection. The data analysis method involved two phases. Phase 1 was a line-by-line coding based on the four Kember et al., (2008) classifications. Phase 2, the investigation of common themes in the WRs by thematic analysis, informed by Corbin and Strauss (1990) grounded theory (GT). Corbin and Strauss' open coding identified, labelled and grouped the themes, and the final selective coding conceptualised their relationship. The written content measurement consisted of examining the themes, the relationship between the Kember et al., and the assessments ' reflection.

The literature advises coding for as many potential themes/patterns as possible; a criticism of this study is that it ignores the tensions and inconsistencies between these approaches. For example, Corbin and Strauss' method is theory-generating, applied in the absence of a predetermined framework, to explore concepts hidden in the data directed towards theory development. Thematic Analysis (TA) confirms themes identified by theory from existing literature (Holloway and Todres, 2003). Strauss and Corbin make clear distinctions between GT and thematic analysis in their methodological characteristics. The coding technique is significant to accurate content measurement. In their work in research theory, Strauss and Corbin argued against equating TA with Grounded Theory (GT) for ontological and epistemological reasons. McKay and Dunn's rationale for the TA and grounded theory hybrid is unclear, raising questioning about the data coding credibility and confirmability. The latter is about the conclusions and interpretations of the findings (Holloway and Todres, 2003). Moreover, evidence of quality control, i.e., data analysis by more than one researcher, e.g., confirming the agreement in code application and meaning, is not discussed. Agreement confirmation is an essential aspect of data analysis for establishing the results' reliability (Polit and Tatano-Beck, 2009).

In their results McKay and Dunn identified that three of the fifteen [n=3/N=15] journals showed Kember et al., (2008) level 3 reflection; the remainder were either habitual action, non-reflection [n=3/N=15] or understanding [n-8/N=15]. However, the excerpts used to illustrate student reflection levels were unrelated to public health and health promotion. McKay and Dunn have not established the content (validity) measurement source in their study to the specified educational theories relevant to that testing situation. Therefore, measurement information is applicable for NMP assessment, i.e., how the model is applied to examine students' relationship with reflection and their reflective ability. Also, the measurement of students' achievement in subject theory at each stage of Kember et

al., the reflective process is unclear, which may impede the transferability of the methods used (Torbin and Bagley, 2006; Poilit and Tatano-Beck, 2009). A further limitation is that the application of Kember et al., the reflective framework was focused solely on students' demonstrating understanding of their career aspirations, community and teamwork experiences.

Castleberry et al., (2016) also applied Kember et al., (2008) in their mixed-method study to measure the extent of reflective ability among pharmacy students on the teacher certification program. The sample comprised 132 WR essays collected at the midpoint and final essays submission, exported into NVivo for analysis. A clear statement of attainment of ethical approval is absent, not indicative of rigour or quality research (Parahoo, 2006). In applying the Kember et al., four-category reflection (see McKay and Dunn 2015 for reference) for coding the WRs, two investigators independently identified the 28 themes discussed in the WR essays. Unlike Gulwaldi (2009) and McKay and Dunn (2015) methods for quality control in coding, Castleberry et al., conducted IRR measurement using Cohen's Kappa. The results are stated as 0.74 (p.3), indicating substantial code reliability and coding and integrity (Castleberry et al., 2016).

Castleberry et al., (2016) undertake Wilcoxon's signed-rank test measurement of the program's median difference in mid and end-point reflection. Wilcoxon's rank-sum test analysed the extent to which reflection varied according to gender and experience for content measurement. Castleberry et al., identified twenty-eight themes. The analysis focused upon the prompted teaching related areas- assessment, feedback strengths, weaknesses. The themes were percentage weighted by the number of times they appeared in the WRs. The highest-ranking subject related to teaching activities n= 126 entries (98.5%) of the N=132 occurred in weaknesses. The lowest was feedback, n=118 (89.4%). However, the reported results are unrelated to the measurement of reflection, and the results in this study are disappointing. The study limitations are its overly quantified reporting of the technical measurement efforts; rather than the qualitative content results, despite the stated mixed method design (Castleberry et al., 2016, p.2). The over-emphasis on the quantitative analysis and measurements' makes the study insensitive to the deep, quality reflection measurement via the reflective theory in Kember et al., (2008) s model (Bradbury et al., 2010). Castleberry et al., application of the Kember et al., model is unclear and not replicable as a measurement method in this proposed NMP study.

In contrast to the above-reviewed studies, Cathro and O'Kane (2017) applied Kolb's (1984) four reflective elements: (1) concrete experience, (2) reflective observation, (3) abstract conceptualisation, and (4) active experimentation. Kolb's four components are applied to qualitatively assess the level depth of reflective ability in international communication course among business students. The single-page WRs, representing 10% of the course mark, formed the basis of the study's data. Ninety-eight students enrolled in the course, 31 students consented and participated in the study following ethical approval. The reflective essays were anonymised by a unique identifier before export into NVivo for data exploration, establishing research ethics rigour (Holloway and Todres, 2003). In Cathro and O'Kane's application of the Strauss and Corbin (1990) GT coding methods, the coding method differed from Castleberry et al., (2016) and McKay and Dunn (2015). Cathro and O'Kane used open coding in the NVivo environment. The axial coding is represented by 53 nodes (themes) generated within NVivo. By not statistically establishing the two independent coders' reliability, methodological integrity is questionable (Holloway and Todres, 2003).

Cathro and O'Kane (2017) stated themes, i.e., the theoretical concepts were consistent with the subject of international communication across the WRs, and primarily at the first two levels of Kolb's (1984) framework. That being (1)concrete experience and (2) reflective observation. Of the 31 students WRs, 29 of those analysed reflected about themselves [n=16] or others [n=15]. Only two students achieved level 3 abstract conceptualisation in interpersonal communication, the skill which elicited the most depth of reflection. None of the students WRs demonstrated Kolb's level 4, active experimentation across any business communication themes. However, the Cathro and O'Kane content measurement centred on the students' level of reflection' interpreted by Kolb's four reflective cycle levels, confirming reflective models as a plausible measurement of NMP assessed work.

Barnes and Caprino (2016) used Fink's (2013) model in their qualitative study examining the reflections of undergraduate student teacher candidates (TC) on an in-service learning course for English teachers. Fink's taxonomy contains six categories: (1) Foundational knowledge, (2) application, (3) integration, (4) human dimensions, (5) caring and (6) learning how to learn. Consistent with the qualitative design, consent to participate in the study from a convenience sample of N=8 undergraduate teachers confirmed ethics rigour (Polit and Tatano-Beck, 2009). The data consisted of the students assessed WRs, and the

researchers' applied an interpretivist epistemological analysis. Two authors searched for themes in the WRs, generating forty-one member-checked codes. Their discussion of the topics found signifying that the data generation is rooted in the identified research design philosophy (Holloway and Todres, 2003).

The sampled WRs n=4/N=8 contained all six of Fink's reflective model criteria. Barnes and Caprino (2016) found twelve of the forty-one codes related to foundational knowledge about academic learning. With knowledge reported as the most frequent, although the instances are unreported. The application (10 instances) and integration (12 cases) in the educational context are documented. The Barnes and Caprino article unfolded some of the ambiguity concerning measuring NMP students' reflection. Specifically, generating coding for text mining and the Fink (2013) reflective model's precise application. Omitted in the measurement are the different levels, i.e., superficial or deep; distinguished by the Fink reflective models, which, according to Leijen et al., (2012), is a dimension for determining the WR quality. What is established is credibility in 'fit' between WR content and the researcher's observations and findings in the context of the Fink framework. Therefore, the meaningfulness of the results is increased, and therefore, aspects of the credibility of data analysis methods (Torbin and Bagley, 2004).

2.2.1.2 Studies Examining Structured Clinical Examination

Healthcare courses with a clinical component frequently use (Objective) structured clinical examinations (designated as SCE in this study) to evaluate students' work-safe proficiency. The educational measurement here is concerned with the SCE and the methodology used to measure test items (Markus and Boorsboom, 2013). The review questions focus on the psychometric properties and relate to decisions about *what* is assessed in criterion analysis and 'statistically' how? The aim is to use the literature to find a conceptually oriented description of the psychometric model to address the research question. The fulfilment of the measurement requirements includes the assessment (e.g., sample size) and the methods applied (e.g., selecting an appropriate statistical technique) to establish the SCE validity. In this section, 2.2.1.2, the six research articles reviewed and summarised in Table 2.1.3.1 represent several healthcare disciplines.

Walsh et al., (2010) tested a novice OSCE tool's psychometric properties in year-1, term one undergraduate students on a clinical nursing course. Attainment of ethical approval from the participating universities is stated, indicating that the participants' rights are addressed appropriately (Burns and Grove, 1999). A convenience sample of 565 implied a non-experimental design. The sample comprised participants across four university sites. Walsh et al., used measurement of sample adequacy (MSA) to demonstrate methodological rigour in the sample representativeness. The MAS results ranged from 0-511-0.974. However, two of the 25 OSCE items were <0.60, below the MSA statistics and were, therefore, removed. Moreover, the OSCE tool's statistical measurement, the exploratory factor analysis (EFA), is consistent with the quantitative methodology.

The Walsh et al., (2010) study's strength is the statistical, methodological quality. For example, the Kaiser-Meyer-Olkin (KMO) statistic established the EFA sample size's reliability. Yielding a reported result of 0.940 in n=23/N=25 OSCE items, a value above the minimum of 0.5 indicating EFA should generate distinct and reliable factors (Field, 2013). KMO includes a performance of Bartlett's sphericity test for indications that the variables are unrelated. Principle Axial Factoring (PAF) accounts for differences in measurement, and algorithms compared the results between factors. Walsh et al., reported a statistical significance of $p < 0.001$ that the data are suitable for EFA.

Further, Walsh et al., applied Cronbach's alpha to calculate and examine the items' internal consistency; the calculations ranged between 0.9380 and 0.9429, indicating the OSCE items' high reliability. In combination, these statistical measurements infer the reliability of the results and statistical methods and validity (Greenhalgh, 2007). The Walsh et al., study aimed to report the undergraduate OSCE tool's novice properties via psychometric testing. In EFA, the factor loadings are theory-driven by the OSCE questions construct. For example, factor 1, patient safety, was included n=18/N=23. The highest loading was above 0.71, in n=8 items within the task competencies of knowledge, environmental, physical and psychological care, clear explanation of asepsis principles. A study limitation is the documented method for evaluating competencies, based on the Likert scale of 0=not at all and 4= excellent. However, these findings are unreported in the results or present in the discussion.

The theory-driven OSCE items were developed and reviewed by academic and expert clinicians. The subsequent EFA helped refine, construct and validate the assessment

(Field, 2013). The EFA enabled a determination of the interrelationships among measures. However, EFA may only reflect variation in a smaller number of latent variables. A further limitation of the Walsh et al., (2010) study is the exclusion of Confirmatory Factor Analysis (CFA). The sample size was sufficiently large to apply CFA and the requisite split-half data method. I.e., performing EFA on half the data and decisions about results confirmed and validated by CFA (Field, 2013). CFA is a more robust, multivariate model used when numerous variables are measuring more than one dimension. For example, the OSCE tool is 'novice specific' characteristics and other variables regarding the test item's question and constructs (theories). Including those related to student ability or other latent variables (Osbourne, 2008; Matsunaga, 2010). A significant limitation of the study is that the student's response patterns associated with the latent (novice) factor and the OSCE tools performance are unmeasured and unreported

Wallenstein and Ander (2014) conducted a prospective cohort study and attained ethics approval suggesting ethical rigour in this context (Burns and Grove, 1999). The sampling method is unspecified, and the population involved N=278 year-4 medical students completing their Emergency Medicine (EM) rotation. The OSCE cases, designed by the Department of Emergency Medicine Education Committee, aligned with the national curriculum/learning objectives for EM clerkships and represented a broad spectrum of disease. The study used two of Downing's five validity sources to determine whether the non-traditional OSCE format was valid for assessing EM trainees' clinical skills (Downing, 2003, p. 803). Measurement comprised of (1) the content, (2) response process, (3) internal structure, (4) determine the relationship to other variables and (5) consequences. The OSCE design integrity, e.g., quality control measures, are not stipulated. The instrument's reliability to consistently and accurately measure the concept of interest is questioned (Polit and Tatano-Beck, 2009).

Wallenstein and Ander's (2014) OSCE utilised a nine-item anchored Likert scale to score the assessments, with subscores in each competency area:- medical knowledge, clinical practice, procedural skills, communication, and a 5-item scale to assess professionalism. To test the power of the association and rule out threats to validity, Pearson's Correlation Coefficient measured the correlation between the final OSCE score and the final clinical competency score (Greenhalgh, 2007). The researchers report a statistically significant linear relationship between the OSCE score and ED performance score [r(239) =0.40, p

<0.001]. To detect the actual difference between the OSCE and clinical score, standard deviation (SD) quantified their variations. Positive correlations were reported; the mean OSCE =0.75 [SD=7.8] and the mean clinical score 8.1 [SD=5.4].

The items analysis (IA) within the five-station results were a mean index of difficulty 63.0 [SD=23.0], and the mean index of discrimination was 0.5 (SD=0.21). When compared to best practice, the difficulty index 24 (70%) of the test items were at level 1 (mid-range), the most recommended difficulty level, according to Wallenstein and Ander (2014, p124). Twenty-six (76%) of the test items demonstrated very good discrimination between high and low performers in the discrimination index. Wallenstein and Ander state their provision of evidence for two of Downing's (2003) validity sources; (3) the internal structure and (4) relationship to other variables. A significant weakness in this study is the interpretation and discussion of the OSCEs internal structure's validity claim. A deficiency in this study is the lack of inter-rater reliability of the OSCE, plausibly negating the instruments' internal validity claim. Similarly, the reliability of the correlation analysis is for this reason (Polit and Tatano-Beck, 2009).

Judd et al., (2016) aimed to test a professional competency assessment tool; the sample comprised the Assessments of Physiotherapy Practice (APP) tool from students in year 3 of a 4-year undergraduate program and year-2 graduate entry program. The study obtained ethical approval from two universities, written consent from all participants, and anonymised the APP, indicating ethical rigour. The sampling method is not defined but comprised of n=971 students in simulation and n=289 from clinical practice. The study's replicability is an issue. The reported results do not explain the sample's delineation between the specified groups and related university sites (Polit and Tatano-Beck, 2009). The 20-item APP purportedly measured professional competence across five domains of practice: professional behaviour, patient assessment, analysis, intervention planning and performance, risk management and evidence-based practice. A performance level indicator Likert scale scoring from 0-4 is applied. Zero denoted competencies infrequently/ rarely demonstrated, and four signified excellent.

The APP data was imported into Winsteps software for Rasch analysis to distinguish accurately between student performance levels. A goodness-of-fit calculation confirmed statistical validity between the Rasch model in the APP data, deemed good at > 0.91 in the person and item reliability indexes with the person separation index at > 2.0. The APP

data was described as unidimensional, meaning there was one latent variable of interest, i.e., clinical competence. In demonstrating methodological rigour, the rationale and the choice of the data analysis model and method should be explicit (Russell, 2005). The Rasch model used is unspecified. For example, several extended Rasch models suited and specific to unidimensional data analysis and the goodness of fit calculation determine which Rasch model applies to the data. As the model is unspecified, there may be implications concerning the results' reliability (Bond and Fox, 2015).

The Judd et al., (2016) study has several limitations. The results and the discussion report consistently only one of the three simulation formats, the longitudinal simulation. The APP scoring ability showed flaws when used in shorter simulation formats or at graduate level ability. Judd et al., suggest altering the APP test's competency benchmark to compensate for weaker students in the early stages of clinical training (p.9). Adjusting the benchmark to reflect the less able student is counter-intuitive to the study's construct validity aim (Kane, 2006). The five competency domains do not feature in the results or discussion. It is unclear whether the APP measured the professional behaviour and other specified criterion or at what level; zero or excellent. However, the authors conclude that the APP results demonstrated construct validity and accurate measurement of physiotherapy students' professional competence.

The Yeates et al., (2017) cross-sectional cohort study design examined the impact of examiner cohorts on students' scores. The data sampled were summative OSCEs from an undergraduate medical school, analysed using the Facet Software available in Winsteps. The observed averages were the researchers' primary measure of interest. The raw mean of items within each facet; the defined aspects are (1) students' ability, (2) OSCE stations, (3) site locations and (4) leniency/severity in the examiner cohorts. There is some evidence of ethical rigour as approval was sought and denied. The study performed secondary data analysis- the student OSCE data is anonymised for analysis, suggesting that confidentiality was maintained (Parahoo, 2009). The OSCE data comprised 235 students. The study population consisted of 16 examiners whose OSCE observation varied from 8 to 16 students, thereby allowing facet comparison analysis using psychometrics to explore variance between the four facets. Although the data was secondary, the individual participant's rights should be considered (Greenhalgh, 2007). The OSCE followed the blueprint, designed against the medical curriculum's learning

outcomes suggesting a degree of reliability in the OSCE; however, quality measures for the OSCE as a measurement tool are undocumented.

The learning objectives set for the students were communication, history taking, physical examination, and procedural skills such as data interpretation and patient management. A seven-point Likert scale ranging from definite fail (1 and 2) borderline fail (3) satisfactory (4) Good (5) and excellent (6-7) provided the overall judgement of the student's standard of performance. The Yeates et al., (2017) study's strength is MFRM modelling, a variant of Item response theory (IRT). Yeates et al. rationalise its use and benefits in simultaneous consideration of multiple variants, i.e., student ability, examiner leniency/severity, station location and station difficulty influences (Osborne, 2011). Methodological rigour, i.e., the data model fit calculation in the data analysis, is evident. The MFRM predicted performance well in the mean-square (MnSq) infit or person sensitive and outfit for lucky guess parameters; the expected value is close at 1.0. The MnSq parameters were between 0.5 (overfit), where the model predicts the data too well. The underfit of 1.5 indicates other sources of variance in the data, i.e., students' erratic behaviour. For example, students of high ability scoring well on complex tasks but poorly on an easy task (Field, 2013).

Yeates et al., (2017) report that different stations' influence was more significant, indicating content (subject domain) specificity in the OSCE stations. The results suggest that the OSCE stations ranged in difficulty from 0.88 [SE ±0.06]. A lower logit score tended to reduce students overall station scores. Stations with a higher logit score of 0.48 [SE ±0.07] increased students overall scores. The medical school site location represented a lesser influence. These results ranged from 0.09 [SE ±0.03] to 0.07 [SE ±0.03]. A drawback in the study is that the analysis between the facets, e.g. (1) student ability and (2) OSCE is unreported in the results or visible in the discussion. Additionally, the 7-point Likert scale indicates an acceptable standard of performance. The student's learning outcome or achievement is unmentioned.

Radziej et al., (2017) developed an assessment tool with a content-specific rating scale to assesses a wide range of doctor-patient communication interaction/skills among medical students. Year -1 medical students [N=69] were randomised into two groups to evaluate the new ComOn tool compared to the existing Communication Skills Training (CST) method. The study obtained ethics approval and gained consent from students for

videotaping during the assessed encounter of a consultation. The researchers videotaped 135 students' for grading global performances. The OSCE had a five-minute limit for the consultation-history taking. The researchers used a 10-cm visual analogue rating instrument ranging from zero, very bad, to 10 being excellent.

SPSS accomplished the statistical analysis. The power of this study, for the most part, is the clarity of the described methodology. For example, in the ComOn instrument development, the 13 patient-centred communication items underwent reliability checks. Radziej et al., (2017) document inter-rater correlation using Cohen's Kappa, recording a high agreement in n=3/N=13 (63-77) items, moderate in n=6/N=13 (45-53) and fair n=4/N=13 (33-37). The ComOn items were phrased and scaled according to a five-point Likert scale from zero=poor to 4=excellent to assess specific improvements in communication skills for each student. Spearman's rank correlation matrix measured the associations between the subscales and the ComOn items. The Correlation coefficients were small if $|r| \geq .10$; medium if $|r| \geq .30$; and large if $|r| \geq .50$. Correlation coefficients assess consistency between raters intra-class and estimate the IRR while being less influenced by systematic leniency/strict rater error (Radziej et al., 2017, p6).
Radziej et al., (2017) rejected Cohen's Kappa, widely used for the inter-rater correlation (IRC) measurement for demonstrating the consistency and reliability of the tests used. The study falls short in this aspect. Authors should cite the classic Fleiss's kappa or other methods in demonstrating methodological rigour (Hallgren, 2012). In discussing the results, Radziej et al., report Pearson's correlations within a range of possible scores between subscales for specific communication skills. For example, the conversation structure, patients' emotions, and offers to ask questions were less than 0.44, and $p < 0.001$. The authors conclude that communication competence is shown by a high correlation between the sum score and the global assessment. The five-point Likert scale results, which can assess a more sensitive ability change, is unreported.

The Shinnick and Woo (2018) prospective comparative design recruited n= 15 qualified nurses of > 5 years' experience from Intensive care units (ICU) and the Emergency Department (ED) of two different institutions. Including n=13 student (novice) nurses who passed their summative assessment and were nearing graduation. Fourteen instructors blinded to group assignment reviewed the 28 simulation videos of the sampled population. The study aimed to compare two commonly used assessments: The objective Lasater Clinical Judgement Rubric (LCJR) and the subjective Pass /Fail. The overall aim

was to validate assessment tools specificity to identify experienced nurses, detect novice nurses, and differentiate between the groups in managing patients' with decompensated heart failure (HF). The study is missing a definitive sampling method for all the participants. However, ethical approval was obtained for students' recruitment via email, indicating that the participants' rights are maintained (Parahoo, 2009). Further, to overcome the subjective pass/fail measurement of competence in clinical assessment, the assessment instrument utilised the Lasater Clinical Judgement Rubric (LCJR).

Shinnick and Woo (2018) used 4 of the 11 domains in the LCJR (1) Information seeking, (2) clear communication, (3) evaluation/self-analysis and (4) commitment to improvement, rationalising that not all 11 domains were pertinent. By novice to expert group, for each reviewer, the mean scores were calculated, resulting in a statistically significant $p< 0.001$ for all reviewers, between the groups. An overall mean for an expert was [22.58 ± 8.14], indicating fair ability (0.72) to identity an expert and low specificity to identify Novice [14.9 ± 14.6], an inability of 0.04. The IRR results showed an 85% agreement for subjective assessment Pass/Fail but only a 29% agreement in scoring for the LCJR. Shinnick and Woo cite unfamiliarity with the instrument for the 29% LCJR agreement. The LCJR, 29% agreement, suggests the validity and reliability are unconfirmed. The findings are consistent with others, e.g., Adamson et al., (2012a); Strickland (2013); Kristiansen et al., (2015); Prion et al., (2017) and Roman-Certo et al., (2018). A significant limitation in this study is LCJR measurement and the omission of the student's performance in the four domains relative to the novice-expert achievement.

2.2.3 The Discussion

Essential for any testing or assessment is to impose standards for achievement upon its examinees'. Studies can profit from executing statistical methodological procedures as the gold standard for establishing the test data's validity. The measurement of causal inference of interest is students' reflections, such as those identified by Castleberry et al., (2016). The statistics are applied to demonstrate measurement rigour but lose faithfulness in measuring and analysing the phenomenon of interest. Alternatively, the Wallenstein and Ander (2016) and Shinnick and Woo (2018) OSCEs, aimed to measure student competency in specific aspects of theory and professionalism. The accepted measurement approach is predictive validity, the degree of correlation for the concept's

measured, i.e., theory and professionalism. However, the statistical methods employed are limiting. The results are reduced to the correlation between OSCE score and performance score, rather than findings concerning the students' achievement levels, e.g., proficiency. There are some flaws and limits in the statistical and other measurement methods applied in the reviewed articles, relative to item analysis, the area of interest for this study.

2.2.4 What these Studies Highlighted

These articles highlight that a certified test score loses value if the corroborating evidence sources do not facilitate test score interpretation. Or yield a more appropriate understanding of the score for interpretation (Kane, 2006; 2013). The Shinnick and Woo (2018) novice to expert group study introduced the possibility of capturing competence levels in healthcare education research. Similarly, possibilities are alluded to in scrutinising the levels of reflection in students assessed work using Kimber et al., (2008) and Kolb's (1984) reflective frameworks in the McKay and Dunn (2015) and Cathro et al., (2017) studies. The sampled studies highlighted that any scheme or framework used to measure students' assessed work for reflective content is likely to be imperfect. Gulwaldi (2009) came closest to matching students' responses to the specific criterion of all the papers reviewed. Consistent with this review's overall aim, to identify methods to measure the student's reflection and reflective ability level.

Equally, in the OSCE assessment, the Wallenstein and Abner (2014) study expressed validity aims concerning the test instrument, and Radziej et al., (2017) the test items' capacity to discriminate among students' ability. Yeates et al., focused on the impact of examiner cohorts on students' scores, examining the influence of multiple variants simultaneously. In the study by Judd et al., (2016), student performance measurements in the OSCE on professional competence and Walsh et al., (2010) clinical competency level in the academic theory contexts of interest were promising. Again, the majority reviewed OSCE studies limitations were the statistical measurement method, or the focus on measuring the device, i.e., the OSCE or the procedure relating to the clinical practice skill's. Rather than the targeted theory and level achievement and how it referred to the score.

Moreover, the statistical measurements are conventional CTT driven approaches designed to estimate measurement error and reliability as they apply to the entire OSCE instrument. Traditional statistics are not intended to measure or consider how test items align with an underlying construct (theory) or student ability (Anastasi and Urbina, 1999; Embretson and Reise, 2009). Both Judd et al., (2016) and Walsh et al., (2010) cited alternative psychometric modelling as their data analysis method. In the literature, psychometrics converses as a modern test theory for assessing achievement, used in mastery learning in exams as is the case for its use in the Judd et al., (2016) study. Alternatively, as applied by Walsh et al., (2010), psychometric statistics have utility in documenting that individuals can perform a specific skill, as well as the measurement of ability.

2.2.5 How the Review Helped to Refine the Methodology for this Research Study.

The review draws attention to the (in)consistencies across models, statistics and measurement, suggesting that no single model provides the correct answer to the research question posed. Further, standard measurements effectively control irrelevant variability. Still, they can also restrict a range of observation and potentially relevant measurements to score interpretation and use (Kane, 2006). This study's challenge was to connect the available measurement methods by laying out a theory-driven analysis network leading the inferences about the observed student performance/ achievement. To provide appropriate IVA evidence to conclude the decisions about the test scores based on that performance.

Chapter 3 documents the methodology. However, the overall methodological aims are to address the research question

> How well do the Assessment Processes used in the Non-Medical Prescribing (NMP) Course Validate Certification/ Licensing of Students to Prescribe in Practice?'

Moreover, the measurement methods aim to achieve IVA backing for the following NMP assumptions:

(1) That the knowledge and cognition required to achieve NMP proficiency are demonstrated successfully in the WRs

(2) The student has demonstrated the application of the underpinning teaching and learning theory in the WRs in keeping with the academic expectations and is consistent with the assessed scores

(3) The SCE test items are at the appropriate level, are related to the theory and are construct representative of the NMP criteria for Proficiency, as required by the NMC

(4) The SCE test items' quality provides stable estimates of test-takers cognitive ability and performance

In addressing points *(1) to (2),* the measurement technique must match the learning (theory) target to provide information that explains whether the students have achieved the learning outcome and how this relates to the score. The assessed work analyses must also be mapped carefully and considered against appropriate teaching and learning theory. The subject/domain must provide clear evidence on which to base score interpretation. In doing so, it may ensure standards of achievement and less subjectively measured levels of performance.

For points *(3) to (4),* criterion analysis measurement must be a substantial component of this research project. The focus must be on the precision of measurement, which provides an estimation of the test item, the examinee ability, and how these characteristics interact in the test score context. I.e., can proficiency in the NMP subject be demonstrated as achieved via the SCE criterion?

CHAPTER 3: THE RESEARCH METHODOLOGY, DESIGN AND METHODS

3.0 Chapter Overview

Chapter 2; presented a 'narrative' analysis of the research and statistical methods used to evaluate test measurement precision and reliability. Section 2.3 offers the reviewed papers' methodological limitations concerning measurements and analysis. Chapter 3 contains the justification for the research design, methodology and methods, including the ethical considerations, qualitative and quantitative data collection and analyses. Section 3.2.1 outlines the approach for Theory-Driven Thematic Analysis (TDTA) and the function of Ryle, Mezirow, Benner, and Bloom in the TDTA analysis. RQDA software enabled the coding and text mining of the content of sampled N=28 WRs. Also discussed is the implication of modern test theory (MTT) for SCEs data analysis, the purpose and application of Rasch analysis and item response theory (IRT). This chapter highlights the *ltm* package in R, giving illustrative examples of the command programming language used for the Rasch analysis calculations.

3.1 Reflexivity

Reflexivity, shaped by the researcher's values, experience, theories and understanding, is an influential component of the research design and process. Reflexivity, in turn, influence the research topic, ethics, data collection and analysis, affecting the study (Denscombe, 2014). Chapter 1, section 1.2, discussed the research topic and question, followed by the researcher's values (orientation) and the significance of the research problem. Reflexivity suggests deliberate acknowledgement of the researcher's influence on the participants, including the risks and potential for the researcher's tacit knowledge to impact the data, make false assumptions or miss important information (Berger, 2015). Section 3.1.2 to 3.1.4 sampling, ethics, and recruitment discusses the activities and tools used to mitigate researcher influence on the participants. This section includes the justification for the choice, use and rationale in the case study design. Sections 3.1.5 to 3.4 reflexively outline the researcher's aims in the data collection and analysis methods to reduce researcher bias, subjectivity and improve fitness-for-purpose (Berger, 2015).

Bracketing is a method used for mitigating the researcher's bias or equivocal evidence influencing the study (Tufford and Newman, 2012). Research praxis safeguards that the researcher does not impose their meaning on the data during analysis (Berger, 2015). The researcher's preconceptions, beliefs, and assumptions that informed the methodological choices for investigating the phenomenon of interest are acknowledged in chapter 1, section 1.2.3. Research praxis addressed in section 3.1.2 presents Yin's (2003, 2014) logic model, which in this study helped mitigate some research praxis aspects through the multiple theoretical frameworks defined. However, there are remaining limitations concerning other bracketing types, e.g., the abstract formulation and reintegration, which may enhance secondary data use. See the limitations section 6.4.2.3.

3.1.1 Research Design

The research design is a case study. The rationale for this choice is the framework affords an in-depth exploration of the NMP programme with the NMP assessment as the focus. In such circumstances, Yin (2003) suggests that a single case study is an appropriate design based on its ability to discover an existing theory's fundamental phenomenon or analytical testing. In this study, the 'case' is a single student cohort. The embedded case study was deemed appropriate for this study on the following grounds: Each NMP assessment acts as a unit, for example, the SCE test, which contained other subunits as test item criteria which required a quantitative psychometric method of analysis (Murphy and Davidshofer, 2005; Schultz et al., 2005). The 'embedded' cases, the WR and SCE, offer contrasting information about the NMP assessed work and the individual participant. The WRs as a unit contain theories as subunits that state the conditions under which a phenomenon is analysed and measured (Yin, 2003, 2014). Similarly, the SCE criterion analysis facilitates the examination of the same participant's test achievement, qualitatively and quantitatively, several times on several theoretical and test occasions (Yin, 2014). In this way, multiple sources of test data are presented to meet Kane's IVA requirement. See section 3.4 in this chapter.

3.1.2 Establishing Rigour in the Case Study Design

According to the literature, the single case study design occupies a vexed position in the research discipline. Seemingly held in low regard, the case study is viewed as methodology flawed due to a perceived lack of design rigour compared to the systematic scientific research method (Gerring, 2004). Case study design attends rigour in its logic model, which informs the case study methodology (Gibbert et al., 2008; Yin, 2014). The model helps define the theories underpinning the case (NMP assessment) and the frameworks for analysing the data to represent the links between the research question and processes applied in solving the problem (Yin, 2003, 2014). The model contains the following steps associated with rigour: (1) Propositions, (2) the conceptual framework, (3) the data link to the propositions and (4) the criteria for interpreting the findings (Yin, 2014, p.33).

(1) Propositions direct the study's scope and what is examined. The first proposition concerned students' performance against the teaching and learning theories in the NMP assessment (Yin, 2014). This proposition informed the data collection, i.e., the students' assessed work. Secondly, the propositions are derived from the theory cited in chapter 1 (Yin, 2014). These theories provided descriptors for data coding and a distinct focus on explaining NMP students' assessed work (see section 3.2.1). The logic model and the theory triangulation helped reduced researcher's bias in analysing their achievements (Carter et al., 2014.; Tappen, 2016). (2) The conceptual framework is Kane's (2006) interpretive validity argument (IVA). See section 3.4. The research's primary focus is assessment validity and is the theory underlying the research questions presented in chapter 1. Kanes' IVA furnished the rationale for the validity evidence and introduced some objectivity in discussing the results (Yin, 2014). (3) Linking of data (NMP assessed work), the propositions (theories) and Kane's (2006) IVA as the conceptual framework. Here the logic modelled a strategy for the qualitative and quantitative data analysis choices and techniques.

The design included Modern Test Theory (MTT) and its various tools for measuring the NMP concepts of interest requiring different validity measurements. (See chapter 1). Finally, (4) the propositions criteria for interpreting results, ensure data analysis alignment to theory, is operationally delivered and applied in the discussion (Yin, 2014). In this study, the propositions are relational. The relational logic states that if teaching,

theory, and learning are present (an input concept). Then proficiency (the outcome concept) should be evident in the NMP student's assessed work (Avan and White, 2001; Yin, 2014). Yin's (2014) logic model was crucial as it informed the methodology and methods for data collection, measurement and analysis (Duff, 2012a, Yin, 2014).

3.1.2.1 Research Methodology and Methods

3.1.2.2 Sampling

The sample comprised the assessed work from a cohort of 52 NMP students' in the given year between 31st July 2012 and 31st August 2013. These students consented to their coursework inclusion in this study. The sampling method was convenience (non-probability), called volunteer sampling. Individuals actively agreed to participate in the study. The students' (volunteer) sampling method was the default. Students enrolled in the NMP course were invited to participate in this study (Teddlie and Tashakkori, 2003). Convenience sampling, considered a weak method is subject to self–selection bias. However, this method was deemed appropriate as the study's focus is the internal validity of the NMP assessments. The results are not intended to be externally valid (generalised) to individuals other than those who participated in the study (Polit and Tatano-Beck, 2009). The sampled assessed WR and SCE work included in the study were matched by student ID to maintain student consistency.

3.1.3 Ethics

The HEI Ethics Committees at the author's place of study and employment approved the research, the documentation, present in appendix 1. Ethically, the research is neither dangerous nor intrusive to participants. There are likely threats of coercion. The researcher is an employee in the faculty where the study took place. Therefore, the researcher's presence can be influential to the student's decision to participate. Further, coercion due to student vulnerability, this cohort may wish to enrol in future health education programmes where an encounter with the researcher is possible. At this point, the researcher did not have current or future teaching responsibility or administrative contact with the NMP students' during recruitment to reduce the potential coercion threat (Research Governance, DH 2005a).

Other threats to ethics included accidental disclosure of results to others not involved in the study, minimised by the students' unique identifier (SId). The SId helped eliminate demographic data use to access the NMP assessments, thereby maintaining student anonymity (Caldicott, 2010). As an employee of the university teaching staff in the Faculty of Health and Life Sciences (FHLS), the researcher had access to all students' assessments via the faculty administration office. Data storage was available in locked metal filing cabinets in a locked storage room in the administration office. During data analysis, computer password protection and encryption-maintained test data safety during storage (Data Protection Act, 1988; Caldicott, 2010; General Data Protection Regulation (GDPR, 2016).

3.1.4 The Recruitment Process

The recruitment process used intermediaries to limit the researchers' influence on the student participants as much as possible to ensure the recruitment process was ethically (Featherstone, 2006). The module leader and course team acted as intermediates, distributing the consent forms (see Appendix 2) and study information (see Appendix 3). The intermediary's role was study information distribution only. Students were at liberty to consent (or not) to researcher access their WRs and SCEs as data for research purposes. A self-addressed envelope was also issued with the study information to return the signed consent forms via the faculty administration office. The researcher received the consent forms and not demographic information, e.g., participant name (DH, 2005a; Data Protection Regulation (GDPR, 2018). The inclusion criteria are students who have completed the NMP course, those willing to consent. The rationale for this is that the study's interest lies in those who pass the NMP course. Age, gender or levels of fitness limitations were not applied. Students' were excluded from the study where there was notification and information from the faculty administration office that a student has 'failed' the assessments or students' unwilling to continue after consent.

3.1.5 Data Collection

The rationale for using secondary data to address the research questions was its economic and pragmatic advantages and the data sources' characteristics, including the data's breadth and depth, reliability, and validity. The data are the NMP students assessed summative work. Regulations govern the conduct, standards of assessment and marking practices, which impact the data quality (QAA, 2011, 2014). The post ethical approval data harvesting revealed that the multiple-choice questions (MCQs) contained too few item sets (<n= 20) to yield meaningful data, statistical analysis and stability (Downing and Haladyna, 2006; Bond and Fox, 2015). The MCQs were, therefore, excluded from the data collection.

The assessments serve an essential purpose in capturing authenticated information for determining the validity of the assessments used in the NMP program. Secondary data sources have several advantages; (a) the data are actual NMP tests specific to the research purpose, (b) it is time-efficient, making use already collected of test data and (c) it allows considerable extension of the scope of the study regarding analysis (Smith, 2008). However, there are also disadvantages (a) the researchers have to utilise the original data collected and collection tools. (b) excludes questions critical to the study in the data, and (c) conditions and data quality are traded for control and accessibility (Vartinian, 2011). Each of these can affect the analysis and conclusions (see chapter 6, limitations).

The data collection consisted of two types of assessed work. Firstly, the summative written reflections (WR), students' reflective narratives about prescribing experience, professional conduct and proficiency (NMC, 2006, 2018). The reflective essay aims to facilitate students demonstration of the range of knowledge gained. Students should show safe, effective prescribing based on reflection of an episode of their clinical practice experience (See Appendix 4). Secondly, the SCE test is an authentic NMP simulation designed to test student prescribing practice. Unlike the Objective Structured Clinical Examination (OSCE), the SCE is a single station designed by academic and clinical NMP practitioners (see Appendix 5). SCEs content and test item criteria are blueprinted, charted against legislated NMC mandates and NMP knowledge domains (Hardern et al., 2016). The NMP SCE is assessed as a coherent, meaningful whole task. Students' consult with and examine a patient based on the presenting complaint to find

the diagnosis and prescribe appropriate treatment. The SCE aims to assess a range of knowledge, skills and performance (KSP) criteria (Hardern et al., 2016). Two examiners observe the candidate without interruption; KPS are captured through follow-up questions at the end of the assessment.

The student's assessed work is primary data used for secondary analysis. There is a reasonable degree of confidence in data generation's integrity and confidence in the data's reliability in how it was collected for the following reasons. The assessment process is regulated, and there is an element of fitness for purpose. The NMP assessments were designed to capture information that accurately exhibits a candidate's ability (Cohen et al., 2017). There is strong evidence that the data collected can answer the research inquiry and the study aims.

3.2. Data Analyses

3.2.1 Developing the Code for Content Analysis

For analysing the written text, a frequent method used is content analysis (CA). The research literature identifies the CA process as words, terms, paragraphs, concepts, and semantics. There are many ways of coding the written message for analysis. CA is applied to address the following:-

> Sub-question [1]. Are the NMP Summative Written Reflections (WRs) a Valid Constructive Indicator of NMP Students' Proficiency for Certification and Licensing to Practice?

There are also different code development methods. Concept (or inductive) driven, data-driven, or deductive priori -driven and theory-driven (Gibbs, 2007), specified by Boyatsiz (1998) as distinct frameworks for code development. For concept (or inductive) driven, a code is assigned as a concept becomes apparent in the text. In the data-driven or deductive a-priori research-driven method, the researcher uses someone else's code and framework to process or analyse the information. The researcher accepts the code originators presumptions, biases and projections (Boyatsiz, 1998 p.37). A theory-driven approach generates codes deductively from theory. The former two concepts and data-

driven TA are susceptible to coding projection obstacles, where the researcher reads into or attributes something of their characteristics into the text.

The deductive theory-driven thematic analysis (TDTA) applied in this study developed the coding and created approximately 120 codes under the nine NMP themes. The TDTA generated codes from Ryle (1949), Mezirow (1981), Benner (2001), and Bloom's (1956) theories, as identified in chapter 1, increased coding quality (see appendix 6a to 6d). They are significant to the content analysis as they provide the categories for the investigation's focus (see chapter 1: 1.4). The deductive TDTA approach used criterion-referenced material anchored to prescribing activities, i.e. the NMC (2006, 2018) Standards of Proficiency document for independent prescribing practice. The TDTA helped reduce the researcher's projection bias and established some consistency in judging the observations made in the assessed work (Boyatsiz, 1998).

3.2.2 Cohen's Kappa Statistic for Inter-Rater Reliability

In the absence of prior studies using similar coding tools, reliability needs establishing using statistical analysis (Polit and Tatano- Beck, 2009). The coding may not produce objective, reliable judgements about the WRs content. As the sample is known, an unweighted Cohen's Kappa measured the Interrater reliability (IRR) agreement. The measurement estimates the TDTA coding quality by the raters' ability to independently apply the codes to the same set of subjects' assessed work and achieve consistency (Heiman, 2001; Bakeman and Quera, 2011).

Cohen's Kappa is a summary statistic that assesses the average agreement between two coders, applying coding to text. Cohen's Kappa (K) is recommended over a simple percentage agreement because it calculates the agreement between two raters. Then corrects the probability that raters will agree by chance (Salkind, 2017). Table 3.2.2 lists the values for agreement between independent raters, occurring beyond chance, expressed mathematically as a score ranging from 0 to .20 for less than a chance agreement to 1.0, indicating perfect agreement (Tappen, 2016).

Table 3.2.2 Lists Cohen's Kappa (ϰ) Interpretation values

Kappa Interpretation					
None	Minimal	Weak	Moderate	Strong	Almost Perfect
0-.20	.21-.39	.40-.59	.60-.79	.80-.90	Above.90
% of Data that are Reliable					
0-4%	4-15%	15-35%	35-63%	64-81%	82-100%

McHugh (2012)

Table 3.2.2 Provides a representation of Kappa values, the Table includes benchmark guidelines for values, none to perfect. The percentage (%) calculate agreement as indicators of the quality of the TDTA coding and the raters' ability to apply them.

The literature suggests that the majority agreement should be ideally > .80= strong, with a minimum for acceptance of .60-.79 = moderate, indicating some confidence in these results (Bakeman and Quera, 2011; McHugh, 2012). When a high disagreement occurs across the coding, e.g., the majority coding < .59 (weak), this result indicates investigating the cause as the codes may need revision. The coders must discuss the minor disagreements and reconcile the disagreement (Tappen, 2016). The higher the Kappa agreement (>.80), the higher the agreement and rater accuracy. This study accepts the Kappa value of >.80 as a good indicator of rater agreement (Bakeman and Quera, 2011).

Some factors can affect the Kappa value, including the number of codes in the data set. For example, Kappa's responds to large numbers of codes (> 30) by increasing the kappa value, and the reverse is true for low code numbers (Bakeman and Quera, 2011). However, Bakeman and Quera state that even when Kappa is not perfect in practice, the computed rates are reliable estimates of Kappa values (2011, p 76). See chapter 8 for Kappa results.

3.2.3 Content Analysis (CA) of the Written Reflections

Within the Content Analysis (CA), the researcher explores WRs in an area not understood or extensively examined in the literature. The TDTA units of coding derived from NMC (2018) standards, teaching and learning theory give the theoretical

justification to the analysis and help establish a codable moment in the text (Boyatsiz, 1998). Theory triangulation, i.e., multiple rather than single theory approaches, examines the WR assessed work. To allow a theory-informed analysis account of NMP practice (Carter et al., 2014; Yin, 2014). CA requires both inductive (the researcher immerses themselves in the data) and deductive (code application) approaches to analysing relationships between the theory and the text message (Boyatsiz, 1998; Gibbs, 2007). For example, what themes and concepts in the WR text count as relevant as the cited theory specifies.

There is an acknowledgement that CA rely on subjective judgments, and there is a natural tendency to confirm proposed interpretations (Gibbs, 2007). However, the theories introduced in section 1.4 are a counterbalance to analyses of the WR content (Mayring, 2014). Bias theory selection is also possible. Nevertheless, the objective was selecting the best option, in this study, the teaching and learning theories in nursing and education (Quinn and Hughes, 2013; Wheelahan, 2010; Oermann and Gaberson, 2014). Students select the NMP scenario for reflection. Therefore, they are responsible for demonstrating the required range of knowledge gained underpin safe, effective prescribing practice (see Appendix 6 for assignment brief). The WR text mining and labelling occurred for all codes as they appear in the WR text (see Appendix 6a). Codes repeatedly found across the n=28 WRs are analysed and discussed under umbrella themes–Theme 01 Consultation, Theme 02 Patient Safety and Theme 03 Communication, cited in the literature as significant to safe NMP practice (NHS, 2010; Morgan et al., 2014; Silverston, 2013).

3.2.4 Text Mining using R-QDA

R is an open-source programme and the software for Qualitative Data Analysis (RQDA) and used for code generation of the theorists and models identified in sections 1.4.0 to 1.4.3 for data mining of the WR text. The TDTA codes for these identified theorists are n Appendices 6b to 6d, respectively. The researcher scanned the students' 2500-word summative WRs in their entirety to generate qualitative data. Each WR was numerically labelled to maintain participant confidentiality and anonymity. The data were exported as a text file into the computer program R v3.2.3, the Qualitative Data Analysis [RQDA] package for coding. The text mining was performed on code completion and comprised

of seeking out content and assigning the codes to any variable of interest for content analyses. The RQDA software applied labels the WR text for the TDTA themes and codes. The WR content was examined line by line for text representative of the specified codes. The analyses aimed to determine the overall content validity of the WRs. What is captured and measured regarding levels of knowledge, subject proficiency (or course level mastery, see chapter 1 section 1.4) to establish the assessment's validity (Messick, 1995). Codes not mined in the WR text are discussed in the analysis related to the prescribing practice deficits.

3.3 The Structured Clinical Examinations (SCE)

The NMP program uses simulation to assess the students' prescribing achievement against the NMC (2006; 2018) standards for prescribing. The Structured Clinical Examination (SCE) is designed to match job proficiency and authentically represent the clinical prescribing problem (Pugh and Smee, 2013; Hardern et al., 2016). The underlying assumptions are that SCE is context-rich and accurately assess traits relevant to professional NMP practice. Validity here requires that the SCE are context-rich and (a) contain some related theory and subject domain constructs simultaneously (b) cover the required MHRA (2005), the NMC (2006; 2018) content and (c) assess the desired NMP standards and student traits. The psychometric analysis is pivotal to answering sub-question 2 concerning the SCE test's validity.

> Sub-question [2]: To What Degree are the NMP Structured Clinical Examinations (SCEs) a Valid, Accurate Representation of the NMP Criteria they are Designed to Measure?

3.3.1 The Rasch Model for Analysis: The Purpose and Function

Test item criteria include background questions to assess cognitive skills such as problem-solving and complete tasks to meet the NMP goals successfully. Criterion Validity (CV) measurements indicate test function and examinees' ability in specified domain areas represented by the test item criteria (Lin, 1980). To establish validity, the researcher applied psychometric statistics to the SCE test criterion. Test performance measurement in psychometrics determines the following: (a) whether the SCE test

criterion has enough difficulty to validly test students' ability in NMP knowledge, skills acquisition, and proficiency. That (b) the SCE test criterion can discriminate between students with high, medium and low NMP ability. Indirect measurement of the Latent Trait (LA) or person-ability is estimated using Item Analysis (IA) from the examinees' response. Therefore, IA provides a close approximation of test score meaning (Anastasi and Urbina, 1997; Embretson and Reise, 2009). This study acknowledges that it is not possible to establish test validity undeniably. However, Kane's IVA may reveal that it is at least plausible using empirical measurement (Kane, 1992: 527). The SCE test hypotheses (a) validity and (b) relationships between test item function. For example, difficulty and person-ability will be analysed using the Rasch Model, which will also investigate test items at the test criterion level (Embretson and Reise, 2009).

The Rasch model consists of a series for exact measurement to validate the clinical and theory characteristic of the SCE assessment (Streiner, 2010). IA has two prominent statistical measurements applied in this study. Item difficulty (DIF) and item discrimination (ID). Item difficulty (DIF) is as the term implies how difficult the test items are (Downing and Haladyna, 2006; Bond and Fox 2015). In test design, it is appropriate to ensure that tests are not so hard that students fail to achieve. Conversely, a too easy test fails to provide information about the test taker (Downing and Haladyna, 2006). In contrast, ID quantifies how well the test item discriminates between examinees' with high subject-domain ability and those who are not. Further, IA facilitated the investigation of the test items' performance against external criteria- i.e., the theory presented in 1.4.2 and 1.4.3. IA thus provides counterclaims about the SCE test's (proficiency) assumptions and information for addressing this study's validity question under Kane's (2006) IVA.

3.3.1.1 Interpreting the Test-Item Characteristics: The Item Characteristic Curves (ICC)

The logit (logistic) scale is the mathematical function in Rasch. Expressed as a probability (log-odds), the measurement locates the examinee's ability (latent trait (LA) and variable interest in the test item along the same logit (log-odds) scale. Latent Trait Analysis (LTA) models routinely set ability (or theta) at zero on the logit scale, which equates with the average 0.5 (50%) probability of a person's success on a test item. The

ICC mathematical form illustrated in Fig 3.3.1.1 is logistics. The logit scale can estimate ability from -∞ to +∞ (infinity) on the logit continuum.

Fig 3.3.1.1 The ICC Exemplar for a Single Test Item (jMetrick Software, 2014).

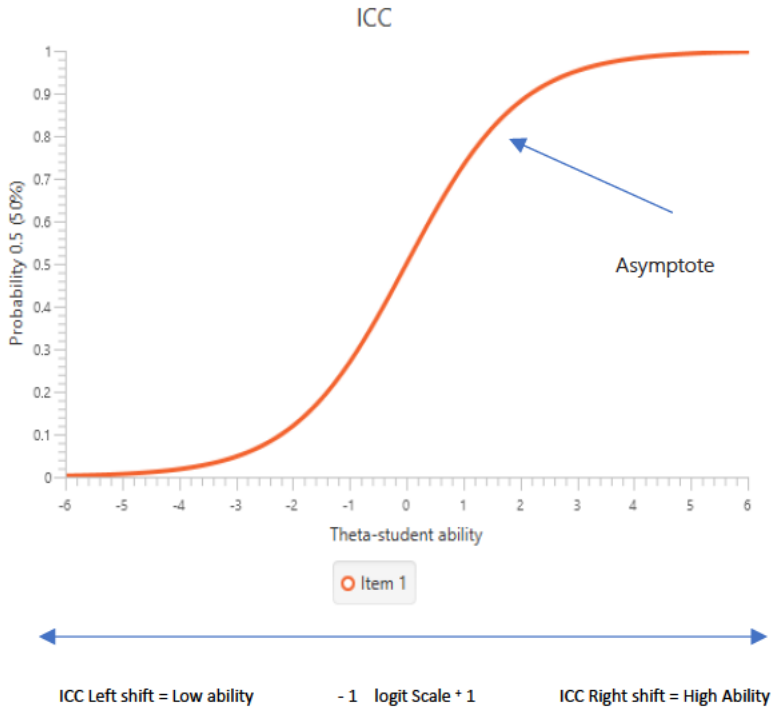


Fig 3.3.1.1 Illustrates the ICC for a single test item: Moving to the right along the logit scale; the higher the examinee ability measured at 0+ on the X-axis, the higher the probability the examinees' will answer the difficult test item correctly.

Lower than negative logits indicate lower than average ability, and higher than average ability is the opposite. The ICC graph's logit formation is an S-shaped curve: Pictorially describing the location of person-ability and item parameter, i.e., the DIF or other variables of interest, at the curve's inflexion. The probability of a correct and incorrect response to the specified test item on the same scale, represented by the uppermost (asymptote) arch of the ICC. Each test item has an ICC, and depending on the LTA model, it will portray different parameters of interest. For example, the items DIF, discrimination (ID) or guessing parameter, along with the examinee ability, are presented on the same logit metric (Embretson and Reise, 2009; Glynn, 2012).

3.3.1.2 The Test-Item Difficulty (DIF) Values

In CTT and IRT, express test item DIF as the likelihood of a correct response. Item difficulty will typically have the following ranges; very easy, easy, medium, difficult and very difficult. The item difficulty is an estimated measure of person-ability. The DIF calculations are the percentage of examinees' with a given ability (trait, e.g., proficiency) in a test who correctly answer the test item (Bond and Fox, 2015). Rasch uses the standard error of measurements (SEM) to define the interval around the estimated measure of difficulty within which it is expected to fall. The empirical rule is that 68% of the sample mean should fall within one SEM of the true mean, which in Rasch modelling, 68% of the time is unknown. Rasch systematically sets the SEM at zero. Therefore, an individual with an ability at zero (logit) will have a fifty-fifty probability of correctly answering the test item. SEM is a correction for potential inaccuracy in estimating exactness and precision in measuring parameters of interest (Hambleton and Cook, 1977). SEM is not extensively reported in this chapter. However, item fit statistics were applied to select the correct Rasch model to address SEM (see results chapter 5).

Difficult test items require a reasonably high person-ability for the individual to answer the test item correctly. Fig 3.3.1.2 illustrates the position of the DIF parameter in a single test item. An examinee with a low capability of -1 has a lower probability of a correct response to a test item (Meyer, 2014). For the examinee whose ability exceeds item difficulty (> 1), the ICC shifts to the right of zero (anywhere between 0.0 too ∞) on the scale. Meaning the person is more likely to pass than fail the test item, i.e., the test item presents no difficulty to the individual. If the test items are too simple, offering no challenge to the individual, the results are negative. The ICC shifts to the left of the logit matric; in difficult test items, the opposite occurs (Anastasi and Urbina, 1997; Embretson and Reise, 2009; DeMars, 2010).

Fig. 3.3.1.2 The Example ICC for a Single Test Item, Illustrating Test Item Difficulty (jMetrick: Software, 2014).

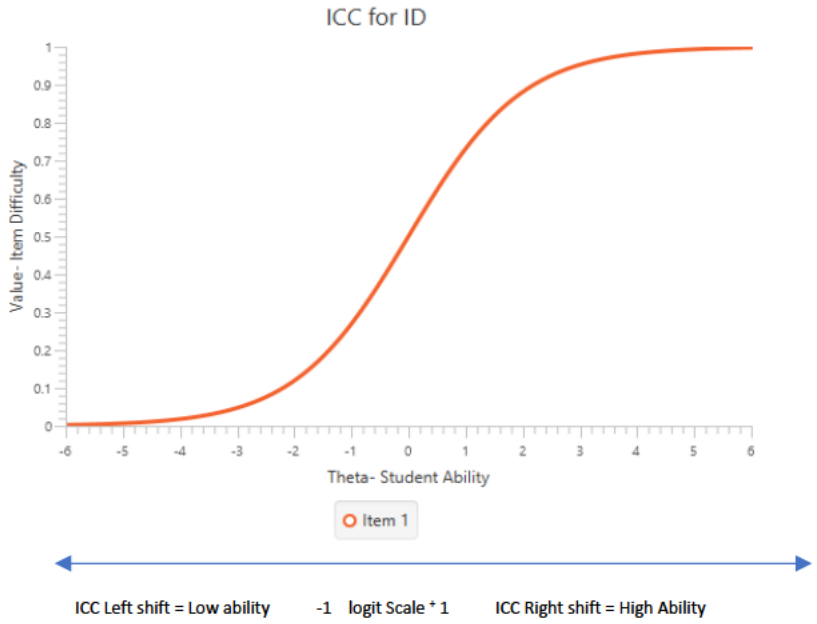


Fig. 3.3.1.2. This curve is representative of a medium difficulty test item which discriminates well. The Item difficulty is represented by the slope of the curve, and the height of the curve on the Y-axis increases with the easy items and shifts to the left of zero. In difficult test items, there is a decrease in the height of the curve and a marked shift to the right of zero the lowest point of the curve as it tails along the X-axis

As Fig 3.3.1.2 illustrates, a single ICC represents one test item. In the presence of the DIF parameter, the ICC's uppermost curve is situated anywhere right of zero logits, with its deflection slope in line with the ability (or theta ɵ) at zero on the X- axis on the logit continuum. The S-shaped curve's inflexion and tail correspondingly shift to the right on the logit ability continuum, acting as an examinee ability locator, shifting as and when examinees' have a higher (or lesser) ability. The ICC's upper curve position on the Y-axis signifies the probability of examinees' responding correctly to the test item (Baker and Kim, 2004). When the item DIF is mid-point on the scale (at zero), the probability of a correct response to a test item for most persons is 0.5 or 50%, i.e. the cut score (Baker, 2001; DeMars, 2010). The test item difficulty should either match the examinee's ability or be significantly above the cut score.

3.3.1.3 The Test-Item Information Function (IIF) Values

The Item Information Function (IIF) is a precision or reliability measurement for standard errors (SE). The IIF also has its characteristic curve, but unlike CTT, single SE measurements and reliability apply to the entire instrument represented by the test score. In IRT, the SE estimates each test item and each examinee location (ability) on the logit scale. The SEs create confidence in the ability estimates the same way that traditional SE measurement builds confidence in the CTT scores (DeMars, 2010). IIFs vary across the ability levels and produce maximum information for examinees' located at the item's DIF level. The height of IIF, where it corresponds with the item's difficulty value along the X-axis, whether positioned at left or right of zero logits, is the precise IIF measurement (Baker et al., 2004; Doucette and Wolf, 2009).

As Fig 3.3.1.3 illustrates, the ideal IIF is bell-shaped in a broader range of tests-items. The bell shape indicates the test effectively measures ability levels over a range of D-level test items, representing ability values between -3 and +3 logits. In such cases, the measurement properties (of ability and DIF) and the amount of information collated from the individual items by the LTA model are at maximum precision at the height of the IIF. However, short wide IIFs overall have lower precision values due to their broader range of measurement. Where this is not found to be the case, for example, if the test items are difficult. The IIF will peak on the Y-axis at maximum precision at the point where most of the test items are in that region, usually where the examinee ability is located along the X-axis (DeMars, 2010; Meyer, 2014).

Fig 3.3.1.3. The Example ICC for a Single Test Item, Illustrating the Item Information Function (IIF) (jMetrick Software, 2014)

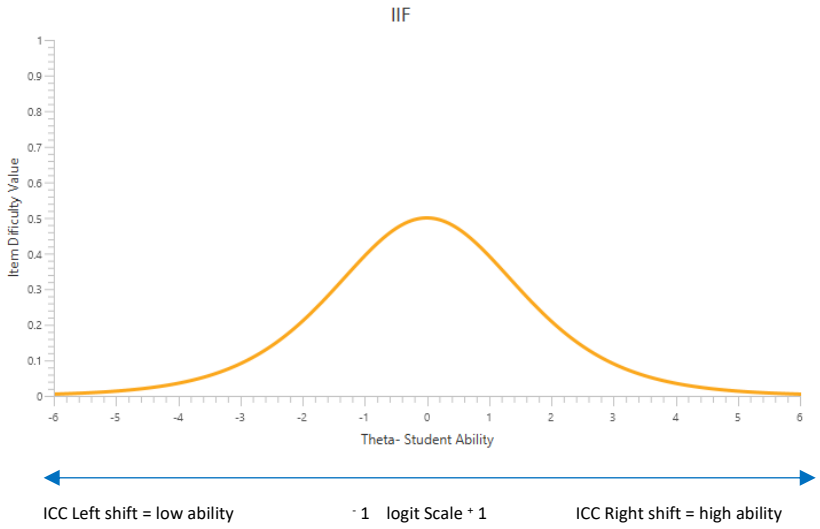


Fig 3.3.1.3. The height of IIF, where it corresponds with the item's difficulty value along the X-axis when left or right of zero (theta) ability along the Y-axis logit continuum, is where precise IIF measurement are located (Baker et al., 2004; Doucette and Wolf, 2009).

The illustration in Fig 3.3.1.3 depicts that when IIF is functioning as intended. The IIF will have several graphical representations of single test items, have high precision measurement within a reasonably confined locus on the difficult [+1 to ∞] to the right of ability. Alternatively, too easy [-1 to ∞] to the left of ability located at zero logits on the scale. The IIF decreases as the person-ability levels move away from the DIF parameter, indicated by a peaked shaped IIF. The IIF is considered precise in a specific area along the logit continuum. IIF can give highly desirable results dependent on the test information's intended use (Baker, 2001; Doucette and Wolf, 2009; Embretson and Reise, 2009).

3.3.1.4 The Test Information Function (TIF) Values

The Test Information Function (TIF) is a global evaluation of the tests' precision, at examinee ability for different test items. The TIF is the totality of all the Item Information Function (IIF) in the test, using the item parameter estimates from the original test datum. TIF also gives the same test characteristics information to describe the entire

test's measurement of examinee ability. The TIF, indicated by the red vertical slope in Fig 3.3.1.4, shows how well (precisely) the test performs its measurement (Hambleton and Cook, 1977). The singularly arranged test item parameters (i.e., DIF) and IIF are plotted from difficult to secure. Unlike IIF, which graphically represents multiple test items on a single ICC, a single line represents the TIFs global test function representation. Affiliated to the IIF, the TIF curve's height also peaks on the Y-axis in the region where the majority person ability resides. As seen in Fig.3.3.1.4, where it corresponds with examinee ability [theta, θ] along the X-axis, the TIF precision estimate is located at the curves maximum height.

Fig 3.3.1.4 The Example ICC for a Single Test Item, the ICC, the Item Information Function (IIF) and Test Information (TIF) (jMetrick: Software, 2014).

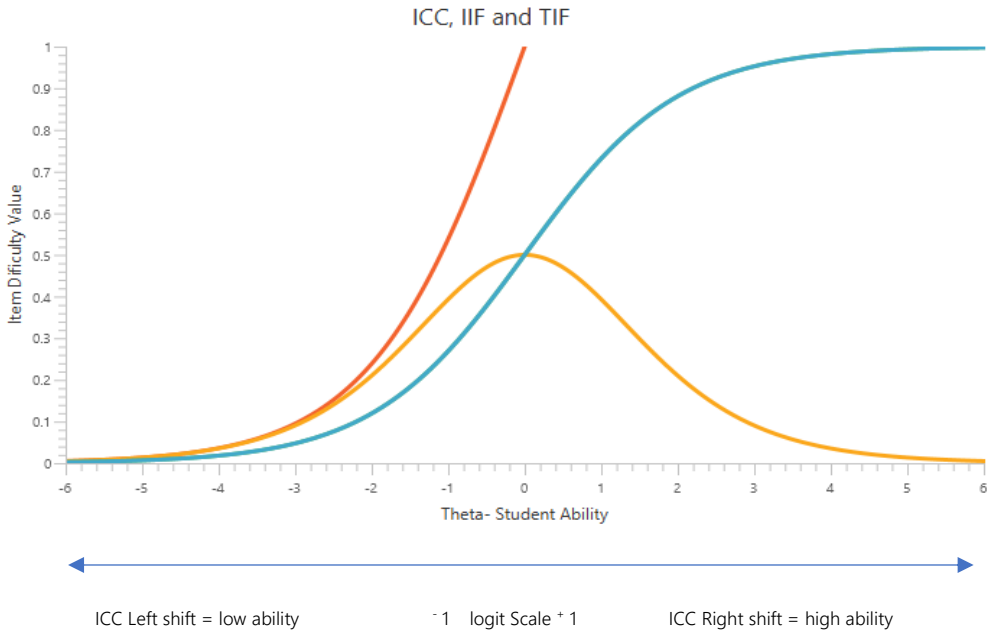


Fig 3.3.1.4 the precision estimate in this figure is located at the TIF (the red line) at 0.04 on the ID values scale on the Y-axis. The measurement is taken where it corresponds with the position of examinee ability at theta (ɵ); at zero along the X-axis.

As illustrated in Fig 3.3.1.4, the TIF and the examinee's ability are reported on the same scale. TIF gives a more worthwhile test validity measurement, providing a (logit) match between the examinee's ability and the test item criterion. Suppose the TIF in Fig 3.3.1.4

curve is high (peaked) indicated by the red line, as in this illustration. In that case, the conclusion is that the ability levels against the items DIF have a reasonable precision estimation. Like ICC and IIF, the parameter estimates in TIF for high ability examinees' shifts the graph to the right of the zero logits point on the ability scale. In the easy test, there is a shift to the left of the zero logits on the ability scale (DeMars, 2010; Bond and Fox, 2015).

3.3.1.5 The Implication of Rasch Modelling in Modern Test Theory (MTT)

The implication of MTT to the analysis methodology in this study is that closer attention is paid to Rasch Modelling, more specifically Item Analysis (IA). IA's impact is the capacity to determine test function and examinee attributes by applying other IRT models (Hambleton et al., 1991; Anastasi and Urbina, 1997; Embretson and Reise, 2009). MTT depicts a more precise relationship between the score and the variables of interest for this study. In MTT, the measurement strategies focus on test-item and examinee ability. Therefore, the test score interpretations are more closely allied to CTT's observed score [X_0] and exact score [X_T] (Shultz et al., 2005).

3.3.1.6 Quantitative Data Analysis using R-itm

The Rasch model (RM) is the most widely used for item analysis (IA). A family of RMs or parameter logistic (PL) models exist to mathematically analyse items to describe the relationship between the person's ability and the test items difficulty or discrimination (Streiner, 2010). The 1PL is the original RM, and the 2PL and 3PL are extensions of 1PL (Embretson and Reise, 2009; Streiner, 2010). The 2PL has enhanced precision on the DIF parameter, and 3PL has an enhanced correction factor for guessing in students with a low ability (Baker and Kim, 2004; Streiner, 2010). The goodness-of-fit calculation describes how well the data fits the RM determined the 1PL suitability for data analysis. See results chapter 5. IA measurement was conducted using the Rasch 1PL model contained in the *ltm* package in the R programming environment. The *ltm* package includes a range of calculations capabilities. The principle of R is to apply commands rather than a graphical user interface. The commands are generally made up of two parts: Objects and functions typed into the R console. Functions create objects.

The objects created contain the analysis outputs, meaning that the objects can have single values, numeric or character (variables). To instruct R to treat data as a numerical object, ""is used (Field et al., 2012). The command "c( )" concentrated the object which contains data variables. An example is, in entering the SCE data onto excel once saved as a CSV file, to export the data into R and save the file, the command took this general form:-

Write.csv (dataframe, "filename.csv")

When multiple variables are used in creating an object, the function command "c( )" is required for the data entry into R. Where multiple commands and stringed variables are needed, '<' is used to instruct 'create from'. For example, stringed variables were required for finding the data file in the ***ltm***-library to create the analysis, for example:-

load(file.choose() pQdmand<-
read.csv(file.choose(),header=T) names(pQdmand)

Table 3.3.1.6 presents an excerpt of the command code for the goodness-of-fit calculation. For R-command codes for the SCE test item analysis, see Appendix 7.

Table 3.3.1.6 The Goodness-of-Fit Command for Rasch Analysis used in R-***itm***

Mandatory Questions:
m1<-rasch(Qdmand,constraint=cbind(length(Qdmand)+1,1))
m2<-rasch(Qdmand)
m3<-ltm(Qdmand~z1)
m4<-tpm(Qdmand,type="rasch",max.guessing=1)

Table 3.3.1.6 contains an excerpt of the function<object c( ) command for the goodness of fit calculation for the Rasch model selection for analysis.

Table 3. 3.1.6 presents the first few commands for the constrained Rasch analysis for the mandatory SCE items to determine 'goodness-of-fit'. The calculation identified that the 1PL Rasch model provided the best fit to the data analysis than the 2PL or 3PL models available in the ***ltm*** software package. The goodness-of-fit result is discussed in Chapter 5.

3.3.1.7 Goodness of Fit Calculation

The **ltm** package is a program within R- applied for the goodness-of-fit-test. A nonparametric bootstrap statistic determined the most appropriate model for the SCE data for the test-item fit. These results are in Chapter 5. In Latent Trait, Analysis (LTA), models identified in section 3.3.1 are used to examine the probability (*p*) of a correct response for each item against the (examinees') ability level [z]. A general model equation for this probability for the mth respondent in the ith item (question) is:

$$P(x_{im} = 1|z_m) = c_i + (1 - c_i)g\{\alpha_i(z_m - \beta_i)\} \qquad [1]$$

Where x_{im} is the dichotomous response to the ith item by the mth respondent, z_m is the respondent's level on the latent ability scale. The guessing parameter c_i represents the probability of getting the test item correct by guessing alone. The discrimination parameter α_i describes how well the test item can differentiate between high/low ability. The difficulty parameter β_i describes where the test item functions on the ability scale as a location index. The link function g is a logit function, the natural log of the odds (logits), which map probabilities on a scale (Field et al., 2013). This calculation correlates the examinee ability to the possibility of correctly responding to the test item (Rizopoulos, 2006; p.2-3). The mathematical application for the 1PL Rasch model allows the calculation of the probability of a correct response for a given latent trait (LT), and the item parameter is as follows:

$$Pr(x_{im} = 1) = \frac{e^{z_m - \beta_i}}{1 + e^{z_m - \beta_i}} \qquad [2]$$

The 1PL model allows test items to vary by how easy the test items are to answer correctly. Where Pr is the probability that the response to the test item $(x_{im} = 1)$ refers to the respondent z_m choosing the correct response on the latent ability (logit) scale on the item parameter β_i (difficulty). The item difficulty acts as the locator of a person's ability on the logit scale; the logit function is '*e'.* The SCE data has two distinct components, non-mandatory and mandatory question sets. Each LTA model 1PL, 2PL and 3PL were applied to the generic and mandatory test items for goodness-of-fit to judge dimensionality. I.e., dichotomous datum or otherwise, identifying the most effective model for explaining the SCE datum (Embretson and Reise, 2009; Bond and Fox, 2015).

For the comparison results of the different models, see chapter 5, Tables 5.1. 2 to 5.1.3.2 for the 1PL model datum.

3.4 . The Structure of the Interpretive Validity Argument (IVA)

The validity argument investigates the accuracy and plausibility of the claims and the assumption made about the student's capability against the cited theory underpinning the curriculum's teaching theory, and therefore, the assessment. Theory triangulation is consistent with Yin (2014) logic model. Theory triangulation also provides lines of evidence for Kane's (2006, 2013) IVA, see chapter 1. The evidence organises the IVA discussion into a coherent and plausible interpretation of the examinee's performance in the assessed score context. The evidence supports conclusions on whether to accept or reject the NMP validity claim.

Fig 3.4 illustrates this research's application of Kane's IVA positioned at the top level of the hierarchies. The second level includes the following. Firstly, the professional body's educational governance and the competency framework in nurse education must meet stakeholder and employee practice education requirements. Secondly, the QAA characteristics addressing the NMP scholarly demand and academic rigour in the master's programme. The third level contains the four prevailing theories used in nurse education and assessment, significant for measuring the content and criterion presence in the assessment and students' assessed responses.

Fig 3.4 The Application and Structure of Kane's IVA: The Teaching and Learning Theories used as Evidence

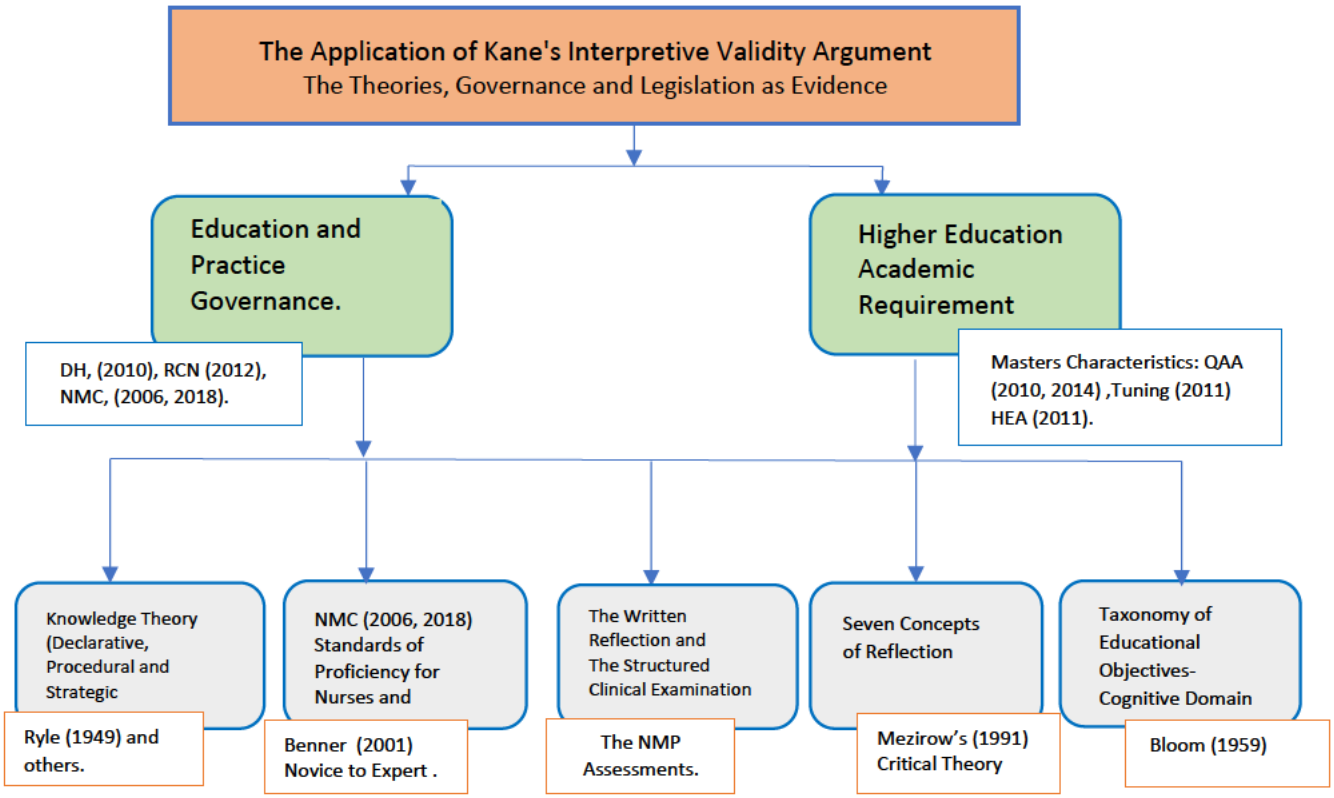


Fig 3.4 illustrates the application of Kane's IVA, identifying academic and nurse education's governances reside. Including the four teaching and learning theories, where these are situated and used in NMP nursing education and practice.

3.4.1 Theories as Evidence Measures in the IVA

Much of the literature discussion related to validly conclude with the degree to which the assessment tasks map the construct (i.e., theory) of interest. An advance of the validity debate in this study is interpretive validity theory and applying the evidentiary processes to justify test score interpretations (Cizek, 2012; Kane, 2013). The IVA works by building on the classic content-criterion validity theories and applies the cited teaching and learning theories as evidence for the logic of an argument supporting the test score claims of student proficiency. The IVA model is more suitable (with MTT measurement) for validating the students NMP-KSPs development and cognitive

learning to provide evidence for the warrants 1 to 4, also presented in chapter 1. In doing so, answer the stated research question(s). Consequently, the identified theories served several purposes, (a) they defined the observed performance domains in the WRs. (b) they provided an effective way of interpreting the WR content, and (c) in the SCE, the theories explicitly laid out the markers for performance measurements. Thus, assumptions about the student and their clinical practices are confirmed or refuted by theory.

CHAPTER 4: THE QUALITATIVE DATA ANALYSIS FOR THE SUMMATIVE WRITTEN REFLECTIONS.

4.0 Introduction

There is an agreement in the literature that students should apply concepts, theories, and other forms of knowledge to clinical situations. During reflection, theory and knowledge are used to identify patients' problems and arrive at decisions about the prescribing interventions and actions to take. This study utilised Mezirow's (1981) seven concepts of reflection to assess students' reflective ability to apply their learning in different clinical contexts reflectively. This chapter presents the results of the content analysis of the WRs using TDTA codes derived from the NMC (2018) standards of proficiency for prescribing and theories presented in chapter 1. The NMC characteristics are operational in examining the text using the codes found across the 28 sampled WRs. This chapter addresses the following,

Sub-question [1]. Are the NMP Summative Written Reflections (WRs) a Valid Constructive Indicator of NMP Students' Proficiency and Certification/Licensing to Practice?

4.1 Results and Findings

Intermediaries distributed study information to a cohort of ninety students, and 52 students consented. From the convenience sample of N=52, NMP students assessed work n=28 were randomly selected. The N=28 sample size is considered appropriate for the case study design (see Methods, chapter 3). Two raters independently provided coding applied to each of the sampled students' work (see chapter 3). Inter-rater agreement (IRA) was conducted on the N=28 written summative and Inter-rater reliability (IRR) calculated using Cohen's Kappa (ƙ). Table 4.1 contains an extract of the Kappa results as an illustration for discussion. For complete results, see Appendix 8.

Table 4.1 Cohen's Kappa (ϰ) Results for 2 Independent Raters' in the Application of the TDTA **Codes to** the WR Content

A Simple Random Sample n=10 of N=28: Percentage Agreement Tolerance =0						
Student Nos		Agreement (%)	Kappa (Unweighted)	*Z*	p-value	Strength of Agreement
1	XX01	87.4%	0.759	11.7	0.01	Almost Perfect
2	XX06	90.8%	0.756	11	0.01	Almost Perfect
3	XX09	86.1	0.692	10.6	0.01	Strong Agreement
4	XX11	92%	0..817	10.3	0.01	Almost Perfect
5	XX13	89.6%	0.803	9.4	0.01	Strong I Agreement
6	XX16	87.4	0.769	10.8	0.01	Strong Agreement
7	XX18	94%	0.872	10.8	0.01	Almost Perfect
8	XX21	86.2%	0.71	10.5	0.01	Almost Perfect
9	XX23	88.5%	0.81	11.9	0.01	Almost Perfect
10	XX26	88.1%	0.766	10.3	0.01	Strong Agreement
Average Agreement 89.01%						

Table.4.1 contains an excerpt from the Kappa results in appendix 8. n=10 randomly selected from the N =28 inter-rater agreement (IRA) calculations. Based on the above sample IRA are > .80 the accepted range - strong to almost perfect agreement (McHugh, 2012).

The ϰ answers how well the raters agree when assigning codes to text. Perfect agreement is evident when ϰ equals 1, indicating that both raters distribute codes the same (Salkind, 2014). A ϰ value of equal to zero suggests that the agreement is no better than a chance agreement (Bakeman and Quera, 2011). As seen in Table 4.1, the IRA Kappa results ranged from 0.692 (strong % agreement), the lowest to 0.872, the highest % value, almost perfect (McHugh, 2012). In Kappa, the results for the z and p-value columns relate to a significance test, and the results z= (> +2 and p =(> 0.5) indicate that the agreement between raters was substantial (Bakeman and Quera, 2011). Kappa uses the z-score to calculate Confidence Intervals (CI). Therefore, CI's were obtained for estimating that the ϰ statistic is a reliable measurement of IRR coding. The CI measurement addresses Kappa's certainty that it reflects an unbiased agreement measurement of in-text code labelling. The most accepted CI value is 95%, and in Kappa, the range is set at 95% or 99% (Field et al., 2012). The accepted CI for this study is 95%. The calculated results obtained for the N=28 sampled data are at 95% (see Appendix 8 for results). The p-value is a probability associated with the Kappa test statistic and the likelihood that the

rater agreement has occurred by chance. The literature identifies the need for a predetermined cut off point customarily set at 0.5 (Tappen, 2016). The p-value results across the sample are > 0.5, indicate that the IRA has not occurred by chance.

4.1.1 The Non-Medical Prescribing (NMP) Subject Domains: Theme 01- The Consultation.

The NMP students are responsible for selecting the prescribing episode for reflection, using the prescribing episode to contextualise the NMP criteria within an 'aspect' of their prescribing practice (see section 3.2.3). The reflective content drives the codes that appear in the text. See Appendix 6. The expectation is for a well-demonstrated proficiency achievement in those self-chosen criteria. The analysis begins with Ryle's (1949) knowledge theory demonstrated within the WRs, followed by the remaining theories listed in this chapter 1: 1.4.

Table 4.1.1 TDTA Code Key for Theme 01: NMP Subject Domain- The Consultation

Code	Descriptor	Code	Descriptor
1	Allergies	14	Medications History
2	Adverse Drug Reaction [ARD] Reporting	15	Medicine Altered by, e.g., age etc.
3	Cost Effectiveness	16	Medicines Licencing, monitoring, supply
4	Consent	17	Misuse Potential of Drug
5	Consultation	18	Over the Counter [OtC] Medications
6	Cost Effectiveness	19	Pharmacokinetics of Medicines
7	Disease Severity Assessment	20	Precautions- Contra indications
8	Drug Interactions	21	Prescribing Safely
9	Identify Desirable/ Undesirable Outcomes	22	Presenting Condition
10	Knowledge of NPF Formulary	23	Principles of Evidence-Based Practice
11	Knowledge of Product	24	Public Health Medicines Use
12	Medical Conditions- natural progress	99	Drug Mode of Action
13	Medical History		

Table 4.1.1 displays the TDTA codes for theme 01 consultancy and the code descriptors for each competence criterion derived from the NMP proficiency framework. For the complete codes see Appendix 6).

Whichever consultation model students referred to in the text, the remit is to explore the relevance of the model to NMP practice through reflection on the experience. The consultation models' application is argued, defended, and internalised. The model's principles transferred into any clinical and practice setting described (Charlton, 2007; Silverston, 2013). The reflection should ideally attend to either or all types of knowledge

theory, the impact of the consultation models in their NMP practice, and the new knowledge derived from the model's application.

4.2. Ryle's (1946) Knowledge Theory: Declarative (Know-that) Knowledge

4.2.1 TDTA Code 5, The Consultation

NMP Students (St.) recalled (declared) knowledge was evident in the naming of the consultation models. Students should add a rationalise, i.e., know-why the chosen model is appropriate for their consultation. For example, if the consultation was difficult and there was no explicit symptom complaint, a general inductive model is relevant. To define the reason for the patient's attendance, to consider or out rule other problems identified and choose an appropriate treatment plan. When there is a diagnosed condition, and the patient has signs and symptoms of being unwell. The NMP practitioner can apply NMP subject-domain, knowledge and theory to quickly identify and resolve symptoms with appropriate medications, treatment or intervention. In this case, a more focused enquiry consultation model may be more suited. As illustrated in table 4.2.1, the reference to consultation models (Code 5) was found in n=15 of the sampled N=28 in several places. Where present in the WR text, was limited to either a statement, i.e. 'a consultation' or naming a specific model, e.g., Neighbour's (1987) or the clinical environment where the consultation took place.

Table 4.2.1 The Consultation: Code 5, Results in Summary

Knowledge Theory		Student No's
Declarative Knowledge; Know-that	**n=15**	02, 08, 09, 12, 13, 14, 16, 17, 19, 23, 24, 25, 26, 27, 28
Procedural Knowledge; Know-how	**n=0**	
Strategic Knowledge; Know-why	**n=0**	
Code not present in the text	**n=13**	01, 03, 04, 05, 06, 07, 10, 11, 15, 18, 20, 21, 22
Sample analysed	**N=28**	

These (n=15) students referred to some of the principles of an effective consultation' giving details of the environment, i.e. ward, outpatient's clinic in which the consultation

took place. For students to demonstrate proficiency, declarative knowledge should be supported by considering the model's biomedical, physiological approaches and principles to (reflectively) reason its use in the specific NMP situation.

4.2.2 TDTA Code 1, Allergies

As illustrated in Table 4.2.2, the students n=10 of the sampled N=28 addressed this specific competency criteria allergy (code 1). Where present in the WR, in a statement, "I checked if the patient had any known allergies" (see, for example, Appendix 6e p. 240). This NMP criteria 'allergies' is a fundamental part of nursing and prescribing. The condition is documented as a standard patient history investigation in clinical practice and heavily influences NMP care management. The information on allergies as part of the Medications History (code 14) is closely associated with maintaining patient safety (code 21). Allergies are an essential part of the global information gathering process within the consultation. If incomplete in the historical patient information, it can lead to diagnostic and prescribing errors. Other potential sources of prescribing errors are related to drug interactions (code 8) and over the counter (OtC) medications (code 18). Errors or omissions in these code categories are attributed primarily to practitioners obtaining an incomplete medications history. However, in NMP, a comprehensive assessment must include allergies, given its implications on new prescription drugs and existing drug regimes. Where code 1 is present in the WR, it is a declarative fact, i.e., a single word embedded in a concise sentence.

Table 4.2.2. TDTA Code 1 Allergy, Results in Summary.

Knowledge Theory		Student No's
Declarative Knowledge; Know-that	**n= 10**	04, 07, 08, 12,13, 14, 22, 23, 25, 28
Procedural Knowledge; Know-how	**n=0**	
Strategic Knowledge; Know-why	**n=0**	
Code not present in the text	**n=18**	01, 02, 03, 05, 06, 09, 10, 11, 15, 16, 17, 18, 19, 20, 21, 24, 26, 27,
Sample analysed	**N=28**	

Once the student affirms the presence of allergies, the subsequent discussion terminates abruptly. Declarative knowledge, i.e., the questioning of the patent about any allergy, is present. An exploration of allergies, for example, its presentation such as urticaria or

milder symptoms such as a skin rash, blistering or itching, is unaddressed. A significant and potentially missed problem is medication history. A common cause of drug-related fatality is anaphylaxis[1] or hypersensitivity and hyper-susceptibility. This criterion is essential to NMP as it can compromise new drug treatment.

Furthermore, an exploration of unprescribed medicines should include OtC, as they can have profound implications for patient safety and the potential for drug interactions (code 8). There is adequate exposure in theory and practice in the undergraduate and postgraduate curriculum and nursing training related to identifying and recording patient-reported allergies. Also, the subsequent procedural knowledge is contextual to the course of action. I.e., what the practitioner does with this information, the prescribing measures taken in the presence of a known allergy, is missing where the code identified in the WRs.

4.2.3 NMP Subject Domain- Theme 02: Establishing Options (Patient Safety).

Establishing Options (Theme 02) takes an overview of patient safety in the context of NMP. Establishing options is an opportunity for the student to demonstrate NMP subject learning in areas related to making prescribing safer in response to reports on medications error and harm to patients. The purpose of patient safety in the NMP curriculum is for students to understand their role in minimising the occurrences and the impact of any prescribing events that may cause the patients' harm. Patient safety is paramount to the overall consultation (Theme 01). Some students have addressed aspects of this theme in their WRs. See Table 4.2.3 for results.

SEE END NOTE
[1] Anaphylaxis

Table 4.2.3. TDTA Code Key: Theme 02: NMP Subject Domain Patient Safety (Establishing Options)

Code	Code Descriptor
25	Physical Examination
26	Differential Diagnosis
27	Effects of Multiple Pathologies
28	Medications Contraindications on Treatment Options
29	Establishes and Maintains Plan for Reviewing
30	Holistic Needs Assessment
31	Monitors Effectiveness of Treatment
32	Most Appropriate Treatment
33	Working Diagnosis
34	Requests, Interprets Relevant Diagnostic Tests

Table 4.2.3 displays the TDTA codes for theme 02 Patient Safety and the code descriptors for each NMP competence criterion derived from the NMP proficiency framework.

4.2.3.1. Knowledge Theory: Procedural (Know-how) Knowledge.

4.2.3.1.1 TDTA Code 25, The Physical Examination

The Physical Examination (PE) is a necessary NMP feature for differentiating between normal and abnormal bio-physiology to enhance the probability of correct diagnosis and subsequent prescribed treatment. PE is foundational to the underlying condition's exposure, the cause of the patients' current ill health, and facilitating the treatment plan. The PE supports the differential diagnosis (code 26) and the concluding working diagnosis (code 33). The PE can be a rapid but comprehensive head-to-toe assessment or a fast, focused Airway-Breathing-Circulation-Exposure-Disability (ABCDE) assessment using the ALERT ™[2] (2002) protocol. As a guide to the practitioner, the ABCDE assessment identifies which systems require further examination, e.g. cardiovascular, respiratory or others. The ABCDE is customarily carried out in emergencies. Alternatively, a broad category assessment technique can include the following. Inspect- (Observation) – Palpation (Touch)–Percussion (Sound produced by striking the body surface)– Auscultation (listening to sounds), e.g. breath or heart sound produced by the body.

SEE END NOTE

[2] Acute Life-threatening Event, Recognition and Treatment (ALERT)

Depending on the patients' presenting condition, past medical history, PE and other information gained during the consultation, knowledge and thinking are essential.

Table 4.2.3.1.1 TDTA Code 25, Physical Examination (PE), Results in Summary.

Knowledge Theory		Student No's
Declarative Knowledge; Know-that	**n= 16**	03, 04, 07, 08, 10, 11, 12, 13, 14, 15, 16, 19, 21, 24, 25, 26, 27, 28
Procedural Knowledge; Know-how	**n=02**	17, 20
Strategic Knowledge; Know-why	**n=0**	
Code not present in the text	**n=10**	01, 02, 03, 05, 06, 09, 18, 22, 23
Sample analysed	**N=28**	

In Table 4.2.3.1.1, eighteen of the N=28 sampled students declaratively referred to the PE in the text. Within the PE, the NMP is responsible for ascertaining a reason to initiate further investigations (code 10). To prescribe (code 21), appropriately treat (Code 32), diagnose (26) or refer (Code 56). Students 8, 12, 13, and 17 stated that they used a focused PE assessment applying the A-B-C-D-E, and n=10/N=28 students reported that their PE consisted of inspection and observation. Of whom n=2 in the application of the protocol, taking the biometric measurement is identified. These students knew how to apply the biometric recordings. For example, St 02 and 26 interpreted the objective (biometric recordings) and subjective (patient responses) and other findings to determine the cause of the presenting condition in the WRs text. In their work, St.17 and St.20 documented in their appendix the complete assessment technique and the rationale for committing to the chosen PE.

St.17
"...At the end of the objective examination, the NMP summarised and documented patient -X's pain region/symptoms... due to the number of 'red flags'...., Based on both the subjective and objective examination (Appendix 2) ..., alerted the NMP to the possibility of more serious disease, such as a tumour..."

St.20
"....., The Physical examination..., [including] recordings such as Blood Pressure was very important as women on contraceptives have an increased risk of having Thrombosis[3] and cardiovascular complications[4]

[3] Thrombosis
[4] Cardiovascular Complications

Procedural knowledge is useful in guiding and directing NMPs in what intervention is needed and how to reach a solution. Procedural is acceptable as this knowledge type aligns with the real world of clinical practice. Cognitive thinking processes about anatomical, pathophysiological diagnoses, i.e., the cause of the presenting condition and effect, the deteriorating health, are vital to the PE. Declarative knowledge should preferably be in conjunction with procedural knowledge to reach a correct differential and working diagnosis (code 26 and code 33, respectively). Procedural knowledge on its own can be problematic as it can hinge on the performance and knowing only the next-step sequence of a psychomotor skill. In practice, the procedure skill can give the appearance of proficient, safe and efficient practice, even in the absence of underpinning cognitive knowledge of anatomy, bio-physiology and the pathological process of disease.

4.2.3.1.2 TDTA Code 34, Requesting and Interpreting Diagnostic Tests

Requesting and interpreting diagnostic tests is essential for diagnostic accuracy and for making a definitive diagnosis. The probability or working diagnosis of a disease is estimated. A positive test result acts as an indicator that the diagnosed condition is present. Tests can be invasive, i.e. those requiring a break into the skin, such as blood tests, biopsy, or body orifice exploration as in endoscopy. Otherwise, non-invasive tests, i.e., Electrocardiogram (ECG) X-Ray or Multi-Resonance Imaging (MRI) scans. For code 34 presented in n=09 of the sampled N=28 WRs, the test requests varied from non-invasive to invasive or both. Students St. 03, 12, 22, and 27 communicated declarative and procedural knowledge by checking for tests requested. The excerpt below illustrates code 34 as presented in the WRs.

> st 03, 05, 11,15 “results from routine and standard screening tests were present”, or stating the test requested
> St. 22, 27, urea, electrolytes, and St. 12‘ haemoglobin’ ,

Table 4.2.3.1.2 TDTA Code 34, Requesting and Interpreting Tests, Results in Summary

Knowledge Theory		Student No's
Declarative Knowledge; Know-that	**n= 07**	03, 05, 11, 12, 15, 22, 27
Procedural Knowledge; Know-how	**n= 02**	01, 23
Strategic Knowledge; Know-why	**n= 0**	
Code not present in the text	**n= 19**	02, 04, 06, 07, 08, 09,10, 13,14, 16, 17, 18, 19, 20, 21, 24, 25, 26, 28
Sample analysed	**N=28**	

Table 4.2.3.1.2 summarises the N=28 sampled WRs; n=2 students' WRs demonstrated that they know that clinical test results are significant. In requesting appropriate tests, students' established that they knew how to do so as 'a matter of procedure'. For the most part, know-how evidence represented the mechanical, technical activity of obtaining a test or seeking the result. Knowing that a diagnostic test is needed and knowing how to request the appropriate tests allows the practitioner to use know-that, know-how knowledge to achieve the desired proficiency outcome. The WRs content did not extend to demonstrating an understanding, as Bloom (1956) defines it, by interpreting the results in the context of NMP practice. In the excerpt from St. 01, there is evidence of 'know that' regarding requesting the test result for prescribing treatment, not 'why,' i.e. to assess the bisphosphonate treatment's efficacy. However, these tests work as a translation of the body's malfunctions due to illness. The student (st.01) did not discuss the associations between the test results, symptoms, disease and the medication in their WR.

St. 01

"...., Jane's latest requested blood tests results indicated that her Procollagen type 1 N-Propeptide[5]; a marker of bone turnover had not reduced as one would expect in a patient taking bisphosphonates[6].....lady had sustained two vertebral wedge fractures five months ago whilst lifting....' '.... fragility fractures.... highly suggestive of osteoporosis [7] ...' this enabled exclusion of other illnesses that could have caused ... fractures; such as myeloma[8]..."

This student, 01, did not demonstrate that they know why the test results were significant to the intended or future medications prescribing. Student 01 named and reviewed the

SEE END NOTE

[5] Type 1N-Propeptide (T1NP).

[6] Bisphosphonates

[7] Osteoporosis.

[8] Myeloma.

result of a specific blood test, the Procollagen type 1 N-propeptide (T1NP). However, concerning the T1NP serum concentration levels found in the test results, the reference intervals range for normality and abnormality is not discussed in the context of prescribing. Neither was the correlation between serum P1PN levels and an (or any other) specific or plausible underlying cause for fragility (stress) fractures. Bisphosphonates' optimal effect requires adequate calcium and vitamin D intake before and during therapy. Here knowledge of product (code 11) requirement is missing as an alternative rationale for the unreduced bone turnover. In the NMP context, exploring medications history (code 12) may highlight the absence of vitamin D in the treatment plan or omission due to patient compliance. Assessing the patient's diet and lifestyle under holistic needs assessment (code 30) ensures treatment compliance. Holistic needs assessment identifies desirable or undesirable outcomes (code 09). Alternatively, there may be an issue with clinical effectiveness (code 03) concerning bisphosphonates as treatment. In the St 01 WR example, know-how to request investigations is clinically specific, domain-focused, but insecure in a clinical NMP practice context. The coding of WR content did not suggest that the clinical practice techniques are refined or to a level of proficiency.

4.3 NMP Subject Domain-Theme 03: Communicating with Patients'

A consequence of governance and policy is ensuring patients have a good experience, a sound practitioner-patient relationship, practitioner-patient communication. Patient involvement in medication prescribing decision-making is an NMC governance and NHS policy requirement. The communication strategy aims to develop robust patient-practitioner relationships to ensure medication compliance. The healthcare literature consistently reports that patients' satisfaction is perceived as good communication that influences treatment compliance. See, for example, the Kings Fund (2017) Young et al., (2017). Communication is vital to delivering safe, quality care, thereby improving the patients' healthcare experience and outcome. Table 4.3 contains the TDTA codes and code descriptors for the NMP Theme 03 Communication.

Table 4.3 TDTA Code Key, Theme 03: NMP Subject Domain-Communicating with Patients

Code	Code Descriptors
36	Adapts Consultation style
37	Checks Patient Understanding of Treatment
38	Clear Instructions
39	Communication with Patient
40	Communication Skills
41	Cultural Religious Implications of Prescribing
42	Deals Sensitively with Patient
43	Existing Medications, Contraindication on Treatment Options
44	Explains Nature of Patient Condition Treatment
45	Helps Patient- Informed Choices
46	Listens to Patients Beliefs, Expectations
47	Negotiates Outcome- Patient /Prescriber
48	Potential Risk-Benefit of Treatment
49	Referral- Preventative Measures
50	Non-Drug Treatment Option

Table 4.3 displays the NMP subject domains for Theme 03, and the communication TDTA code descriptors for each competence criterion derived from the NMP proficiency framework.

In pre-registration education, communication is a core competency. Most nurse NMPs should have a high degree of transferrable skills, underpinning knowledge and practical experience regarding therapeutic communication. Therefore, NMP students should demonstrate effective patient communication with ease, and subsequent prescribing should be successful. In the context of these results and NMP practice standards, this was not the case. A revisit to this study's core competency proved to be a worthy investigation and analysis.

4.3.1 Knowledge Theory: Strategic (Know-why) Knowledge

The strategic knowledge is less explored and discussed in the literature, although this component has been repeatedly identified as essential and is representative of higher-order thinking (HoT)skills. A skill necessary for recognising the causal relationship between the phenomenon and for planning appropriate interventions. In the literature, strategic knowledge is a reliable maker of inquiry, systematic hypothesis testing, innovation, and new knowledge generation, a definitive characteristic of proficiency (expert and mastery) achievement.

4.3.1.1 TDTA Code 42, Dealing Sensitively with Patients

Dealing sensitively with patients requires the communication skill of listening and emphasises that the NMP practitioner adopts a holistic approach to prescribing. NMP students should demonstrate that they know why it is essential to consider the patient's age, gender, psycho-emotional and cultural needs during the consultation. As illustrated in Table 4.3.1.1, the data mining retrieved code 42- Deals Sensitively with Patient from n=7 written reflections. Five students addressed this code declaratively in their work. Only n=2 students of the sampled N=28 convincingly showed know-why knowledge of this communication principle.

Table 4.3.1.1 TDTA Code 42, Deals Sensitively with Patient Results in Summary

Knowledge Theory		Student No's
Declarative Knowledge; Know-that	**n= 05**	07, 08, 09, 12, 14.
Procedural Knowledge; Know-how	**n= 0**	
Strategic Knowledge; Know-why	**n= 02**	16, 19.
Code not present in the text	**n= 21**	01, 02, 03, 04, 05, 06, 10, 11, 13, 15, 17, 18, 20, 21, 22, 23 24, 25, 26, 27, 28
Sample Analysed	**N=28**	

The NMP may know that dealing sensitively with patients is the best way to achieve their prescribing aims and outcomes. They may also know how to interact sensitively with the patient. However, students may not know why this code criterion is significant to the patients' experience and its implications. Nor the possible impact this may have on NMP practice in achieving patient concordance with medications and other treatment plans. Student 19 demonstrated know-why this code criterion serves as the basis for attaining collective goals for medicines prescribing to treatment outcome.

St.19

"... Including patients in the decision-making process of their care...asking patients to actively participate in discussion and make important decisions about their care is the way forward for me in achieving concordance with Mr K was to give him enough time to understand what I wanted to achieve for him and more importantly enough time for him to express his concerns".

Student 19 contextualised the patient's consent, and agreement with medications was necessary, considering the patient's concerns- relevant for achieving the prescribing outcome. The above excerpt indicates that the student knows why communication quality can have adverse clinical consequences if it leads to patients failing to adhere to the treatment plan (code29) and take the prescribed medication.

4.3.1.2 TDTA Code 50, the Non-Drug Treatment Option

Conducive to the prescribing process is the practitioners' ability to facilitate the patients understanding of their ill-health condition. Particularly as the patient and their families' health beliefs are a significant factor in successful prescribing practice, i.e., taking or not taking the prescribed medication. For this (code 50) NMP criterion, as seen in Table 4.3.1.2, code 50 is identified in eight n=7 of the N=28 students WRs. Only one of whom (St.16) demonstrated know-why knowledge in this code category.

Table 4.3.1.2 TDTA Code 50, Non-Drug Treatment Option Results in Summary.

Knowledge Theory		Student No's
Declarative Knowledge; Know that	**n= 04**	04, 09, 20, 21,
Procedural Knowledge; Know-how	**n= 02**	27, 28,
Strategic Knowledge; Know- why	**n= 01**	16
Code not present in the text	**n= 21**	01, 02, 03, 05, 06, 07, 08, 10, 11, 12, 13, 14, 15, 17, 18, 19, 22, 23 24, 25, 26.
Sample analysed	**N=28**	

The excerpt from Student 16 demonstrated the student's know-why knowledge in consideration of non-drug treatment options was an important factor in treating their patient.

St.16

"...Patient A was informed of some psychological intervention such as Cognitive Behavioural Therapy (CBT) [9] that could be helpful for him, so as to establish alternative treatments and ways of coping with his symptoms. The treatment of psychosis [10] is not restricted to drug interventions; this should always be supported by psychological and social interventions, carefully planned for the individual.... Patient A and his parents were advised that we would not prevent them from seeking spiritual healing for their son".

SEE END NOTE

[9] Cognitive Behavioural Therapy (CBT)

[10] Psychosis

Disease characteristics can be very troublesome to the patients. Some patients may have the notion that medications, in general, are more harmful, addictive or poisonous and should not be ingested. In the presence of polypharmacology[11] (code 19) and multiple pathologies[12] (code 27) and in the absence of prescribing the lowest dose for the shortest possible time, non-drug treatment options should always be considered. The expectation is that NMP students' explore the issues around the patient's perceptions regarding their presenting condition (code 22) and treatment intervention (Code 32). Exploration ensures that the patients' values and beliefs are not counterintuitive to prescribing safely (code 65). In context to providing appropriate treatment (code 32) or establishing a review plan (code 29) for concordance, adherence and compliance.

The analysis revealed that evidence and the presence of strategic knowledge across the sampled N=28 written reflections are not prevalent in this NMP cohort. This knowledge type is key to developing deep learning and insight into understanding the subject domain's principles and other theory's clinical practice functions. Strategic knowledge should underpin practice and practical performance, procedural process or technique. This knowledge is fundamental to safe, proficient practice. The absence of strategic knowledge can position students into over-reliance on the formalised structure of procedure and performance, i.e., codified as standardised competency procedures. Strategic knowledge theory explains theory practice interrelatedness for identifying the causality and considering an appropriate diagnosis (code 33), treatment and intervention plan. Consider, for example, cardiopulmonary resuscitation (declarative plus procedural) and the resultant (strategic) medications required. In combination with theory to support, this activity stimulates high-inference deliberate learning. Section 4.2 and 4.3 verified that students have solid declarative knowledge and procedural knowledge; however, minimal strategic knowledge achievement is demonstrated across the N=28 WRs.

4.4 Mezirow's (1981) Seven Concepts of Reflection

Mezirow’s (1981) framework has an advantage over the other modules previously outlined for analysing reflection. Mezirow’s reflective levels are operationally defined and

SEE END NOTE
[11] Polypharmacology
[12] Multiple Pathologies

conceptualised, and it provides a more precise definition and frame of reference for determining the students' reflective development and achievement. Therefore, here is a sufficiently robust framework for examining the differences in reflectivity, the levels, and components essential to critical reflection demonstrated within the students WRs. The TDTA coded and text mined the seven concepts of reflection to evidence and analyse the students' reflective disposition, e.g., metacognitive, self-corrective, logic in decision-making, and problem-solving other abilities in the students described NMP events in their WRs.

Fig 4.4 Mezirow's (1981) Seven Concepts of Reflection.

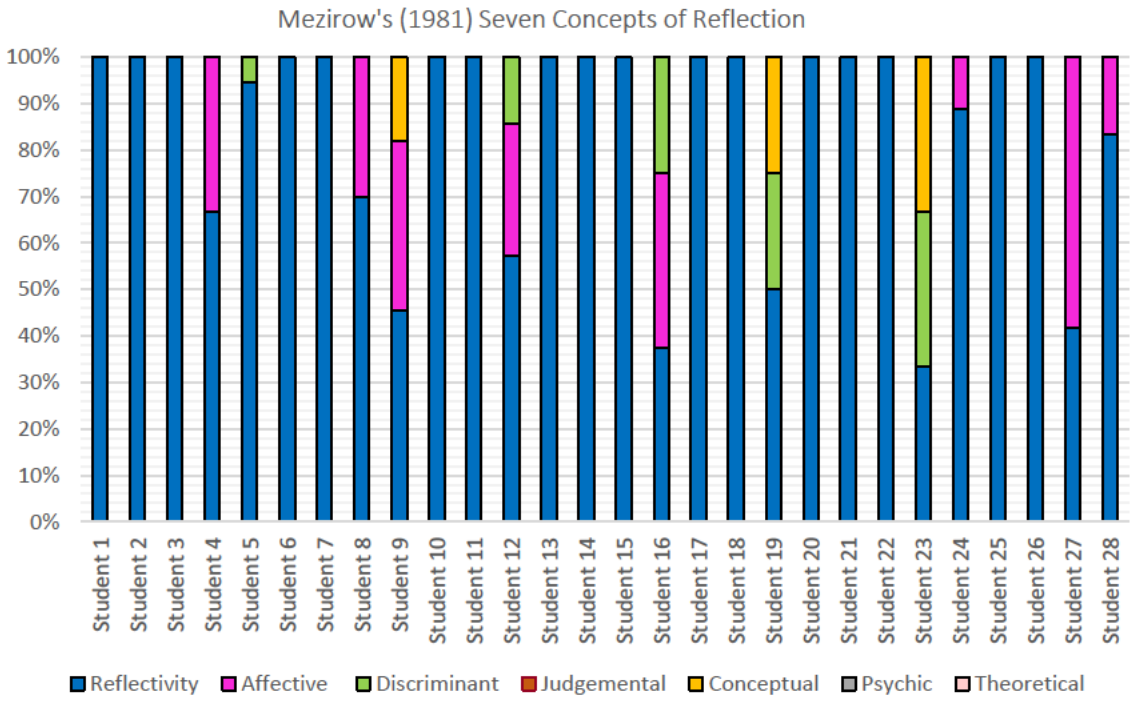


Fig 4.4 Summarises the Seven Concepts of Reflection criteria where coded and identified across the N=28 sampled Students' written reflections. Where Mezirow's (1981)criteria are demonstrated, only occasionally did some students reach the higher-order thinking and more challenging levels of reflectivity espoused by Mezirow (see chapter 1, Table i.4.3 for lexicon).

4.4.1. Mezirow's First Concept: Reflectivity

Mezirow (1981) defines his first concept, reflectivity, habitus, and non-reflective customary or traditional practice. Although the student may be aware of a specific practice behaviour (a habit), a form of seeing, thinking or acting, there is no critical consciousness. For example, the individual may have good practice sense and familiarity with the practise field. Familiarity is a condition of participation that, like prevailing (habitual) customs and traditions, equates with the achievement of an assessed task in theory and practice. Fig 4.4.1 illustrates those students whose WR content were attributed to reflectivity, concept 1- level 1. Level 1 is identified more than once in all the students' work, some students [n=17] scoring 100% in this reflective category.

Fig 4.4.1 Mezirow's (1981) Concept 1: Results in Summary.

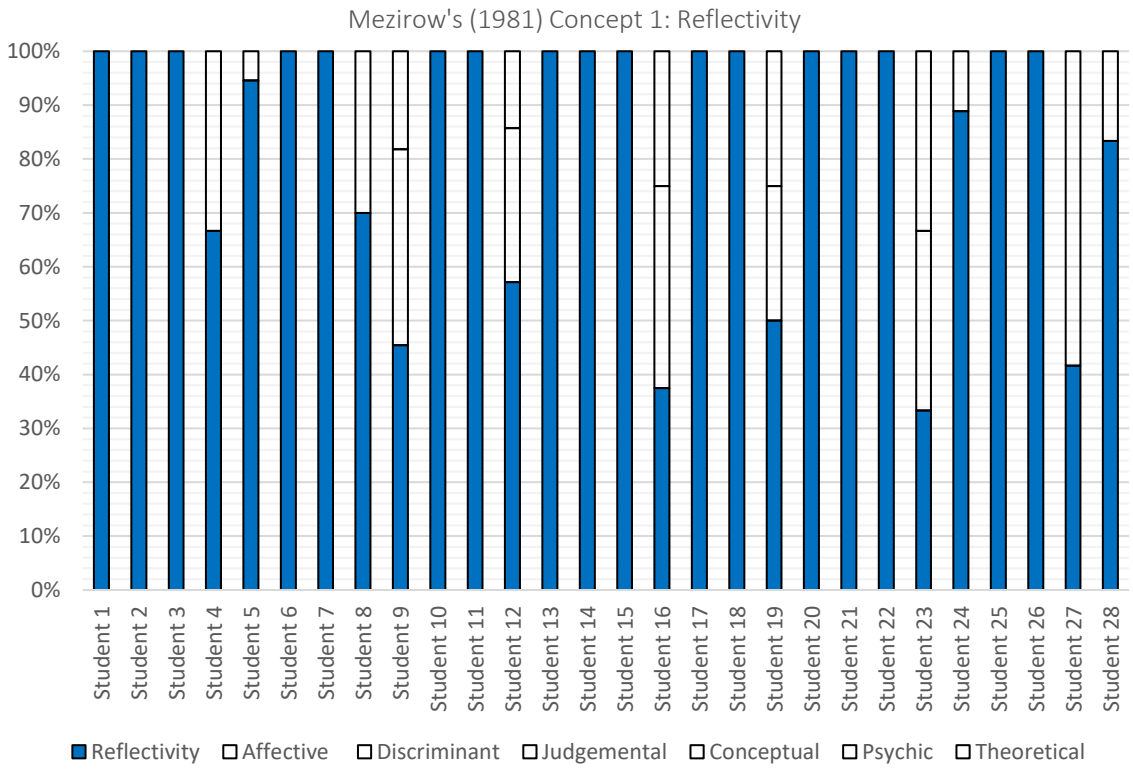


Fig 4.4.1. illustrates the Mezirow's Reflectivity criteria coded and identified across the N=28 sampled WRs. Of the n=8 /N=28 students' who presented this criterion in their WRs, the lowest % value for the Reflectivity criterion was found in St. 23 at 32% and the highest % ratings were identified in n=18 students assessed work at 97-100%

In Fig 4.4.1, the results showed level 1 reflection as the most frequently present code, in N=28 students WRs, with a content rating of >70% in n=21 /N=28 in this reflective category. The following excerpts typify reflectivity found in the written reflections, consisting of descriptive information, limited statement of fact, or reference to their feelings.

St.13

"…, the reflection and analysis of this consultation has shown me how far I have come in developing the necessary skills and knowledge to prescribe safely for my patients".

St.15

"…, I believe using reflection in everyday practice will aid me to become an expert practitioner. I feel reflection will help me to rationalise and deal with stressful situations and complex patient situations ".

The example descriptors form students13 and 15embody non-reflection (level 1). Devoid of the critical consciousness typified at reflectivity level 5 to 7 required to evidence the development of NMP learning, abilities and practise (see Appendix 6). These students limited themselves to declarative knowledge about reflection and novice clinical and theoretical application. The individuals' beliefs, subject knowledge base, and theories underpin their understanding, which situates their practice remained unarticulated. The overall result was that many students failed to reach more challenging reflection levels—for example, level 2 (Affective Reflectivity). Particularly, levels 5 to 7 representative of the critical theory underpinning reflective practice. The levels 5 to 7 of reflectivity are the attributes AP, expected for NMP.

4.4.1.1 Concept 2, Mezirow's Affective Reflectivity

Affective Reflectivity, outlined in Chapter 1, Table 1.4.3, is an awareness of an individual's feelings about the way they are thinking or acting, and habits for doing so were found in n=8 of the WRs. At this level, students express their feelings, demonstrate an awareness of those feelings and the potential impact they may have in their perception of the practice event. Awareness of feelings requires an ability to assess the efficacy of thought, actions, habits, ways of doing things, and questioning the clinical activity. As discussed in

chapter 6, several authors contest that attending to feeling is not consistent with the reflective philosophy. However, in Affective reflectivity, the learner looks for explanatory answers to understand a claim or action and challenge it. Fig 4.4.1.1 illustrates written reflections where Concept 2 -level 2, Affective Reflectivity, are attributed to students' work.

Fig 4.4.1.1. Mezirow's Concept 2: Results in Summary.

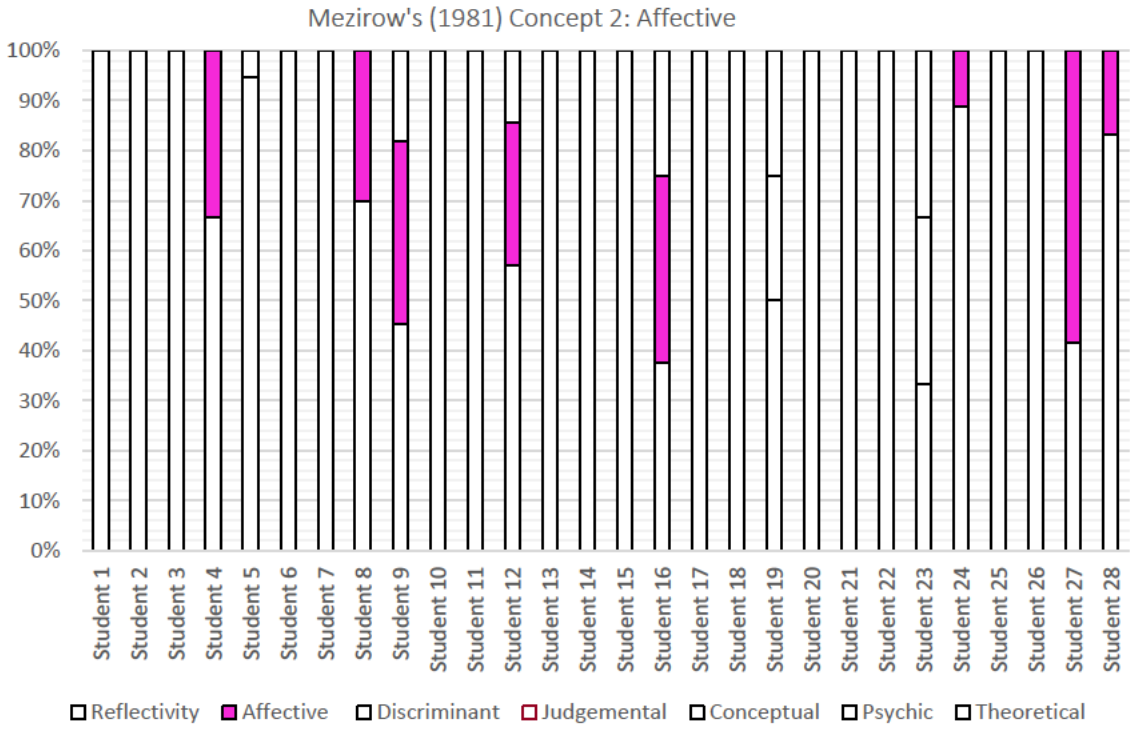


Fig 4.4.1.1 illustrates the Affective Reflectivity criteria coded and identified across the N=28 sampled WRs. Of the n=8 /N=28 students' who presented this criterion in their WRs, the lowest % value for Affective Reflectivity criterion was found in St. 11 at 12% and the highest % ratings were identified in St.27 at 58%.

A total of n=8 of the sampled N=28 WRs showed evidence of Affective Reflectivity, an awareness feeling about self, as seen in the excerpt from St. 04 below and St.08

St.04

"...., I felt awkward as it became clear to all, that the Dr had not read Jane's medical or clinical notes fully, prior to the appointment. We had arrived separately and as a result; we did not discuss Jane's case prior to going inside as would often happen routinely".

St.04 also identifies the routine practice of discussing the patient's care before the appointment. Still, it does not address the willingness to share and debate prescribing practice (code 74). The NMP code criteria are relegated to "the Dr neglecting to read the patient's medical notes before the consultation". Further, this incident does not support the theory in assessing patients' needs or the related health information applied to plan the patients' care (code 30). The planning and planned achievement for the patient are not reflective within the NMP role. Also overlooked is code 30, the holistic, humanistic approach, which ensures that healthcare practitioners perceive the client's unique needs. The excerpt from St.08 implies Mezirow's concept reflectivity. The student describes their feelings about their way thought and actions are processed. However, attempts at assessing the causes for these thoughts and actions and St. 08's professional relationship is, i.e., prescribing responsibility is in the situation, is not followed through in WR discussion.

St.08

"….It would be inappropriate to suggest prescribing more medication if there was a lack of compliance with the current medications as prescribing more would not solve the problem and have wider financial and poly pharmacy implications..".

The student 08 statement 'lack of compliance with the current medications' merits a thoughtful reflective exploration. A search for explanatory answers to understand the non-compliance claim and cause. Also, the rationale for this thought process toward the patient. An investigation of the patient's social determinants relative to the 'current medications' (code 14) may provide a physical, psychological or cultural-religious reason (code 41) for the 'lack of compliance'. The student recognised the implications for 'polypharmacology'. However, non-compliance to medicinal treatment indicates this is not a risk factor. Attention to the rationale for the patient's non-compliance could enhance the written reflection quality to level 5, conceptual reflectivity. If associated with the patient's holistic needs (code 30), lifestyle (code 41), values (code 46) or other prescribing reasons.

4.4.1.2 Mezirow's Third Concept: Discriminant Reflectivity

Discriminant Reflectivity represents the efficacy of the individual's perceptions, thoughts, actions and habits. This reflective level requires identifying the immediate cause with detailed reasoning and decision-making. This level anticipates that students will address some NMP subject knowledge. The students' reflectivity in the clinical context should be firm, along with evidence of prior skills gained in their NMP role development. Fig 4.4.1.2 provides a summary of the results for students' who demonstrated Discriminant Reflectivity in their WRs.

Fig 4.4.1.2. Mezirow's (1981) Concept 3: Results in Summary.

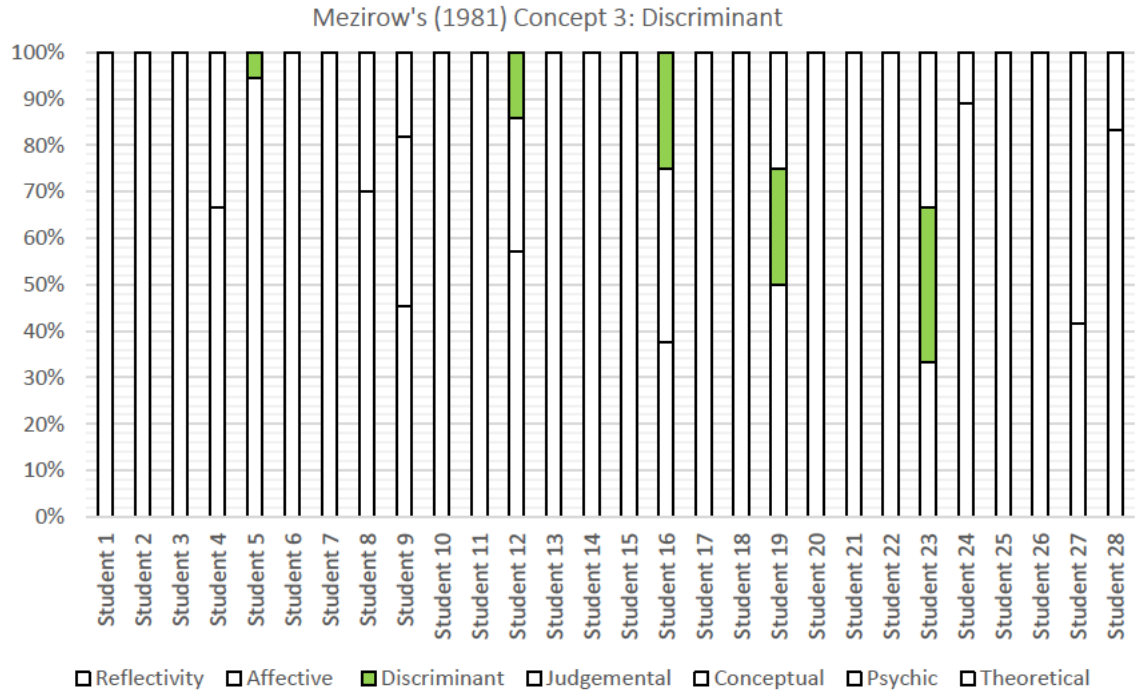


Fig 4.4.1.2 Shows the results for the Discriminant Reflectivity criteria where coded and identified across the N=28 sampled WRs. Of the n=5 students' who presented this criterion in their WRs, the lowest % ratings for Discriminant Reflectivity criterion were St.05 at 3% and the highest % ratings were identified in St.23 at 28%

Student 12 demonstrated an ability to assess their perceptions' efficacy, i.e. the clinical NMP activities to be carried out. Also, the treatment's impact on the patient and in a minimalist way, highlighting the habitual ways of doing things, is that the standard treatment for Acute Kidney Injury (AKI) is intravenous fluid (IV). However, the situational relationships between AKI and Heart Failure (HF), or the suitability and trustworthiness of

the student's perception of the treatment regime, are absent. Along with the confirmed diagnosis of AKI by other tests and the critical evaluation of the outcome of carrying out that care. Typical and representative of habitual ways of doing things found in the n=4 written reflections. As illustrated by the excerpt from student 12, unaddressed is the rationale for the student's cause of concern.

> **St.12**
> "The management of AKI[13] is, therefore, to give fluids…., simultaneously", … "I was also concerned about prescribing fluids to treat the AKI which could have a negative effect on the HF [14] with further fluid overload[15]."

AKI is diagnosed initially by elevated serum creatinine[16] levels, confirmed by a Glomerular filtration test (GFR[17]), and AKI management is intravenous (IV) fluid therapy to restore GFR and increase urine output. The treatment aims to re-establish systemic blood pressure and cardiac output, such as in HF and subsequent reduced cardiac output, the pre-requisite for AKI. Fluid therapy is contraindicated in HF as fluid-overload may impede the HF resolution and adversely affect cardiac function. Although the student articulates a concern for prescribing fluids to treat AKI, this is unrelated to recognising the reality of the context in which these clinical interventions function. For example, there was some discussion regarding the prescribing intervention choice or action, e.g. IV diuretics, but not much reflectivity on the learning experience within the complexity of the care situation. For example, other prescribing options to support the heart function, e.g., intravenous inotropes, which have a dual effect of cardiac support and subsequent diuretic effect to improve urine output, remained unexplored.

4.4.1.3 Mezirow's Fourth Concept: Judgemental Reflexivity

This concept was not found in the Sampled N=28 WRs.

SEE END NOTE
[13] Acute Kidney Injury (AKI)
[14] Heart Failure (HF)
[15] Fluid Overload
[16] Serum Creatinine
[17] Glomerular Filtration Rate (GFR)

4.4.1.4 Mezirow's Fifth Concept: Conceptual Reflectivity

Conceptual reflectivity is the ability to self-reflect, critique, and critically question whether the concepts employed, such as the political, cultural, social, economic constructs and legislation that influence healthcare practice, are adequate for making sound NMP judgment decision-making. Mezirow's fifth concept posited as a higher level as conceptual reflectivity is associated with critical consciousness. The fifth concept challenges healthcare practitioners to be aware of and engage in intellectual argument and debate about the influence of research, policy and legislation on practice. In Fig 4.4.1.4, a small number of students n=3 of N=28 engaged in concept reflectivity.

Fig 4.4.1.4. Mezirow's Concept 5: Results in Summary

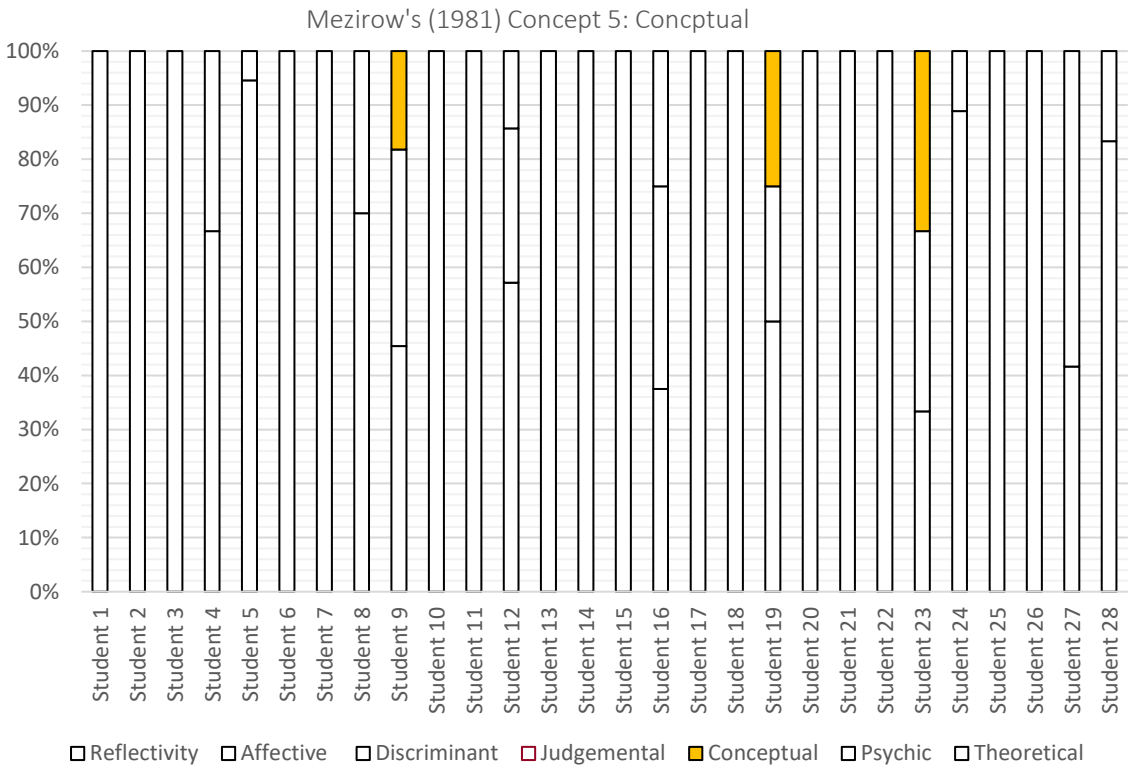


Fig 4.4.1.4 Shows the results for the Conceptual Reflectivity criteria coded and identified across the N=28 sampled written reflections. Of the students [n=3] who presented this criterion in their written reflections, the lowest conceptual reflectivity criterion % ratings were St. 09 at 19% and the highest % ratings were identified in St.23 at 32%

Student 19 begins to question policy and consider it to support and defend decision-making around prescribing. There is also an indication of an assessment and evaluation of nurse prescribing, its benefits to the broader NHS organisation, judgment about thoughts and perceptions for a holistic understanding of this situation.

> **St.19**
>
> "…. if nurse prescribers are to be taken seriously, we should welcome any opportunity to challenge policy and defend our decisions around prescribing, …. The only way to prove that nurse prescribing is beneficial to patients, colleagues and the NHS is to grasp the opportunity to prescribe and prove to those with scepticism, that nurse prescribers can become a valuable member of their team."

Absent from the written reflections are intelligent arguments and debates related to policy and critical questioning of how the policy may defend prescribing decisions. For example, and based on the above excerpt, the NMP problems with the social and cultural policy construct suggesting a perceptual change about the NMP role is required. Attention to the problems inherent to NMP practice, considering possible policy failure reasons, would convincingly address conceptual reflectivity. For example, the student's critical examination of their policy position concerning governance and legislation may help the NMP student conceptualise their perspectives. The critical examination result is questions concerning the socio-political barriers to NMP practice and policy to improve the sceptical attitudes encountered in NMP practice.

4.4.1.5 Mesirow's Sixth Concept: Psychic Reflectivity

This concept was not found in the Sampled N=28 WRs.

4.1.1.6 Mezirow's Seventh Concept: Theoretical Reflectivity

This concept was not found in the Sampled N=28 WRs.

4.5 Benner's (2001) Novice to Expert Model

Benner's (2001) model delineates the differences between the novice nurse who does not have the background experiential (in-depth) knowledge and the expert nurse. Expert nurses can apply knowledge, experience, and clinical reasoning to discern changes in the patient or changes in their own understanding of the clinical situation. These NMP students work as APs with prior nursing, education, clinical experience, and several years of practice beyond initial registration. At the master's level study, students should ideally display higher clinical practice aptitude levels in their assessed work beyond that of novice (See Appendix 4). Fig 4.5 provides a summary of the findings related to Benner's five Novice to Expert criteria present in the sampled N=28 WRs.

Fig 4.5. A summary of Benner's Novice to Expert Model, Results.

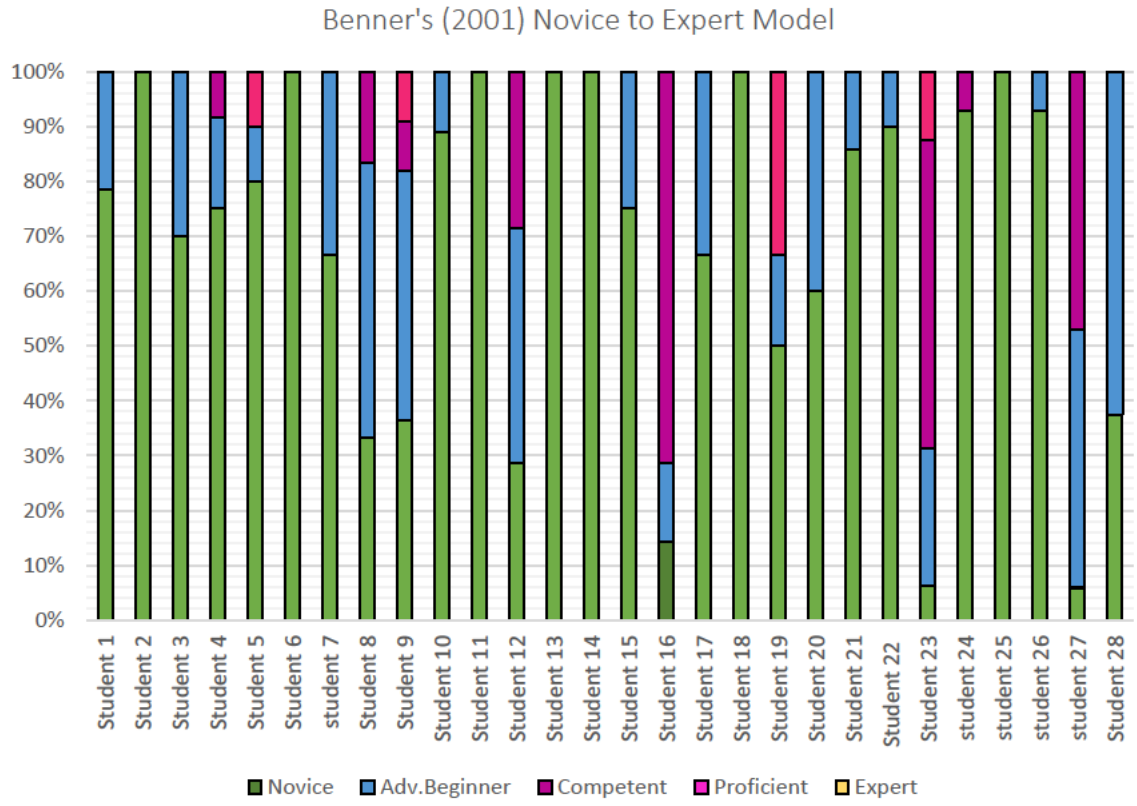


Fig 4.5. Summarises the five stages of the Novice to Expert criteria coded and identified across the N=28 sampled students' written reflections. This chart is gradients of 10%, representing sections of the student work, related to specific TDTA codes where Benners criteria are demonstrated.

4.5.1 Novice: Level 1

Novice level characteristics consist of context-free and rule-governed practice, universally applied without reference to the overall situation. Context-free consists of information processing relevant facts and clinical action, independent of whatever else is (clinically) going on. The novice practice is notoriously rule-governed because the rules determine any or all clinical actions. Novice performance has utility for undergraduate nurses with limited disciplinary knowledge and experience to override rule-governed practice and act proportionally to the clinical demand. Many core nursing competencies are transferable, and many nursing discipline-specialisations should already exist. For students studying for the NMP qualification, their pre-registration education (prior knowledge and experience) and postgraduate nursing practice exposure in areas such as wound care should allow them to move beyond guides, rules and well-versed skills within the prescribing role. Fig 4.5.1 identified the number of NMP students' [n=20] displaying high novice characteristics (% > 60), indicated that few of these students had progressed forward, transitioned toward competent or expert achievement.

Fig 4.5.1 Benner's Criterion Level 1: Results in Summary

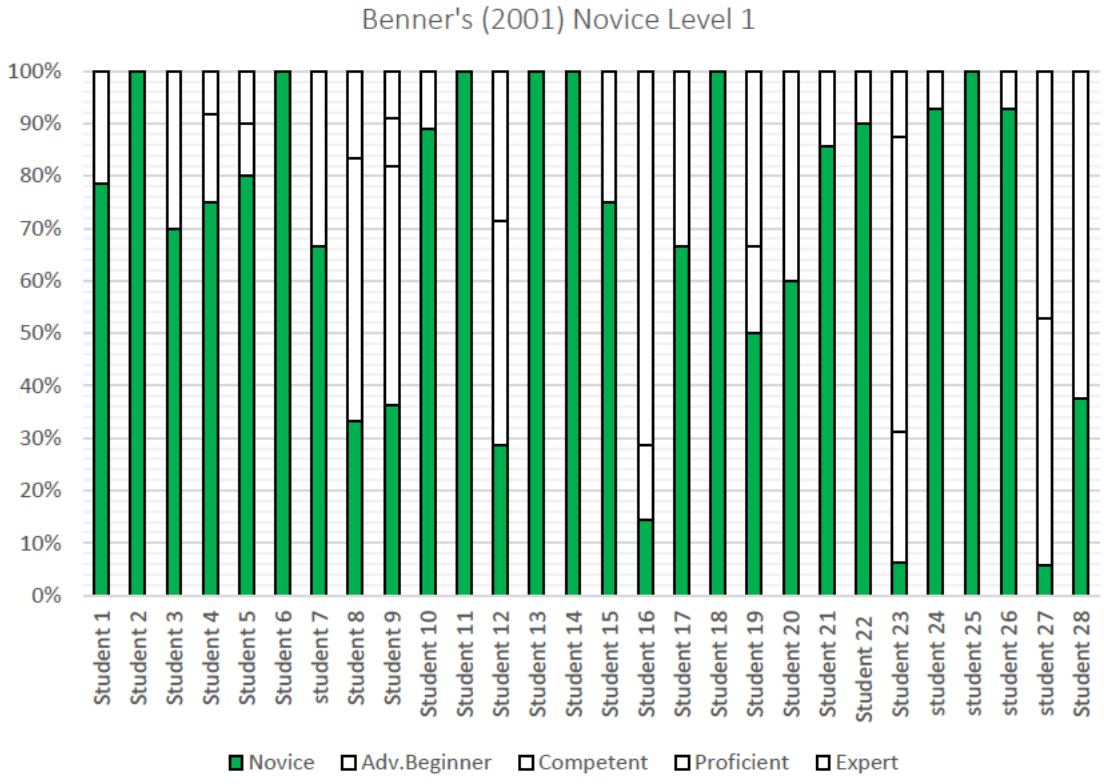


Fig 4.5.1 Illustrates the novice criterion found across the N=28 written reflections. The lowest novice criterion % ratings were St. 23 and 27 at 4% and the highest were St. 02, 06, 11, 13, 14, 18 and 25 at 100% respectively.

The implicit rule here is that these students are new to NMP, and therefore, they are novice practitioners. Habitual practice interprets the NMP event described in the WRs. However, the majority were overly reliant on rules to guide their actions. Without acknowledging through reflection, the underpinning knowledge elements that prelude judgement, critical thinking in action. For example, students 07 and 28 stated the following:

> **St. 07**
> "……., The guidelines go on to suggest…., wounds should be cleaned thoroughly prior to closure…, If a wound requires exploration or debridement,[18] then it should be anaesthetised[19]."

> Moreover, **St. 28**
> "…. I routinely refer to the seven principles of good prescribing advocated by National Prescribing Centre…. I have found this to be a safe invaluable tool for nurse prescribing."

This study acknowledged that students' in a new area of clinical practice might perform to a novice level. However, NMP students' neither practice nor prescribe outside of their clinical practise area, e.g., oncology nursing or area of expertise in exposure to NMP development. Therefore, within the speciality scope (i.e., oncology), the knowledge and skill should exceed novice. The expectation is that students would exhibit much higher practice acumen. NMP students should demonstrate the capacity to explore treatment, interventions and actions to critically examine why they took such actions or made such judgments and not others. Further, the NMP students should have considerable experience addressing many-core transferable clinical nursing elements, such as wound care and dispensing medicines, i.e., use of drug formularies. These concepts are generic learning to nursing and subsequently transferable to NMP care and practice.

4.5.1.1 Advanced Beginner: Level 2

Advanced Beginners (ABs) understand the recently learned textbook theory's contextual meaning and terms of practice. ABs have managed some real clinical situations of note due to their post-graduate status. However, the NMP student should draw on prior

SEE END NOTE
[18] Debridement
[19] Anaesthetised

experience and make conscious and deliberate situational application of the rules. Like procedural know-how, AB performance is marginally acceptable as students at this level can demonstrate acceptable performance and deliberate on and in the implementation of guidelines and rules. The ABs levels of clinical practice are context-free. Benner (2001) asserts that ABs are accountable to the rules and not the subsequent results where rules are applied. Fig 4.5.1.1 provides a summary of the AB criteria as present in n=20 WRs from the sampled N=28.

Fig 4.5.1.1 Benner's Criterion Level 2: Results in Summary

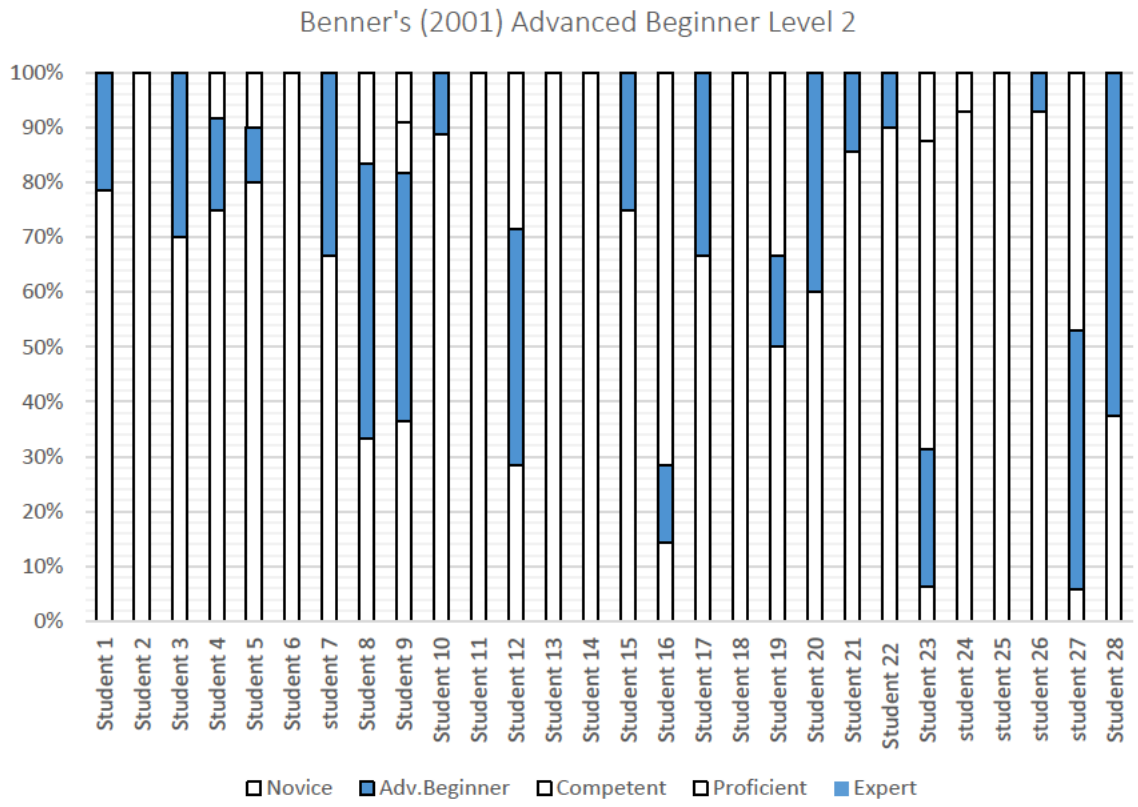


Fig 4.5.1.1 Illustrates the Advanced Beginner [AB] criterion found in n=20 of the sampled N=28 WRs. Where the AB criterion was present in the student's reflections, the criterion % ratings ranged from 8% in St.26 with the highest at 62% in St.28.

Advanced Beginners (AB) are just beginning to make meaningful associations using learned (or observed) clinical behaviours to direct their clinical practice. Alternatively, the significant associations formed are due to a perceived similarity to prior experiences. For example, the nature of that experience depends on the quality of the cumulative experience acquisition. The selection and interpretation of that experience and its

influence on the student's ability to perform fluently and effortlessly based on length of practice in nursing. The following excerpt from the NMP student (st.10) written reflection illustrates the AB characteristics.

St 10

"… When prescribing antibiotics, as an independent prescriber it is important to ensure that I use up to date information…. Local trust guidelines (2014) are in line with the government guidelines "Start smart, then focus" (ARHAI 2011)"…. "This means giving the right drug in the right dose at the right time for the right duration in every patient".

In contrast to the novice nurses' performance and behaviour, the ABs actions are task-driven about demonstrating procedural knowing. That the guidelines exist irrespective of content and know-how, the rules support or influence NMP practice. According to Benner's model, ABs have an increased sense of responsibility enacted through achieving goals by mastering a task- by routine referral to principles such as those described by St.10. The 'up to date' information referred to was not unfolded or integrated into the text to demonstrate the application in the evolving clinical situation described in the WR (see st.10 excerpt). It remained a 'task by routine' and a simplistic referral to being 'up to date. The 'up to date' evidence was not applied to uncover clinical knowledge development aspects or demonstrate how the evidence determines safe, effective NMP interventions. The high number of AB practices (n=20/ N=28) was an unexpected finding. See Fig 4.5.1. The NMP students work at AP status and study at post-registration masters level. These criteria lacked visibility in the WRs analysis.

4.5.1.2 Competence: Level 3

Characteristics of the competent nurse continue to follow the rules designed to assist them in working productively. However, a sense of what is situationally important may be absent. At this level, NMP nurses typically have 2-3 years of experience in a specific setting or specialist discipline. Rule application is context-driven, procedural, but pertinent to the situation. The nurse is conversant in when and where to seek expert advice. According to Benner, the competent practitioner can visualise the outcome for the patient. Based on similar previous events, practitioners' recognise the importance of the rules, is accountable to those rules but not their impact or the result. The criterion

competence found in n=8 of the sampled N=28 WRs illustrates that n= 8/N=28 was functional and procedural. See Fig 4.5.1.2. For example, St.04 in the excerpt discontinued the drug as instructed by the drug formulary without considering the consequence for the patient or prescribing alternative therapy.

Fig 4.5.1.2 Benner's Criterion Level 3: Results in Summary

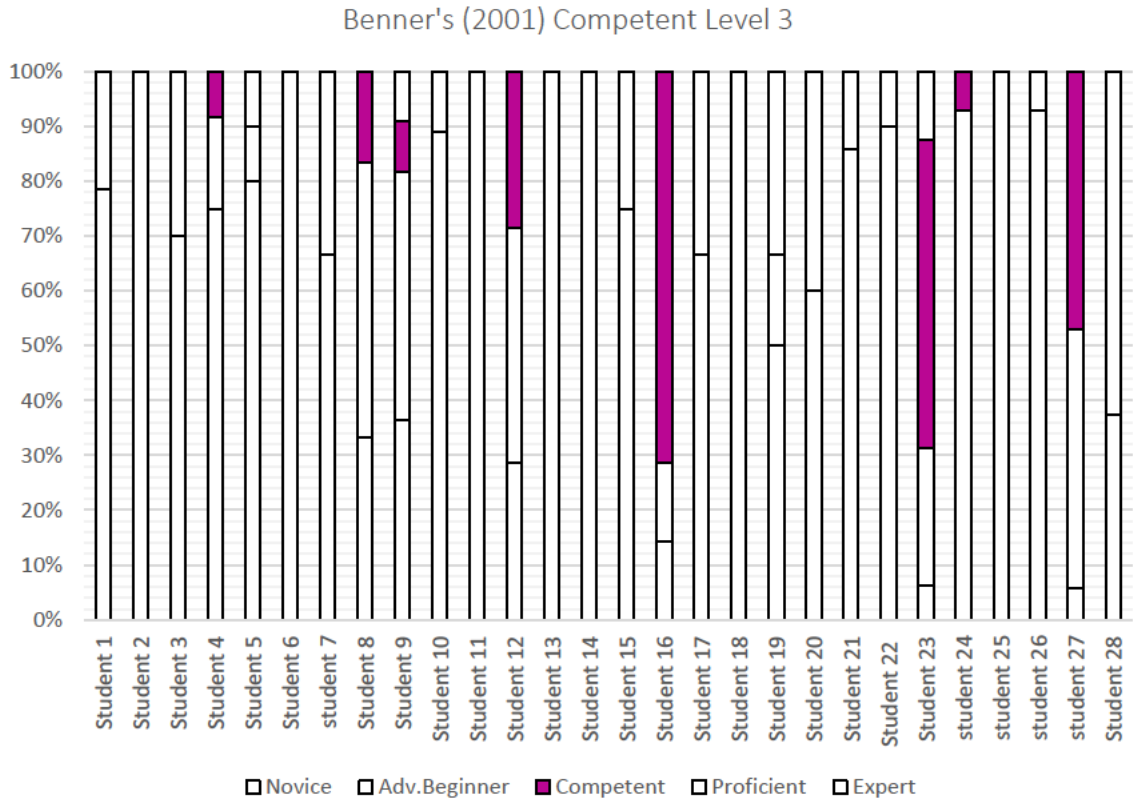


Fig 4.5.1.2. Illustrates Benner's competence criterion coded as present in n=8 of the sampled N=28 WRS. Where this competency coded was present in the text, the lowest Competence criterion % ratings were St. 24 at 6%, and the highest was St. 27 at 45% and St. 23 at 60% respectively.

Standards, policy, and protocol help practitioners perform non-conscious, non-contemplative (know-that, know-how) action. The problem addressed by rule application gives the NMP a linear perspective of the patient's condition. Rules and information are accepted without question. Typical to competent characteristics, there is accountability to the rules but not to the results. The following excerpts highlight how the plan is drawn from guidelines to direct and dictate which of the situational circumstances the student attends. However, there were significant, potentially life-threatening ones ignored.

St.04

"…. I took some time reading the BNF for advice… the most effective strategy for the management of drug allergy is avoidance or discontinuation the recommended treatment to give to a patient…"

Student 04 displays competence, including the classic non-context-driven action, in their management of the patient's allergic reaction (code 1), illustrated by engagement with a performed task. The student used the appropriate formulary (code 85) in discontinuing the medication, applied the information in the clinical context (code 78). Benner (2001) described this as skilled know-how, which de-locates integrated knowledge for practice. St 04's WR did not consider an alternative, replacement drug therapy to treat the patient's presenting condition (code 32). In this context, "the most effective strategy for managing drug allergy is avoidance or discontinuation". However, there are potential and significant consequences of stopping the drug concerning codes 07 the disease severity and code 99 (therapeutic) drug action. How this information concerning allergies shaped or influenced the prescribing decision and action (code 64) following the discontinuation of the is not evident in St.04's WR.

4.5.1.3 Proficient: Level 4

Benner's proficient nurse is characterised as thinking about what to do, triggered by their experience of similar events. Essential 'prioritised' features of a situation will stand out, and others will recede into the background for later consideration. Aspects of care are not compartmentalised or approached as tasked performance. Due to experience, the proficient nurse promptly addresses notable and significant clinical events combined with a clear rule like calculation, decisions made about how best to manage the situation. The proficient nurse is analytical of plans that worked in the past and modifies these plans, homing in on resolving the problems and responding to current events. Aspects of care are not compartmentalised or approached as tasked performance. Such nurses have an extensive repertoire of clinical experience to draw on. The approach to a situation is holistic, i.e., using a broader spectrum of clinical theory and knowledge for care planning, intervention management, and delivery. As shown in Fig 4.5.1.3, identifies the Proficiency criterion identified in n=4/N=28 students' WRs.

Fig 4.5.1.3 Benner's Criterion Level 4: Results in Summary

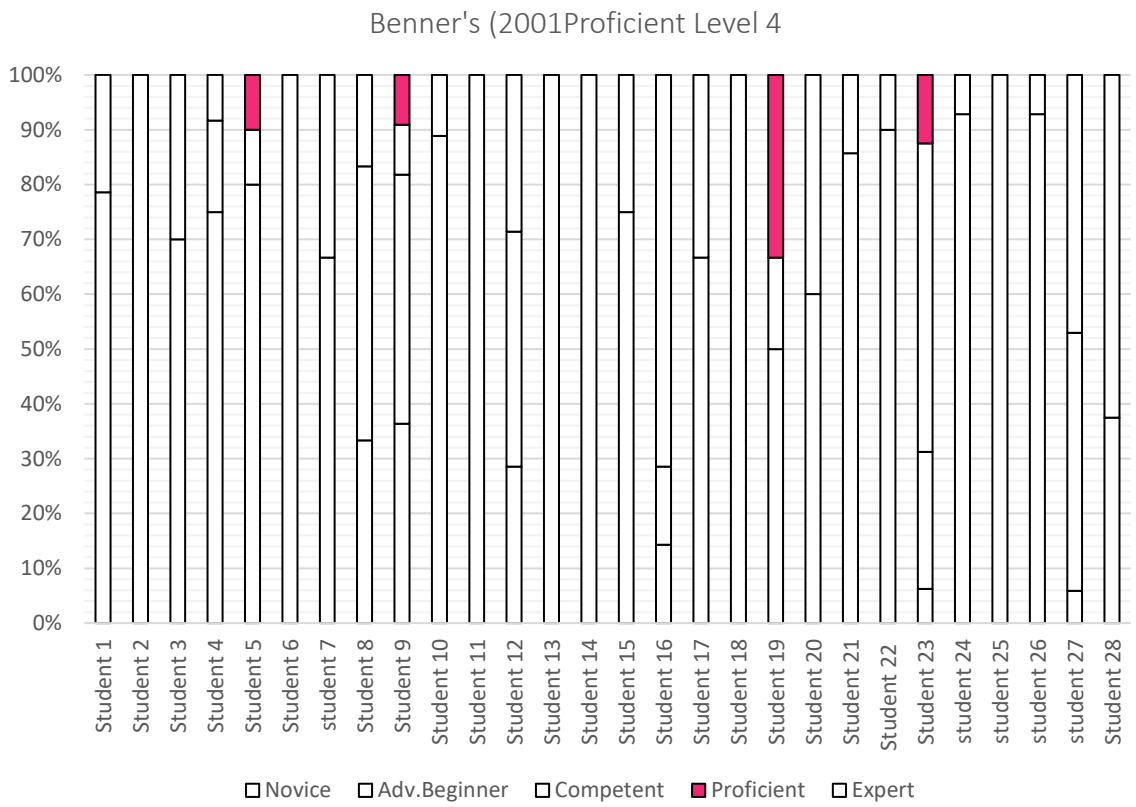


Fig 4.5.1.3 Illustrates the proficiency criterion represented in n=4 of the sampled N=28 WRs. Of the students whose work was coded as Proficiency, the lowest Proficiency criterion % ratings were St.05 and 09 at 10 % and marginally smaller 8% respectively; the highest was St. 19 at 32%.

For these n=4 students' qualities of Benner's proficient practitioner were distinct parts of a few WRs. In comparison to competence, proficiency demonstrates superior performance supported by knowledge application in clinical events. There is a global understanding of the patients evolving clinical case (code 22) based on knowledge and understanding of anatomy related to the PE (code 25). Along with experience to interpret the patients' disease severity (code 07) and changing clinical conditions. The excerpt from student 05 illustrated proficiency in observing, understanding, and interpreting the patient clinical status, situation, and the appropriateness of the response.

St.05 "…with 'Surviving sepsis[20] guidelines …… recommended for fluid resuscitation[21] in severe sepsis……. consideration was given to this patient with cardiac history" …. "The administration of Normal Saline 0.9%[22] will raise serum chloride[23] which in the already compromised……., the patient will experience metabolic acidosis, and this may lead to a poor outcome for the patient."

The proficient practitioner sees the relationships between aspects of the situation. For example, "the patient's cardiac history," and in assessing the patient's history, there are reflective thoughts about "recommended fluid resuscitation…." Unlike students 04 and 08 in section 4.5.1.1, student 05 identified the patient's biochemical response to the normal saline and discussed it in the context of the patient's confirmed elevated serum chloride. The adverse effect this may have on the patient is in line with the medication's precautions and contraindications (code 20). The application of their experiential knowledge to the salient features of the situation, i.e., "metabolic acidosis [24]". The subsequent "… a poor outcome for the patient…." demonstrates that the student has a perspective on which physiological aspects present are significant. These are addressed well in the following WR excerpt, where reflexivity on experiences is flexible and consciously used.

4.5.1.4 Expert: Level 5

Missing from the Sampled N=28 WRs is Benner's fifth criterion of expert practice.

4.6 Bloom's Taxonomy of Educational Objectives: The Cognitive Domain

Bloom's (1956) educational objectives is a structure that promotes mastery learning. These BT objectives depict the expected increase in cognitive complexity and in-depth learning. At the masters level study, the academic attainment requires higher cognitive skills evidence of more in-depth understanding and knowledge transfer to various NMP clinical duties and contexts. Bloom's taxonomy exhibits the six domain categories illustrating specific knowledge and skills acquisition gained through the

SEE END NOTE
[20] Sepsis
[21] Fluid Resuscitation
[22] Normal Saline
[23] Serum Chloride
[24] Metabolic Acidosis

programme of study. Fig 4.6 summarises the cognitive taxonomies codes' results as present in the sampled N=28 WRs.

Fig 4.6 Bloom's Taxonomy of Educational Objectives: The Cognitive Domain Results in Summary

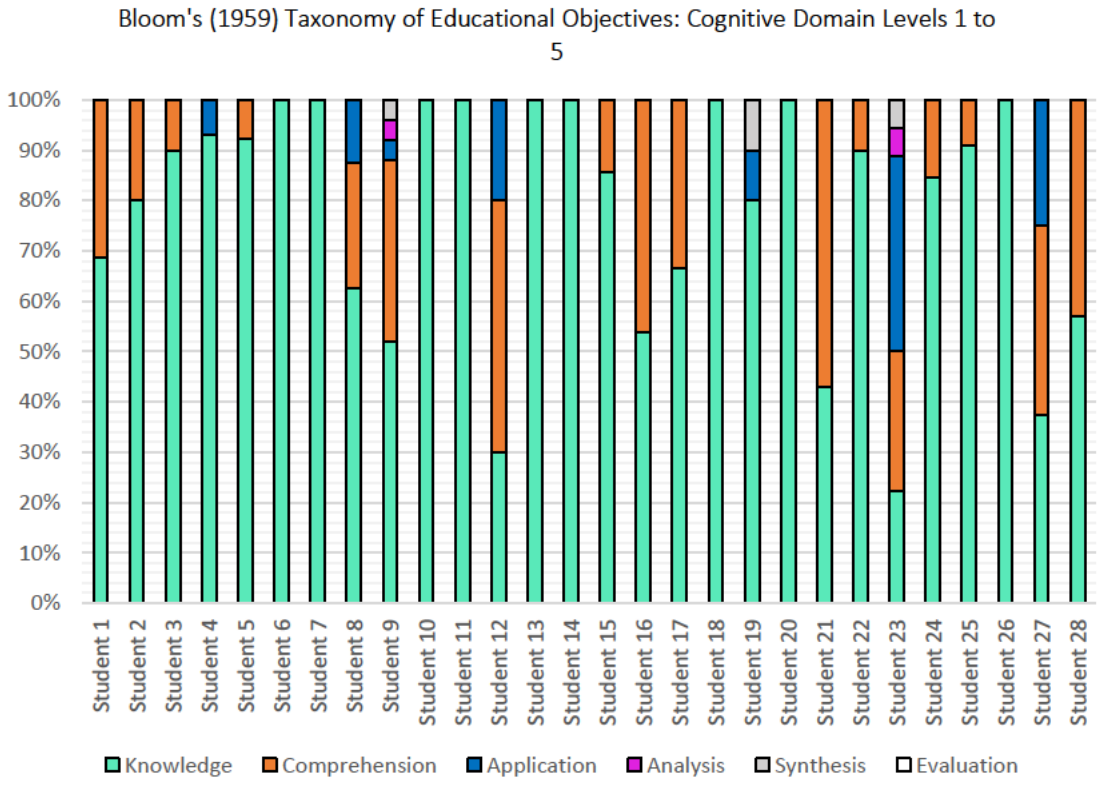


Fig 4.6 Illustrates the educational taxonomy's cognitive domain criterion, coded and as presented the sampled N=28 written reflections. This chart is gradients of 10%, representing sections of the student's work related to specific TDTA codes where Bloom's criteria are demonstrated.

4.6.1 Bloom's Cognitive Domain Taxonomy Level 1, Knowledge

The taxonomy of knowledge represents the central elements of mastery and what students must know to be soundly acquainted with in the specialist healthcare field. Cognisance may include facts about the test's name, medical or NMP terminology, the vocabulary or the language representative of the practice area appropriate for conversing in the clinical practice context or reading medical data. Fig 4.6.1 Illustrates Level 1, Knowledge attainment as represented by the sampled N=28 WRs.

Fig 4.6.1 Blooms Cognitive Domain of Knowledge: Results in Summary

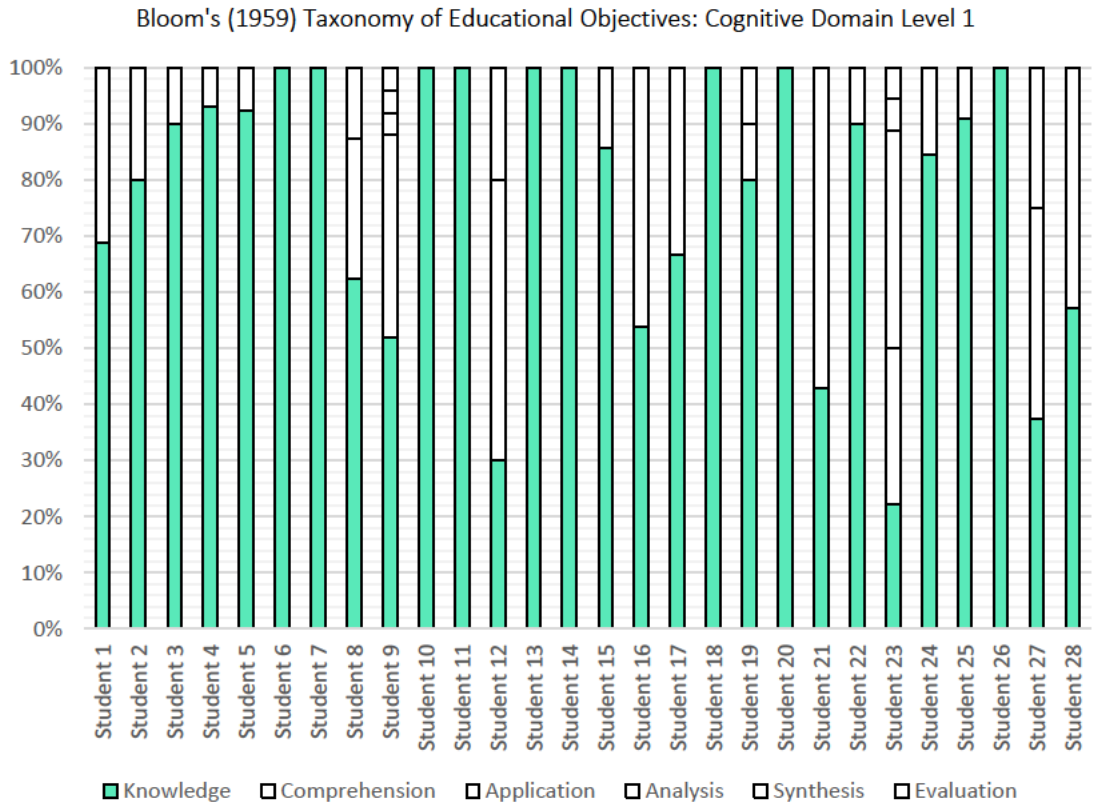


Fig 4.6.1 Illustrates the knowledge taxonomy found across the N=28 written reflections. Of the students whose work was coded as Knowledge attainment, the lowest % ratings were St. 23 the highest at 22% and were St. 04, 06, 07, 10, 11, 13, 14, 18, 20 and 26 at 100% respectively.

Knowledge may include facts about the test's name. However, this occurs in isolation, as separate discrete elements of declared knowledge. The tests are of limited value without detailed knowledge of the clinical rationale for the test. The test's rationale but crucial is the interpretation of those tests or results (code 06). For example, an X-ray will illustrate the state of bones or other conditions such as chest imagery, inflammation, or tissue damage. Interpreting the X-ray or other radiographic imaging (code 34) requires knowledge application synthesis and evaluation. However, interpreting the X-ray or other radiographic imaging (code 34) requires knowledge application, synthesis and evaluation. To confirm bone, tissue, or organ damage, arrive at a diagnosis (code 33) and prescribe appropriate treatment (code 32). The information is of limited value unless applied to the prescribing action.

The excerpt from St.22 is an example of the terminology or vocabulary representing the language of the discipline used to converse intelligently. Medical terminology such as 'purulent discharge', 'erythema' and 'phalanx of the hand...' is the universal vocabulary for healthcare practitioners to communicate with colleagues and express clinical information. Students can recall, recognise and use the terminology such as names of tests or the medical terminology but may not fully understand, as indicated by the declarative statements made in St.22's excerpt

St.22

Both patients' wounds showed no obvious purulent[25] discharge; however, the four cardinal signs of infection; erythema[26], pain, swelling and warmth…., were all present in their affected hand. Both patients were experiencing limited movement in the phalanx of the hand that was affected by the bite.

In the students' WRs coded for this taxonomy, there was no probing beyond recall. In places, meaningful interpretation of the information was lacking and, in some cases, inaccurate. For example, the significance of the assessment for infection, the implications of 'limited movement in the phalanx' (code 26), the differential or (code 33) the working diagnosis is not explained or explored in the MNP context.

4.6.1.1 Bloom's Cognitive Domain Taxonomy Level 2, Comprehension

Bloom's classification of comprehension represents an understanding of the communication's real message, requiring translation, interpretation and extrapolation. In translation, the student can put the information in other terms, usually giving meaning to the parts of what is communicated- often in isolation; the meaning determines the context in which the ideas appear. However, the information is, according to the definition akin to recall, declarative knowledge. Interpretation requires the student to think about the relative importance of the ideas presented, such as the context of reflection, the clinical relevance and interrelationships of disease or symptoms described within the text. In extrapolation, the student states what the clinical matter's truth is believed to be,

SEE END NOTE

[25] Purulent

[26] Erythema

including the consequences and the ramifications. At the same time, there is sensitivity to the factors which may change or render inaccurate their clinical predictions, thereby compromising patient safety. Fig 4.6.1.1 summarises Level 2 comprehension attainment as represented by the sampled N=28 WRs.

Fig 4.6.1.1 Blooms Cognitive Domain of Comprehension, Results in Summary

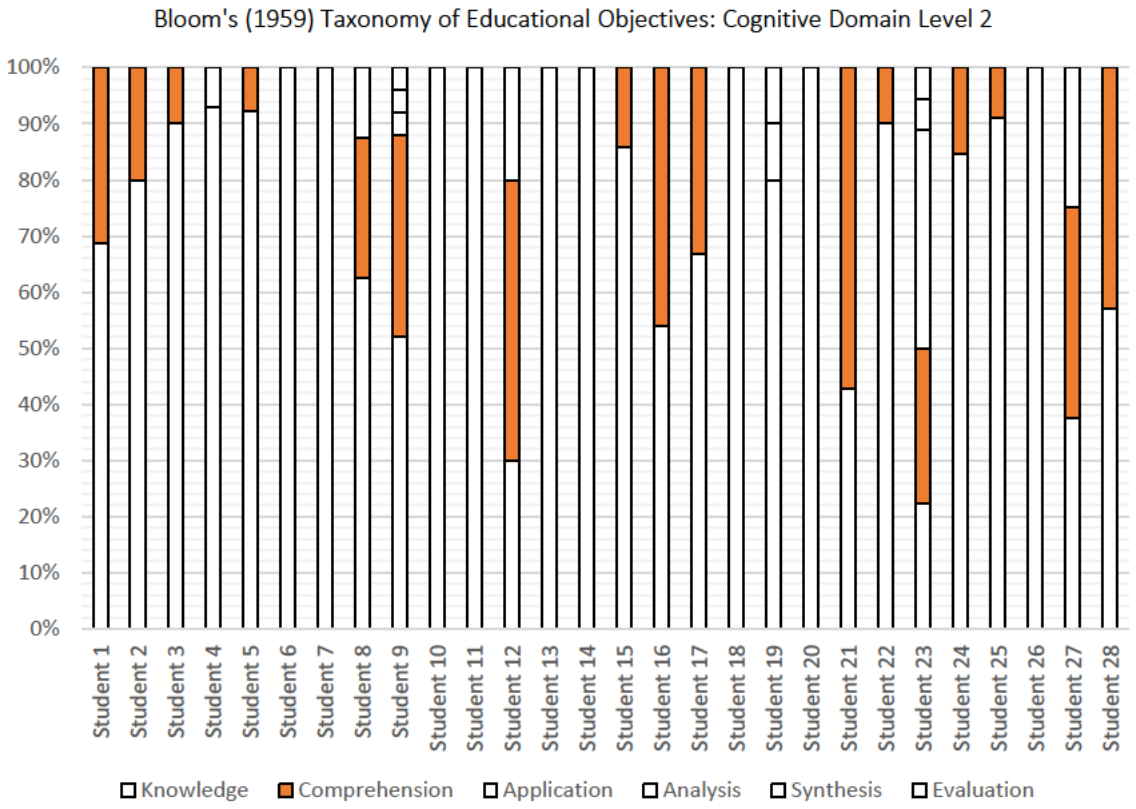


Fig 4.6.1.1 displays the comprehension taxonomy found across the N=28 written reflections. Of the students whose work was coded as comprehension [n=17] attainment, the lowest % ratings were St.5 at 8% and the highest was St. 21 at 55% respectively.

An individual's translation behaviour within their WR is in isolation. Reflective thought about the importance of the ideas presented is absent. The listing of clinical and medical terminology implies interpretation, i.e., thought about the relative importance of the ideas presented. Conversely, the pre-requisite or relevant knowledge upon which interpretation is dependent is missing in the WR of student 24. The excerpt exhibits the translation type of comprehension behaviour characterised in this cognitive domain taxonomy in the n=17 of N=28 sampled WRs.

St 24

…" In the minor injury's unit, I saw a patient that had inverted his ankle[27]. The patient inverted his ankle playing sport and had presented due to pain and swelling around the ankle"… On assessment he was tender over his AFTL[28] and CFL[29] but had no bony tenderness to his posterior Malleolus or distal aspect to his Fibula. He was X-rayed showing no fracture and I decided to give him an ankle brace for support, refer him to physiotherapy and advise regular Paracetamol.

Student 24 demonstrated translation at the level akin to textbook factual recall or Ryle's (1949) declarative knowledge. There is correct, appropriate and contextual use of medical terminology to establish and achieve interpretation. However, as a sub-component of comprehension, the student should denote meaning to each of the terminologies used, i.e., the tenderness 'over his AFTL and CFL' relative to the 'inverted ankle.' See endnotes for definitions. Alternatively, give brief but precise examples of their function and type of pain in the context of disease severity assessment (code 07), i.e., inflammation within the immediate or adjacent text, to demonstrate these concepts' comprehension. For example, the function of the various ligaments surrounding the ankle and forming part of the joint capsule surrounding and lubricate the joints (code 12). In the absence of a fracture, the student has correctly offered 'an ankle brace' as a non-medical treatment option (code 50) to immobilise and support the injury site, preventing further trauma. However, due to the physical symptom, the ankle swelling and subsequent pain do not correspond with codes 39 and 42, which address dealing sensitively with patients.

The essential characteristics of demonstrating interpretation are that students understand significant ideas and their clinical interrelationships in their WRs. For example, st.24 stated the ' X-Ray showed no fracture'. However, there was no reference to the evidence base for treating such injuries (code 81). Collectively missing from this WR were the related discussion about the alternative appropriate treatment options (code 32), i.e., ice packs, limb elevation, and prescribing practices relative to ankle inversion treatment. Or a reference to recommended non-steroidal anti-inflammatory (NSAID) drugs or topical creams (code 80). Suppose either paracetamol or NSAID pain-relieving agent are prescribed. In that case, a discussion of the effect, reason, and rationalisation for the presenting condition 'pain' is expected (codes 11 and 19). In this taxonomy's

SEE END NOTES

[27] Inverted Ankle

[28] Anterior Talofibular Ligament (AFTL)

[29] Calcaneofibular Ligament (CFL)

extrapolation component, the NMP student should detail the conclusions drawn from their prescribing actions and indicate the consequences and implications of the treatment suggested (code 48). The WR should give the reader a more precise indication that they are capable, proficient and safe in their practice.

4.6.1.2 Bloom's Taxonomy Cognitive Domain Level 3, Application

According to Bloom, previous knowledge is essential to this taxonomy. However, the knowledge possessed and recalled does not mean knowledge applied. Knowledge application is the ability to use learned material or theory in a different context or situation. Bloom's criterion application evidences students' achievement of theory to support NMP practice and for problem-solving.

Fig 4.6.1.2 Bloom's Cognitive Domain of Application, Results in Summary

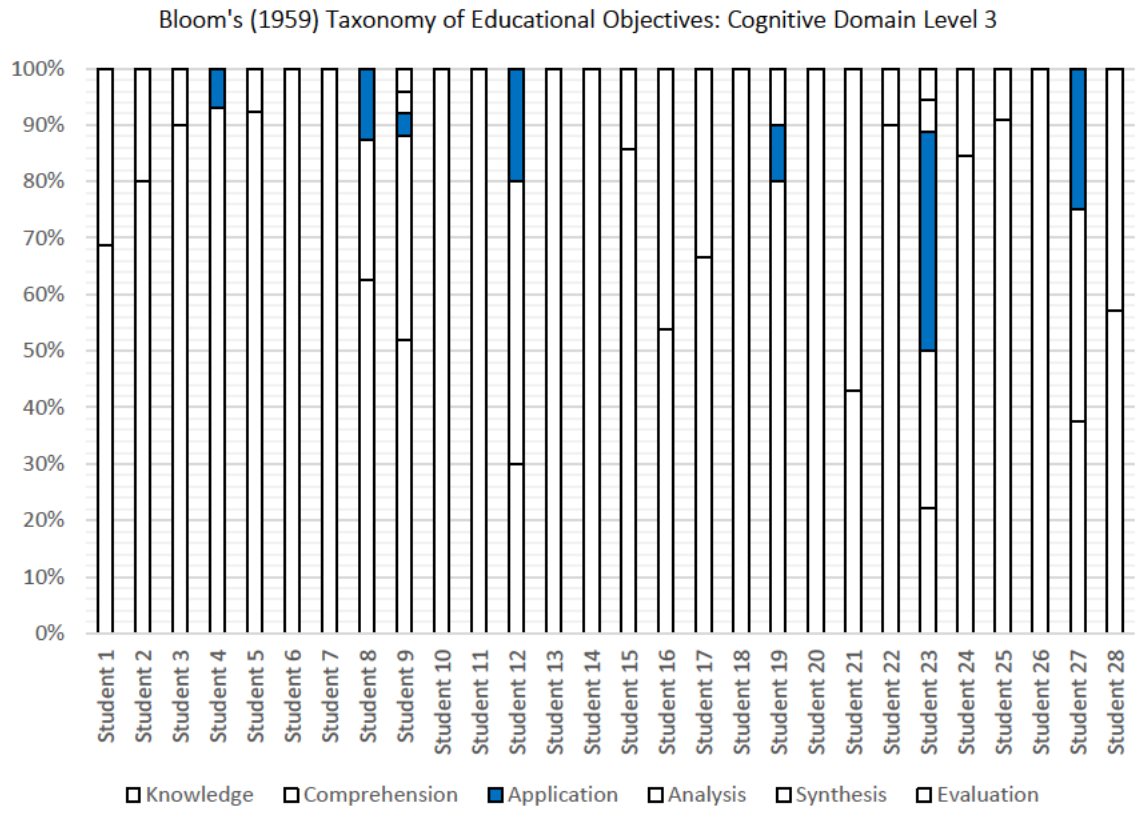


Fig 4.6.1.2. Displays the taxonomy 'Application' as found across the N=28 written reflections. Of the n=7 students whose work was coded as application attainment, the lowest % ratings were St.4 at 8% and the highest was St. 23 at 45% respectively.

Of the N=28 sampled written reflections, n=7 of the students presented examples of application in their prescribing practice; unfortunately, this level of use is not demonstrated consistently across the n=7 students. One-way evidence of application may occur directly relates to drug or product selection for prescribing. For example, drugs or products designed for primary use in specific diseases and for treatment purposes, such as found in the WR of St. 27.

St.27

"NICE guidelines……Theophylline as a next stage option for unstable COPD [30]….., a potent bronchodilator and this would have been an appropriate option if I felt Broncho-constriction[31] was the primary cause of the symptoms experienced. However, I felt that inflammation was the issue which is why I felt oral corticosteroids[32] where the most appropriate option…. a maintenance dose of prednisolone outweighed the risks."

The above excerpt is an example of an NMP theory application intentionally used for a clinical problem. The student has logically identified various related factors about theophylline to treat (or not to treat) the presenting unstable condition of Broncho-constriction. Not the primary cause of the COPD instability and therefore discarding Theophylline[33] as a treatment and therapy option (code 28). The decision not to decongest the chest to prevent further exacerbation of COPD led the student to conclude that Theophylline would not be helpful. As the cause of the patient's unstable condition was "inflammatory". In the potential risk and benefit assessment (code 48), the student opted to treat the patient with corticosteroids, the most appropriate treatment (code 32).

4.6.1.3 Bloom's Taxonomy Cognitive Domain Level 4, Analysis

This taxonomy emphasises grasping meaning and intent in knowledge-theory use for analysis of other relevant material. The analysis addresses why a solution works or identifies errors in logic to explain why a solution fails to work to resolve a problem. In presenting ideas, the student should demonstrate critical judgment to determine how well the concepts are interrelated and relate to the problem resolution argument. Fig 4.6.1.3 summarises the results for analysis as coded in the sampled N=28 written reflections.

SEE END NOTE

[30] Chronic Obstructive Pulmonary Disease (COPD)

[31] Broncho-constriction

[32] Corticosteroids

[33] Theophylline

Fig 4.6.1.3 Bloom's Cognitive Domain of Analysis, Results in Summary

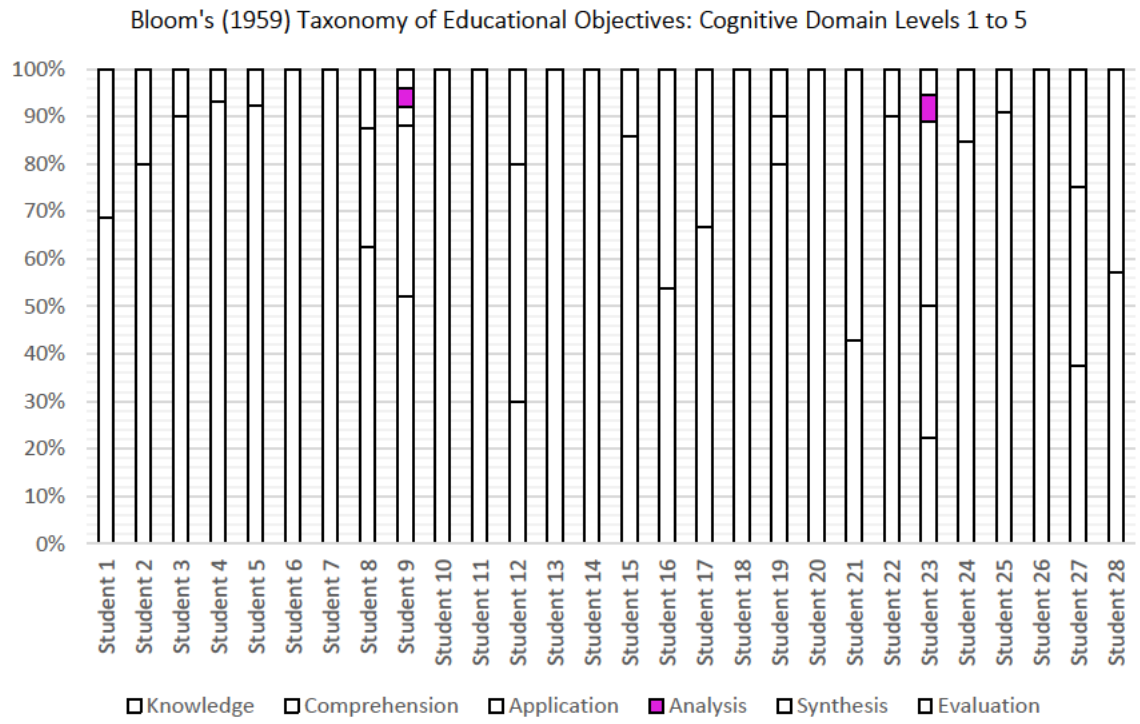


Fig 4.6.1.3. Summarises the results for the taxonomy application, as found across the N=28 WRs. Of the n=2 students whose work was coded for the attainment of analysis, the lowest % ratings were St.9 at 4%, and the highest was St. 23 at 6% respectively.

In the sampled N=28 written reflections, the overwhelming majority (98%) fell short of demonstrating this taxonomy achievement. The n=2 Students who demonstrate this taxonomy achievement went beyond knowledge application and further utilised theory to analyse the problem. See, for example, the excerpt from student 09.

St 09 "...in the treatment of hysterical patients. Monoamine oxidase[34] (MAO) enzymes can break down many nitrogenous molecules present during the ingestion of food, as a result, inhibits the monoamine which can cause a build-up of tyramine[35] an amine found in certain foods like fish, cheese etc. which may cause a dramatic increase in blood pressure.... adding another potential long-term condition. With this in mind, it is too complex and pointless in managing the current risks presented ...to the patient...however its beneficial sedative benefits are noted. 10mg Diazepam[36] three times per day for one week only and review.... this is short term....., due to their highly addictive nature."

SEE END NOTE

[34] Monoamine oxidase (MAO)

[35] Tyramine

[36] Diazepam

In achieving Bloom's analysis criterion, students actively look for relationships between terminology, i.e. MAO inhibitors interpret their function in increasing the BP. The student may need to recall the specific facts about MAO. Still, analysis requires the student to examine the cause-and-effect relationships (code 27). The student analysed the clinical observation, the increased BP, and the theory concerning MAO to propose the possible causes being the MAO's inhibiting effect, determining this as the best reasoning based on the clinical recordings (code 11). However, the critical missing factor about NMP is the pathophysiologic[37] effect in MAO inhibition. Specifically, the build-up of tyramine, the causative agent. Its impact, vasoconstriction leading to an elevated BP' and the subsequent implications to prescribing action (codes 19 and 21).

4.6.1.5 Bloom's Taxonomy Cognitive Domain Level 5, Synthesis

The Level 5, Synthesis was not found in the N=28 sampled WRs.

4.6.1.6 Bloom's Taxonomy Cognitive Domain Level 6, Evaluation

The level 6 taxonomy, Evaluation, was not found in the N=28 sampled WRs.

The discussion chapter 6 applies Kane's Interpretive Validity Argument (IVA) as a framework to describe and illustrate the IVA procedure. The main aim is to provide evidence concerning the inferences and claims made about students based on the assessed score.

SEE END NOTE
[37] Pathophysiologic

CHAPTER 5: THE QUANTITATIVE DATA ANALYSIS OF THE STRUCTURED CLINICAL EXAMINATIONS (SCE)

5.0 Chapter Overview

Chapter 4 presented the content analyses of the summative written reflections and discussion of the results of the levels of reflection and achievement in Ryle's, Benner and Bloom's theoretical concepts, introduced in chapter 1. Chapter 5 presents the analysis and discussion of the SCEs criterion validity measurement, the item difficulty and student ability results following the Rasch 1PL model's application.

5.1 Introduction

Kane (1990; 1994a) argues- that in test scores used to make conclusions about professionals' readiness for practice, the validity question should ask what exactly the test has measured. The validity premise here is to provide credible supporting evidence of the scores used to indicate professionals' readiness for practice. Or show what may be wrong with the Structured Clinical Examination (SCE) to demonstrate invalid use. This chapter addresses Criterion-Referenced (CR) measurement using the psychometric analysis to determine validity in research sub-question 2.

> Sub-question [2]. To What Degree Do Structured Clinical Examinations [SCEs] Accurately Represent and Measure the Criteria they are Designed to Measure?

5.1.1 The Sampled SCE

The sample consisted of N=28 SCE exam papers containing 42 test items, grouped into the following three NMP domains: (1) Consultation skills and history taking, (2) Decision-making and rationale and (3) legal prescription writing. Subject Domain 1 contains twelve test criteria, one of which is mandatory, an essential criterion for awarding a pass-score on test items. Domain 2 includes fifteen test criteria, eight of which are mandatory, and domain three also contains fifteen test criteria, all of which are mandatory. The test criterion in domains 1 and 2 included oral response options within the marking to allow

the examiner to question the examinee to elicit additional answers to the test criteria. The examiner awards marks against predefined criteria under the headings 'stated', 'incomplete answer', 'not stated, or 'incorrect' scored as a (dichotomous) Pass/Fail. For the SCE test item, see Appendix 5.

5.1.2 Goodness of Fit Results

The Bootstrap test was 0.74 (p< 0.001). The statistically insignificant p-value indicated that the 1PL model best fitted the SCE data, Tables 5.1. 2 to 5.1.3.2 present the analyses' results. Table 5.1.2 for M2, the unconstrained 1PL where no imposed restrictions on the fixed or random parameters were the best fit (Field et al., 2013). The item fit for comparing more than two IRT models was established using a likelihood ratio test. See methods chapter 3 for discussion.

Table 5.1.2. The Goodness-of-Fit Results for the Non-mandatory SCE Questions for the N=28 Sampled

Model	BIC	Comparison	LRT (p-value)
M1 - Rasch [1PL] Constrained	523.3	M1 / M2	18.41 (<0.001)
M2 - Rasch [1PL] Unconstrained	508.8		<0.001
M3 - Two parameter	545.77	M2 / M3	29.54 (0.03)
M4 - Three parameter	573.43	M2 / M4	5.78 (0.997)

To account for model fit, the Bayesian Information Criterion (BIC), a statistical model, was applied to account for model fit and parameter relevance measurement (e.g., difficulty and discrimination). The smallest BIC value makes the model selection. In the Table 5.1.2 example, the lowest BIC of 508.8 suggested the Rasch model unconstrained was appropriate (von Davier and Carstensen, 2007). Table 5.1.2.1 similarly, with the non-mandatory questions in Table 5.1.2.2, the unconstrained 1PL Rasch model best fits the data. The lowest BIC overall is 272.48.

Table 5.1.2.1 The Goodness-of-Fit Results for the Mandatory SCE Questions for the N=28 Sampled

Model	**BIC**	**Comparison**	**LRT (p-value)**
M1 - Rasch [1PL] Constrained	278.73	M1 / M2	10.16 (0.001)
M2 - Rasch [1PL] Unconstrained	272.48		
M3 - Two parameter	294.48	M2 / M3	8 (0.309)
M4 - Three parameter	303.17	M2 / M4	9 (0.874)

The p-value for the bootstrap test was 0.435, not statistically significant at <0.5. Meaning that the data modelling and chosen model was appropriate, confirming that the 1PL model best fitted the SCE data. Table 5.1.2.2, as with the non-mandatory and mandatory questions, the unconstrained 1PL Rasch model provided the best with the lowest BIC of 921.85.

Table 5.1.2.2. The Goodness-of-Fit Results for All SCE Test Questions for the N=28 Sampled

Model	BIC	Comparison	LRT (p-value)
M1 - Rasch [1PL] Constrained	934.83	M1 / M2	16.89 (<0.001)
M2 - Rasch [1PL] Unconstrained	921.85	NA	NA
M3 - Two parameter	1016.62	M2 / M3	65.63 (0.009)
M4 - Three parameter	1047.27	M2 / M4	38.89 (0.608)

The p-value for the bootstrap test was 0.945, confirming that the 1PL model fits the SCE test item criteria in all the data measured.

5.2 The Analyses Results: The Mandatory SCE Test Item Criteria

The SCE criterion was dichotomously scored (pass=1/fail=0). Therefore, test item measurement provided evidence on the difficulty or ease of the SCE criterion and whether the SCE can safely be presumed to represent the NMP (proficiency) attributes in practice. The NMP student's ability or KSP level on the NMP program is an unknown quantity. It is also unknown whether the SCE test is at the appropriate DIF level, i.e., for AP-level or mastery at which SCE is supposedly set. The mandatory and non-mandatory SCE criteria listed in Tables 5.2 to 5.2.3.1.2 were analysed using the 1PL Rasch model. In determining SCE criterion validity, the researcher evaluated examinee ability and the test item parameter for difficulty estimates in answer to the validity question:

> What is the likelihood (probability) of an examinee with a given ability correctly responding to a test item with a given difficulty?

The tabulated SCE difficulty values and results are illustrated graphically and discussed in the following text. According to psychometric theory, the DIF index $P\ (x)\ value$ in Table 5.2, or ease index describes the % of students who correctly answer the test items. These values range from 0-100%, and the higher the percentage, the easier the answer.

Table 5.2 Mandatory SCE Test Items - Decision Making and Rationale for the N=28 sampled SCEs

	Question (Pass/ Fail)	Difficulty [β_i]	$P(x = 1 \mid z = 0)$
Qm 4	Gains consent	-2.73351	0.996837
Qm_13	Has reached a decision, including consideration of any potential differential diagnoses	-1.52313	0.961043
Qm 14	Recognises need for further examination/ /investigation	-2.03005	0.986244
Qm_15	Has considered the practical aspects of the treatment plan, whether this is suitable for the patient	-1.52292	0.961026
Qm 17	Has identified possible cause/causes	-1.82771	0.979094
Qm_19	Recognises the limits of personal competence and makes recommendation for referral where appropriate	-1.39907	0.950001
Qm_20	Has considered the relevance of all other treatments/ Medication the patient has already taken or is taking	-1.39955	0.950049
Qm 21	Has considered the relevance of patents medical history	-1.66352	0.970717
Qm_22	Has considered possible cautions/contraindications or previous ADRs	-1.66299	0.970685
		Discrimination parameter 0.10 to +1: value = unable to discriminate	

Table 5.2 The equations are explained as follows. X means response (X) made by the examinee to item (*i*).
β_i refers to the difficulty of the item. X = *1* refers to a 'correct' response to an item.
P is the probability of examinees' correct response to *i* .
P (X =1) is predicted by student ability using the logistic regression
z =0 is a selection filter for which terms to report, i.e., include everything between 0-100%
(Embretson and Reise, 2009).

The information in table 5.2 highlights that students correctly answered a high proportion of the mandatory NMP decision-making and rationale questions. The ICC in Fig 5.2 depicts the difficulty calibrations of test items presented in Table 5.2. Primarily the relationship between ability and item DIF is situated at -1.3 to -1.8 for Qm-13, Qm-15, Qm-17, Qm-19, Qm-20, Qm-21 Qm-22, a 95%-97% probability of the examinee zero logit ability being correct on these test items. Qm-4 and Qm-14 are at DIF levels of -2.73 and -2.03, indicating an even lower DIF of 98% - 99% probability of examinees' with the limited ability (i.e., registering at zero logits and below) giving a correct answer on these test items.

Fig 5.2 illustrates that all the test items were shifted to the left of zero ability, indicating that the test items in this data set are, in general, of low DIF level. Therefore, persons with ability at zero logits will have a high probability of passing the test item. The test items of note are Qm-4 and Qm-19. The ICC for Qm-4 climbed sooner < 0.2 on the Y-axis with the inflexion curve at -.3.0. The uppermost part curve is approximated at 0.89 (89%), almost 90% probability on the Y-axis. SCE item Qm-19 was the most difficult. However, the ICC inflexion is situated at -1.39, relatively low on the ability scale. Although marginally better than Qm-4, with a p-value of 0.99 (99%), item Qm-19 still had a high probability of 0.95 (95 %) of being answered correctly by examinees' with an ability (or theta=Θ) of zero logits.

Fig 5.2 Mandatory SCE Criterion: ICC Decision Making and Rationale for the N=28 sampled

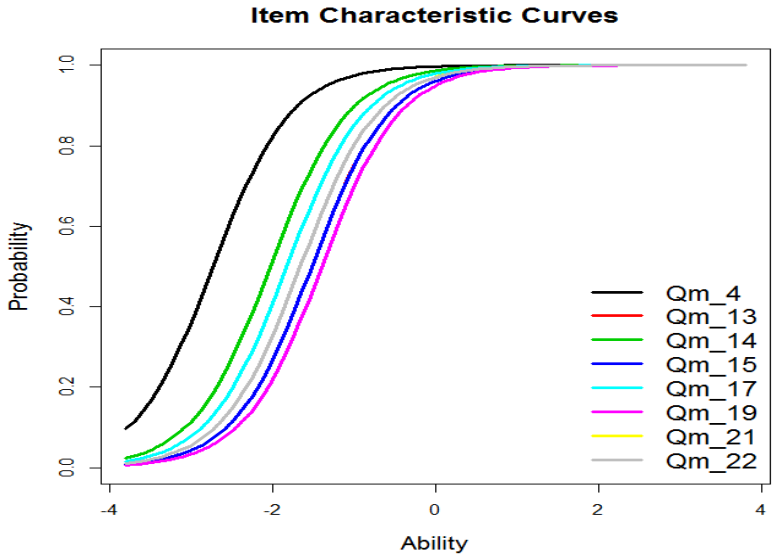


Fig 5.2 The ICC compares individuals on the same test item. These ICCs are very close together and some are obscured (yellow, green)

Test Qm-4 was the easiest, followed by Qm-14, Qm-17, Qm-22, and Qm-15, with Qm-19 the most difficult in this data set. The ICC for Qm-15 obscures items Qm-13, suggesting that they have the same DIF estimates of -1.52 on the ability scale. These values equal a probability of 0.96 (96%) of correct response in examines with zero ability on the logit scale. The same is true of ICC Qm-21, similarly, obscured by the ICC for Qm-22; again, suggesting the same DIF parameter estimates, which are positioned equally at -1.66 on the ability continuum. Therefore, there is a probability of 0.97 (97%) of examinees' with zero (logit) ability to answer these test items correctly. As ICCs represent the relationship between ability and DIF on given test items. As several test items tap into the same ability range and overlap, the ICC obscure the other (Doucette and Wolf, 2009).

Fig 5.2.1. IIF for Decision Making and Rationale in the N=28 Sampled SCEs

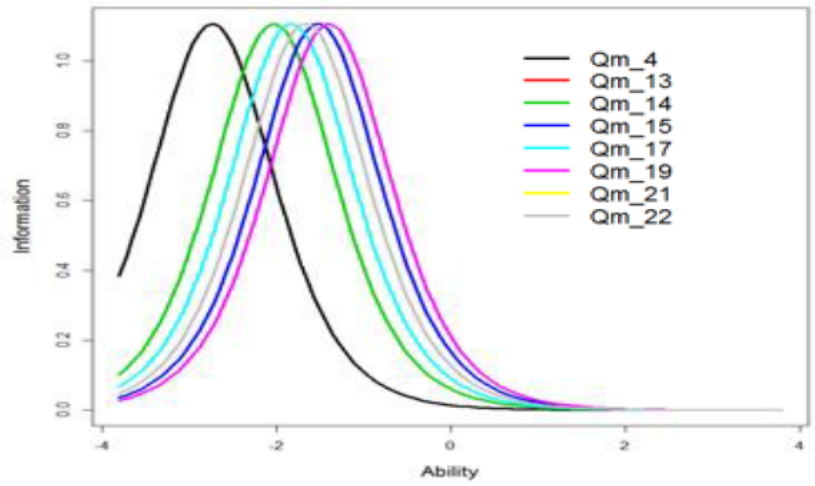


Fig 5.2.1 The height of the IIF corresponds with the ICC information in Fig 5.2.
There is an extreme shift to the left of zero logits, indicating that the ICC is correct

In Fig 5.2.1, the IIF is consistent with the ICC measurement of low difficulty values in this data set. The IIF for Qm-4, Qm-14, Qm-17, Qm-22, Qm-15 and Qm-19 reached maximum height approximated at 0.89 on the Y-axis within a narrow locus on the logit scale. The IIF estimate is consistent with IIF rubrics; different test items have provided varying amounts of information. The graphs are tall and narrow at 1.9 on the Y-axis, corresponding to measurement sensitivity (accuracy) regarding the test items low DIF parameter in determining the examinee location, i.e. a lesser ability of -1.30 to -3 along the X-axis.

Fig 5. 2.1.1 TIF for Decision Making and Rationale in N=28 Sampled SCEs

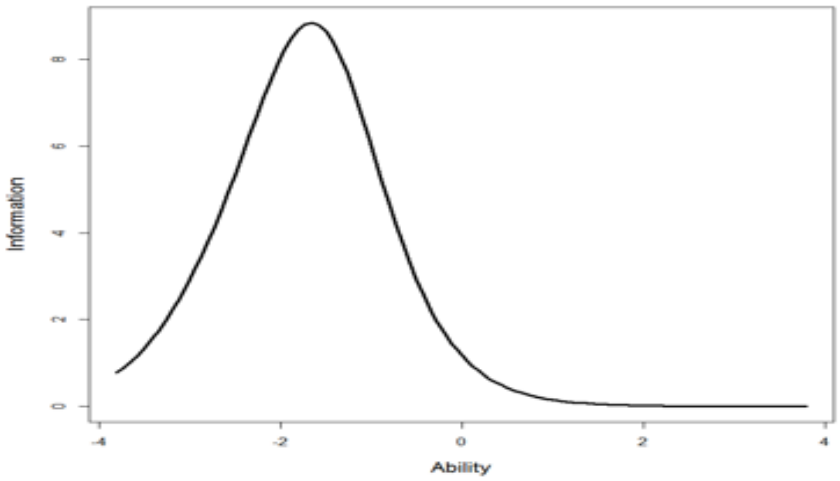


Fig 5.2.1.1, the test information function (TIF]) summarises the IIF and is pictorially similar. TIF is peaked at 8.9 within a narrow locus of the low ability values of -1.8 to -2.7 along the X- axis. This TIF measurement is, therefore, considered a reasonable estimate of precision measurement of low examinee ability and reflecting that the test items were easy across the test items in this data set.

5.3 Legal Prescription Writing

Essential to meeting the legal requirement for writing prescriptions is that the delivery instructions are clear about what, when and how of product distribution. Therefore, prescriptions should be legible (Nuttall and Rutt-Howard, 2011). The literature reports a persistently high rate and frequency of drug prescription errors due to writing illegibility, dose and route ambiguity. Practitioners' not writing prescriptions according to the formulary or the omission of the prescriber's signature and date (Kings Fund, 2013). All of which singular or combined can seriously compromise patient safety (NPSA, 2007a; Calligaris et al., 2009; Kings Fund, 2013). Table 5.3. contains SCE items about the legal requirement for prescription writing.

Table 5.3 Mandatory SCE Test Items- Prescription Writing for the N=28 sampled

Question	(Pass/Fail)	Difficulty $[\beta_i]$	$P(x = 1\|z = 0)$
pQm_28	Patient Name	-13.1539	1
pQm_29	Patient Address	-13.1539	1
pQm_32	Name of drug /product	-13.1539	1
pQm_33	Appropriate drug/product	-13.1539	1
pQm_36	Dose	-13.1539	1
pQm_37	Frequency	-13.1539	1
pQm_39	Directions for use	-13.1539	1
pQm_40	Signature	-13.1539	1
pQm_41	Date	-13.1539	1
			Discrimination parameter 0.10 to +1: value = unable to discriminate

In the SCE questions listed in Table 5.3, the students (N=28) scored correctly across all these test items. The item DIF was -13.1539 on the logit scale across this data set. Each test item had a probability of 1 (100%) of a correct response by examinees' with zero (logit) ability. As the DIF and ability values across this data set are identical, the resultant ICCs placed all the test items on a single curve. The test items in Table 5.3 make an accumulative contribution to the IIF and, therefore, the TIF estimation of examinee ability against the DIF measurement. Although acceptable for TIF, the IIF would also provide the item information on a single curve. The data in Table 5.4, the IIF was of limited use in terms of the item information function's purpose due to the extreme negative values (Doucette and Wolf, 2009; Bond and Fox, 2015). The ICC, IIF and TIF have produce graphs at the extreme left of person-ability located left of zero. Entirely off the logit scale at -13.15, indicating that these are undemanding test items relative to mastery learning.

Table 5.3 contains the minimum information required on a written prescription for complying with the legal and universally accepted prescribing requirement. Inaccurate, incorrect or missing information about one or more of these test items is considered a prescription error (WHO, 1994; Aronson, 2009).

Table 5.3.1. Mandatory SCE Criterion- Prescription Writing Cont'd for the N=28 sampled

	Question (Pass/Fail)	Difficulty $[\beta_i]$	$P(x = 1\|z = 0)$
pQm_30	Age	-2.20309	0.98845
pQm_31	Date of Birth- Only legally required if child under 12 years	-2.81652	0.943562
pQm_34	Contact details	-3.14766	0.958817
pQm_35	Quantity	-4.33553	0.987074
pQm_38	Duration of treatment	-4.33553	0.987074
pQm_42	Blank space underscored	-4.33553	0.987074
			Discrimination parameter 0.10 to +1: value = unable to discriminate

The ICC in figure 5.3.1, although still expressing test item parameter values, was atypical and did not have the expected logit formulaic S-shape. The flattened uppermost part of the ICC and the natural inflexion at the ability level is atypical compared to other data sets across these test items. This difference is evident when compared to the ideal example in Fig 5.3.1.1 in the previous section of this chapter, possibly because prescription writing presents a limited or even insignificant challenge to nurses in general. More importantly, as general nurses and, more significantly, APs have the advantage of exposure to prescriptions through years of experience dispensing medication to patients. Diligence over incorrectly written prescriptions is necessary for practice safe medicines administration as it prevents nurses from legally dispensing medications (NMC 2010a, NMC, the Code, 2015). Therefore, the NMP students have a sound comprehension of what constitutes a correctly written prescription. These test items present a very low (possibly insignificant) practical purpose or cognitive challenge.

DIF [β_i] values in the ICCs for this data set shifted left of zero, set at <-2.2 and below. Thus, confirming both low test item difficulty and limited examinee ability is required to pass these test items, which changed the ICC inflexion and shape (Doucette and Wolf, 2009). Despite the ICC's altered form, the IRT function of locating ability along the X-axis was unchanged; examinee ability is at -2.81 for pQm-31. With a 0.94 (94%) probability of correct response to this test item in persons with zero ability on the measurement scale. Item pQm-34 had a DIF of -3.14; an examinee with zero logit ability had a probability of 0.95 (95%) of passing this test item. Test questions pQm-35, pQm-38 and pQm-42 had

identical person ability locations along the X-axis and probability values of 0.98 (98%) of examinees' with zero logit ability passing these test items as illustrated in Table 5.4.1.

As the parameter values are identical, the ICC for pQm-42 displaced the ICCs for pQm-35, pQm-38 in Fig 5.3.1. There is an almost flat horizontal line for pQm-42, indicative of incredibly easy, a defectively functioning test item or both. Note that the DIF values are identical at -4.33553 across these test items, presenting no real challenge to qualified nurses. Particularly APs with extensive experience of reading, interpreting and, in some cases, troubleshooting the content of written prescriptions. The negative -2.20 logit DIF indicates that these examinees' ability surpassed that of the set standard, with a high probability of 94% to 98% correct response to the test item (Meyer, 2014).

Fig 5.3.1. ICC for Mandatory prescription writing in N=28 Sampled SCEs

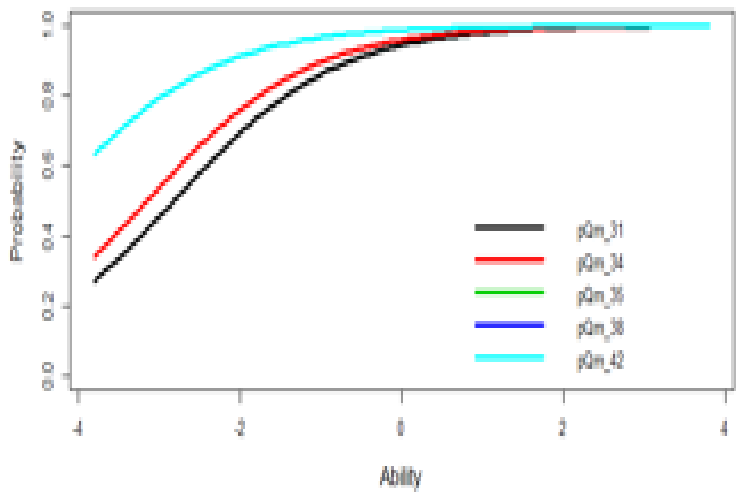


Fig 5.3.1. shows a flattened ICC with a left shift (-2) indicating very easy items. The flatness of the curve suggests that these test- items do not discriminate among students with high, moderate or low ability. Test items Q36 and 38 are displaced by Q42 as the results are the same.

For the test items in Table 5.3.1, it is natural to presume high proficiency levels among the examinees'. However, these test items in this data set are low in the DIF parameters compared to student ability. The rationale for this low DIF level is that these students are conversant with written prescriptions. Due to frequent exposure to prescription errors and resolving the prescription error problems in their nurse's role, the test items propose to measure. These students, as qualified nurses, have significant exposure to prescriptions

and prescription errors. These test items measure no more than prior experience and memorised knowledge of the procedures to troubleshoot prescription errors.

Fig 5.3.1.1 IIF for Mandatory prescription writing in the N=28 Sampled SCEs

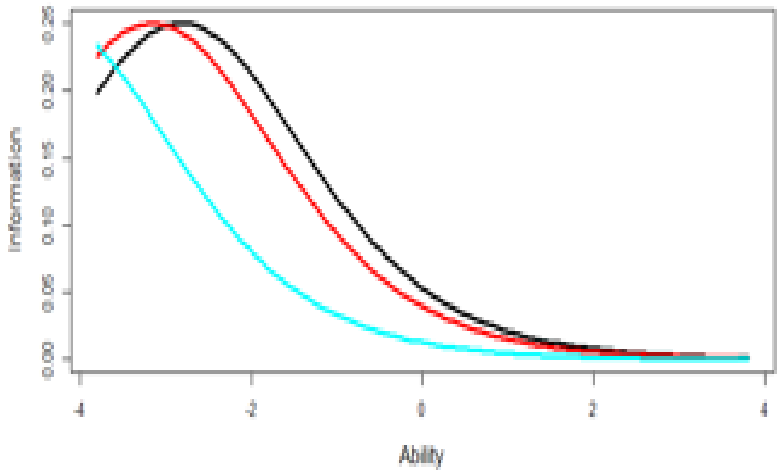


Fig 5.3.1.1 The IIF had good precision measurement and was correlated with the ICC information. A good test item is 0.10 to +1 or above; these items are -2 to – 4; a low level of difficulty.

The IIF in Fig 5.3.1.1 shows only the curve's upper asymptote. The IIF is outside the logit scale, which ends at -4.0 logits, beyond a useful DIF range for the test to represent any purposeful practice challenge. Extending the X-axis to include a broader range of -0.0 to ∞ ability, the upper curve at 1.9 on the Y-axis would make the values visible (Doucette and Wolf, 2009). The IIF had a substantive shift to the left of zero (ability) along the X-axis, consistent with Fig 5.4.1.1 and the ICC measurement of -2. 30 to -4.33 ability on the measurement scale.

Fig 5.3.1.2 TIF for Mandatory prescription writing in the N=28 Sampled SCEs

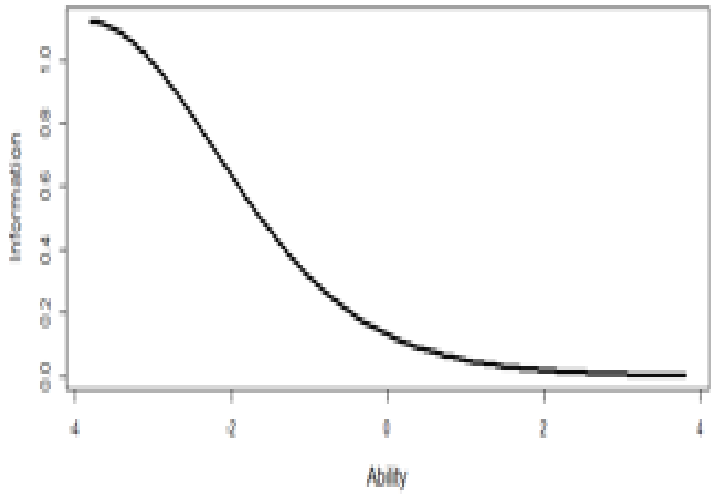


Fig 5.3.1.2 The TIF shows that the lower asymptote is off the logit scale due to a large amount of β_i (the *D* parameter) information located along the X-axis, at -4 and beyond the low person-ability levels.

Consistent with the ICC, IIF findings displayed in Fig 5.3.1.1 and TIF in Fig 5.3.1.2, the examinee's ability was at -2. 30 to -4.33 logits, and the N=28 students had a 94-98% probability of answering the test items in this data set. The IIF and the TIF results placed this data set outside the range of test item usefulness. I.e. very low on the DIF parameter, indicating the need for test item removal or improvement of the test items in this data set.

5.4 Generic SCE Test Items: Decision Making and Rationale

The NMPs clinical practices and the decisions made must be substantiated by applying the best available evidence about therapeutic effects and efficient use of relevant clinical literature (Sackett et al., 1996; Vitale and Exley, 2016). The general SCE test items are the universal qualities of desired decision-making skills in prescribing expected in NMPs to underpin patient safety and quality care requirements (Latter, 2008; The Kings Fund, 2017). Although not defined as mandatory, the examinee is required to achieve a pass-score. Either pass or fail without compensation in these test items to achieve an overall pass score on the SCE test entirely. Table 5.4. depicts the numerical values for an item DIF across this data set.

Table 5.4. The Generic SCE Test items- Decision Making and Rationale for the N=28 Sampled SCEs

	Question (Pass/Fail)	Difficulty [β_i]	$P(x = 1\|z = 0)$
Q_23 -	Makes treatment choice based on cost and clinical effectiveness	-1.58339	0.978521
Q_24	Explains the potential side effects and provides information to the patient	-1.33757	0.961804
Q_25	Uses Source of information available, appropriately (NPF/ BNF)	-1.91901	0.990325
Q_26	Identifies that the patient has a central role in decision-making	-1.04456	0.925489
Q_27 -	Has recognised the need to ensure the patient understands the information given	-1.1353	0.939244
	Discrimination parameter 0.10 to +1: value = unable to discriminate		

Table 5.4 locates these test items parameters at -1.04 to -1.91 on the DIF parameter measurement. Students with zero logits ability had a 0.92 (92%) to 0.99 (99%) probability of correctly answering these test items. The DIF values displayed in the ICCs in Fig 5.4 confirmed that Q-26 was the most difficult of this data set. The DIF value was at -1.04, the closest to zero logits of all the test items in this data set. Examinee's with zero logit ability had a 0.9 (92%) probability of correctly responding to these test items. Item Q-25 was the easiest; the DIF parameter is at -1.91; the chance was 0.99 (99%) for an examinee with zero logit ability responding correctly to this test item. Test items Q-23, Q-24 and Q-27 presented no challenge to these NMP students, with β_i values at -1.58, -1.33 and -1.13 in that order. An examinee with zero logit ability correctly responding to these test items were 0.97 (97%), 0.96 (96%) and 0.93 (93%), respectively. The best test item is of this data set was Q-25, albeit an effortless item.

Fig 5.4 The Item Characteristics (ICC): Decision Making and Rationale for the N=28 Sampled SCEs

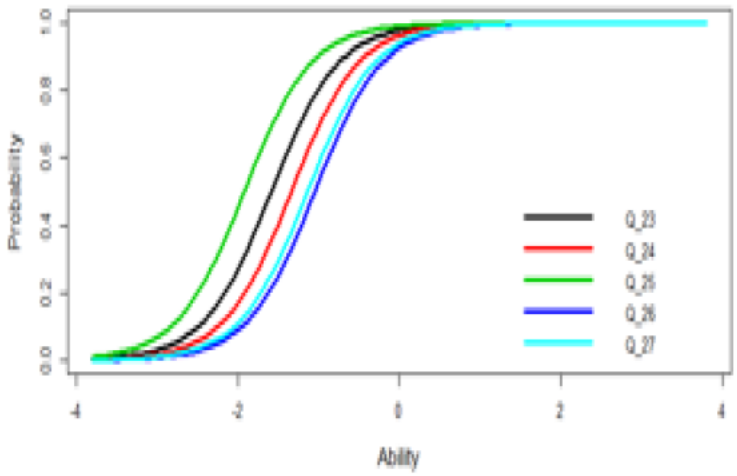


Fig 5.4 ICC, the slope of the curve of this ICC is well below the zero logits and flattened indicating a high probability of a correct response. These test items are easy

In Fig 5.4.1, the IIFs uppermost curve (peak) is a maximum of 1.4 on the Y-axis, with a narrow focus along the X-axis of -1.1 to -1.9 (ability) logits. Test items Q-23, Q-24 and Q-27 IIFs were visible, with β_i values at -1.58, -1.33 and -1.13 shifted left of the logit scale, indicating these test items were also unchallenging. In test Q-25, the maximum peak is -1.9, close to -2.0 logits on the ability scale - the easiest of the test items in this data set. There is evidence of measurement precision in IIF for these test items.

Fig 5.4.1 Item Information Function (IIF): Decision Making and Rationale for the N=28 Sampled SCEs

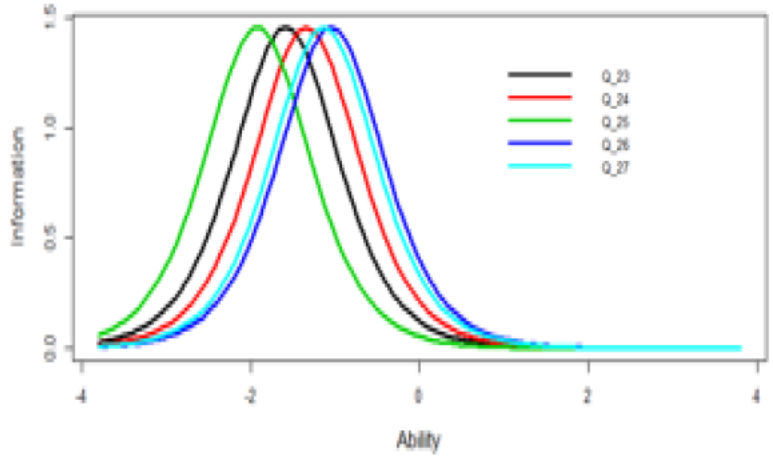


Fig 5.4.1 The IIF indicated that these test items had an overall low ability threshold indicated by the peak of the curve at between -1.1 to -2; this corresponds with the ICC

Fig 5.4.1.1 illustrates the TIF's uppermost curve (peak) at a maximum of 6.4 on the Y-axis with a narrow focus along the X-axis of -1.1 to -2.0 ability. Representing how well the individual test items functioned, collectively capturing the SCE's overall design and function. For this dataset, the examinee ability and test DIF overall were low. The TIF indicated sensitivity in measurement, and the information is congruent with the ICC information and the graphical IIF representation.

Fig 5.4.1.1 SCE Test Information Function (TIF): Decision Making and Rationale for the N=28 Sampled

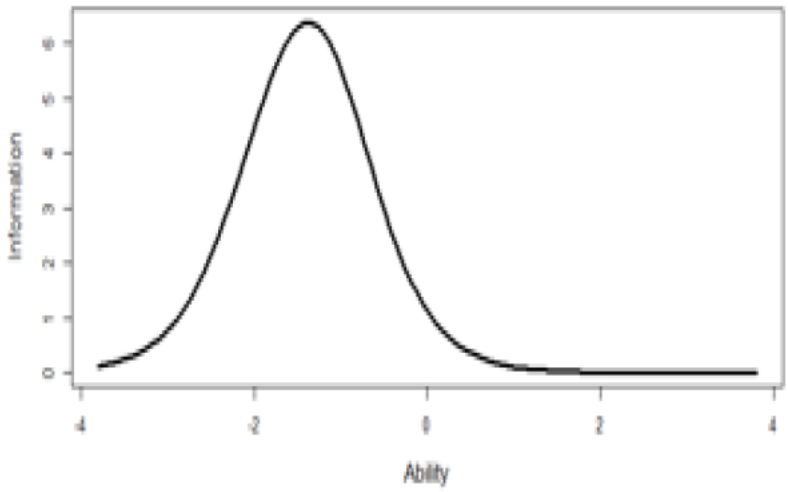


Fig 5.4.1.1 The TIF peak corresponds with the IIF in Fig 5.4.1, with the peak of the curve at between -1.1 to -2; indicating the precision of measurement in the test items and confirming the overall low ability threshold.

5.5. Generic SCE Test items Continued: Consultation and History Taking

Consultation and History-taking feature significantly in the NMP assessment processes. These SCE test items in Table 5.5 are associated with patient-centeredness and the consultation skills of eliciting and giving information (Silverston, 2013). Q-1, 'Greets patient', Q-2 'introduces self and role', and Q-3 'identifies the reason for patient attendance', had a DIF parameter identically positioned at -11.45 on the X-axis. The probability of a correct response to these test items from a student with zero ability on the logit scale was 1 (100%). With the DIF parameters at -11.45, the test information for these items was off the logit scale. These values again indicate simple test items.

Table 5.5 The Generic SCE Test Items- Consultation and History Taking N=28 sampled SCEs

Question	Pass/Fail	Difficulty $[\beta_i]$	$P(x = 1\|z = 0)$
Q_5 -	Identifies patients concerns and expectations	-1.7312	0.986798
Q_6 -	Uses Appropriate verbal/non-verbal communication skills to Elicit information	-1.44162	0.973208
Q_7	Seeks clarification appropriately checks patients understanding	-1.44161	0.973207
Q_8	Demonstrates respect for patient /client views	-1.57536	0.980655
	Discrimination parameter 0.10 to +1: value = unable to discriminate		

Table 5.5 showed that test items Q-6 and Q-7 are nearer to the DIF parameter estimate, with Q-6 at -1.44162 and Q-7 at -1.44161. Question 8 was the most difficult, with the DIF value at -1.57 and closest to zero ability on the logit continuum, closely followed by Q-5. The difficulty estimate was at -1.73. Overall, there is a 97% to 98% probability of students with zero ability to respond correctly to these test items. In Fig 5.5, concerning test items Q-6 and Q-7, the more difficult of the two is Q-7. Test item Q-7 is graphically placed nearest to zero logits at -1.44. Examinees' with zero logit ability had a 0.97 (97%) probability of correctly answering this test item. For Q-5 and Q-8, the likelihood of a correct response is 0.98 (98%) in examinees' with zero ability on the logit continuum. This test items DIF parameter for Q-5 is at -1.73 and for Q-8 at -1.57 on the X-axis; persons with low ability would answer these questions with ease.

Fig 5.5 ICC for test items Generic SCE Test Items for the N=28 Sampled SCEs

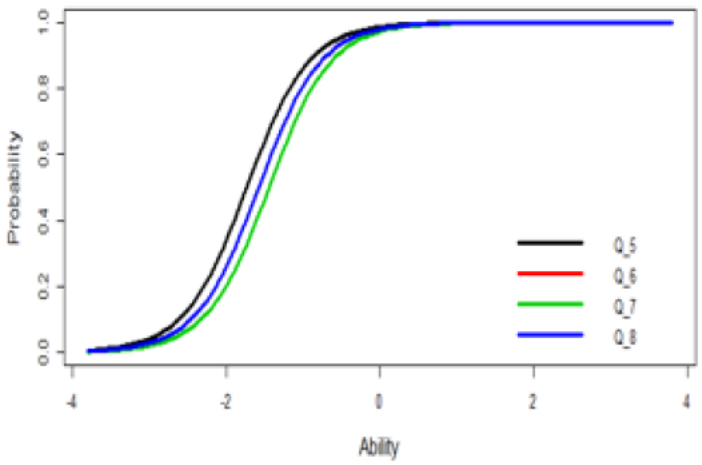


Fig 5.5. The ICC test items are close to identical and do not display difficulty values, in the results at -1.4 to -1.7 (logits) ability. Q6 is hidden as the ability values are a replica of Q7.

The IIF in Fig 5.5.1 provides information parallel to the ICC's in Fig 5.5. The ICC peak for Q-5 is at -1.73 ability on the logit scale was graphically depicted as the furthest away from zero ability on the X-axis, followed by Q-8 located at -1.57 ability on the X-axis. Test item Q-7 with a DIF value of -1.44 ability on the logits scale was the test item nearest to zero logits, corresponding to the ICC information. Therefore, the most difficult of this data set. The IIF indicated the reliability of the ICC information as it was consistent with the ICC information values.

Fig 5.5.1 IIF for test items Generic SCE Test Items in N=28 Sampled SCEs

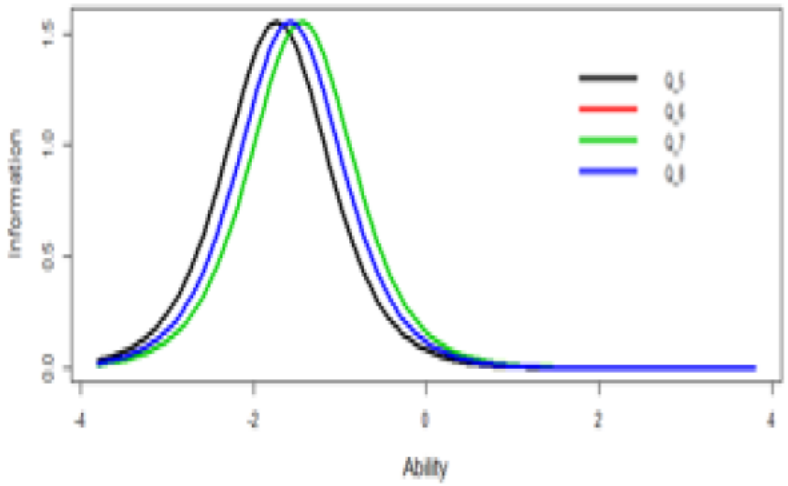


Fig 5.5.1 IIF has a very narrow focus, mirroring that of the ICC indicating low item difficulty. Q6 IIF remains hidden as the ability values replicate those of Q7.

Fig 5.5.1 illustrates that IIFs uppermost curve (peak) is at its maximum at 6.0 on the Y-axis with a narrow focus along the X-axis of -1.44 to -1.73 ability. The TIF in Fig 5.5.1.1 graphically agrees with the ICC information in Fig 5.5 and the IIF in Fig 5.5.1, which confirmed that these test items are of low DIF levels.

Fig 5.5.1.1 TIF for Generic SCE Test Items for N=28 Sampled SCEs

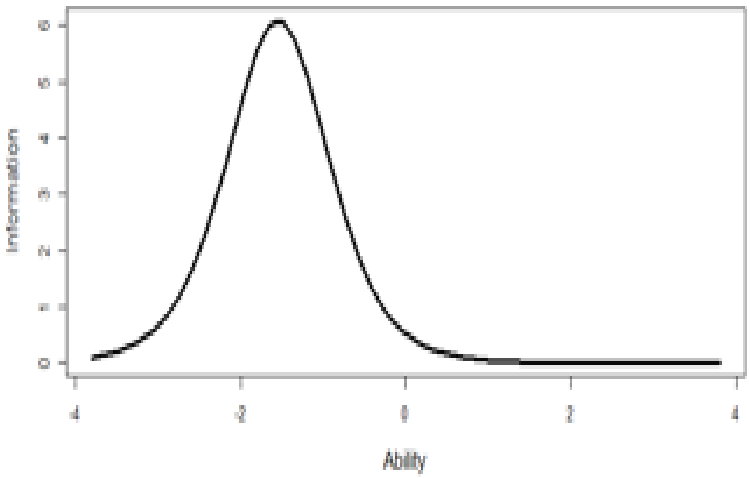


Fig 5.5.1.1 TIF shows its peak at the lower end of the logit scale -1.4 to -1.7, Synchronous with the IIF; at low person ability levels indicating measurement precision.

5.6 Generic Test Items- Consultation and History Taking Continued

Table 5.6 for test items Q-9 to Q-18 of the Consultation and History Taking domain identified that these test items suited examine ability below zero on the logit scale. The overall DIF values were at -1.66 and lower at -2.89. The ICCs for all these test items were consistent with the information in Table 5.4.3. The test items shift left of zero logits on the ability scale, specifying that all the test items were simple. The uppermost slope of the ICC curves for this data set is flat. The inflexion locates the position of person-ability on the logit scale. However, the ICCs for all the test items are skewed left of student ability on the logit scale X-axis. Indicating the test items in this data had a low DIF parameter measurement. See Figure 5.6.

Table 5.6. Generic Test Items- Consultation and History Taking Continued in N=28 Sampled SCEs

Question	(Pass/Fail)	Difficulty $[\theta_i]$	$P(x = 1\|z = 0)$
Q_9	Demonstrates efficient and logical approach to information gathering	-2.89029	0.962139
Q_10	Uses skills of Negotiation to achieve concordance	-2.6129	0.931688
Q_11	Demonstrates awareness of the ability of the patient to concord with a treatment plan	-1.98998	0.879741
Q_12	Closes consultation properly	-3.23524	0.962139
Q_16	Identifies further information required for holistic care	-2.17331	0.897827
Q_18	Makes treatment choice based on an agreement with the patient	-1.66922	0.841472
	Discrimination parameter 0.10 to +1: value = unable to discriminate		

In Fig 5.6, there is a flattening of the ICC's uppermost curves on all test items in Table 5.6. Moreover, the normal position of the inflexion point is flattened. The ICC shows a bias in the DIF value, favouring examinees' with low ability. Students' in general with this category will do exceedingly well on these test items. In Fig 5.6, the ICCs in Q-18 is the most challenging test item with a graphically illustrated DIF value of -1.66. Consistent with the datum on Table 5.6, the ICCs for the test items in this data set are positioned representatively to the left of zero along the logit continuum.

For Q-18, the probability of a correct response from examinees' with low capability is 0.84 (84%). Compared to the rest of the test items in this data set, this suggested that some examinees' had not done well with this test item. See chapter 6 for discussion. Question -12 was the easiest with a DIF parameter of -3.25 and a probability = 0.96 (96%) for examinees with zero (logit) ability correctly responding to this item. Test items Q-9, Q-10, Q-11 and Q-16 were also easy with DIF values in the order of -2.89, -2.61, -1.98 and -2.17 and probabilities of 0.94 (94%) 0.93 (93%) 0.87 (87%) and 0.89 (89%) respectively. An examinee with zero ability will correctly respond to these test items. The best functioning SCE question was Q-18 followed by Q-16, although both were undemanding test items.

Fig 5.6, ICC: Generic Test Items- Consultation and History Taking Cont'd in N=28 Sampled SCEs

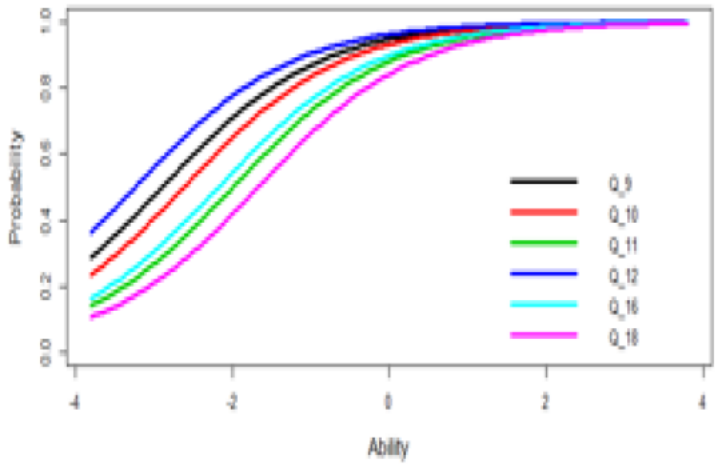


Fig 5.6. the ICC is flattened and at a low (ability, with values falling left of zero at – 1.6 to -3.2; and falling at – 4 logits. illustrating that irrespective of ability individuals will get these test items correct.

In the ICC in Fig 5.6, test item Q-12 was the easiest in this data set. The ICC curve maximised at 0.25 on the Y-axis. The IIF in Fig 5.6.1 correctly locates the difficulty value at -3.25 ability along the X-axis logit continuum, consistent with the numerical data on Table 5.6. Test Items Q-9, Q-10, Q-11 and Q-16, located on the logit continuum with correspondingly high IIF curves at -3.23 (= Q-9), -2.61 (= Q-10), -1.98 (= Q-11), and -2.17 (= Q-16) along the X-axis. Test item Q-18 curve peaks at -1.66. These items IIF is the closest to zero logits of all the test items, appropriately locating pinpointing where examinee ability resides along the logit continuum.

Fig 5.6.1 IIF for Generic Test Items- Consultation and History Taking Cont'd in N=28 Sampled SCEs

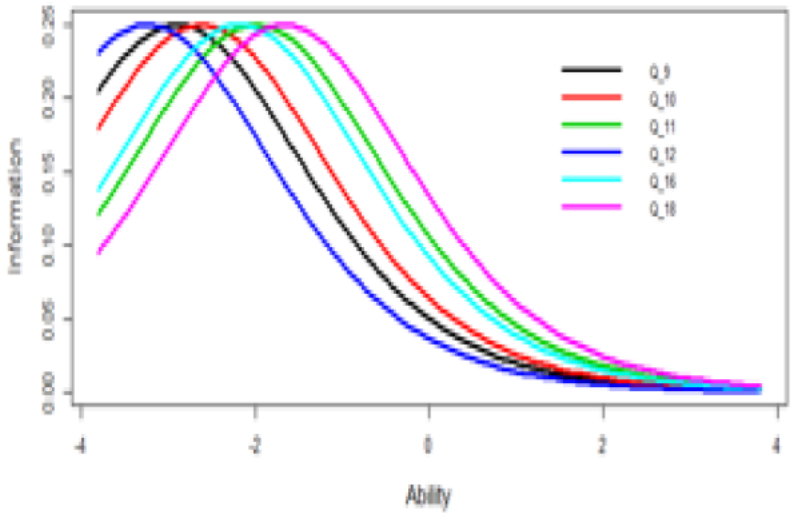


Fig 5.6.1 The IIF demonstrated at –1.6 to -3.2 and falling at –4 logits there is a good degree of sensitivity (reliability) in measurement with the test items in this dataset.

The IIF in Fig 5.6.1 signified that the examinee ability and test difficulty for this data set was overall low. The IIF had foci along the X-axis of between -1.66 and -3.23. The IIF curve correspondingly fell at -4.0 logits and continued to fall at a point beyond the scale. The IIF is consistent with the ICC in Fig 5.6, indicating the precision in IIF measurement and confirms that the overall test items in this data set are suitable for examinees' of low ability. There is a general absence of test items in the SCE that assess examinees' with high ability in locations greater than zero logits on the measurement scale. Test items registering greater than zero logits are indicative of moderate level test items.

Fig 5.6.1.1 TIF: Generic Test Items- Consultation and History Taking Cont'd in N=28 Sampled SCEs

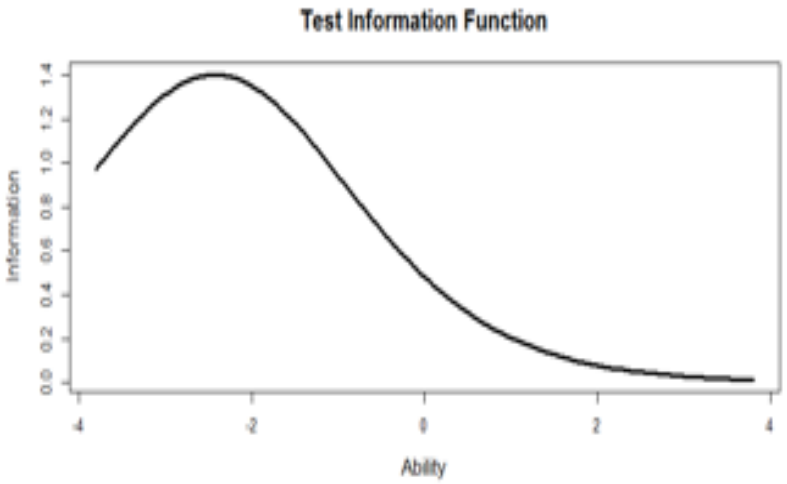


Fig 5.6.1.1 TIF is a mirror image of the IIF in Fig 5.6.1, the measurement of TIF is at – 1.6 to -3.2 and falling at – 4 logits indicates the IIF precision. The test items favour students of low ability.

In Fig 5.6.1.1, the TIF scale is skewed left toward the low ability range. Matching the graphical illustrations in the ICC in Fig 5.6 and the IIF in Fig 5.6.1, a precision indication the items target a low ability level. A valuable feature of item response theory is the TIF. The IIF graphs collectively summarise the precision of measurement for the overall test. Since the TIF summarises the IIF at a given ability level, the overall SCE is too simple for these NMP students' ability. Overall, TIF may indicate that the NMP SCE assessment is specific to students with a low ability range, presenting no real challenge to students with fair, moderate or high capability.

5.7 Item Discrimination (ID) Parameter

In psychometric theory (PT), the test tells how difficult or easy the questions were. The item discrimination (ID) tells us whether the SCE items could differentiate between students who performed well or not, i.e., the item discrimination (ID) index. The ID index ranges between -1.00 and +1.00. Usually, the high-performing students answer correctly for each item more often than the low-performing students, yielding a positive ID between 0.10 and +1.00. Suppose the low-performing students get specific test items correct

more often than the high scorers. In that case, the items will have a negative ID of less than -1.00 and 0.00.

In this study, the SCE test items presented with a DIF index higher than 86%, as illustrated in Tables 5.1.2 to 5.6. These test items are impaired functioning in both DIF and ID parameters. Chapter 6 uses the IVA to discuss the implications of these findings associated with the NMP standards of proficiency claims.

CHAPTER 6: DISCUSSION, CONCLUSION AND RECOMMENDATIONS

6.0 Introduction

This final chapter relates to the research questions, reviews the dissertation's methods, and summarises the measurement results. The major sections of this chapter discuss the findings following Kane's (2006) interpretive validity framework's application to the qualitative and quantitative assessment data. The chapter concludes with the study limitations and closes with the importance of validity to nurse education.

6.1 Summary of the WR Results

Chapter 4 analysed the WR content against the four theories, that of Ryle (1949) and others, Mezirow (1981), Benner, (2001) and Bloom (1956). Applied as a framework to describe and measure in more detail than TDTA alone is capable of, against the student's assessed response. The analysis showed that the WRs content and NMP quality across all four theories were at low levels of achievement in large numbers of the N=28 sampled WRs. The NMC (2006, 2018) Standards of Proficiency for nurses and Midwife Prescribing played a significant role in the TDTA coding for the content validation of the WRs. The theories provided an additional layer to the analysis to determine whether multiple sorts of theory-knowledge transfer, in any given clinical context, was present in the WRs.

The n=20 /N=28 sampled findings were weighted heavily in the lower academic performance levels in Ryle, Mezirow, Benner and Bloom's theories. Suggesting that the WRs content validity is severely compromised. To convincingly prescribe independently and safely, NMPs should demonstrate a sophisticated grasp of knowledge. NMP encompasses masters level study. NMP, as an advanced practice (AP) role, requires sound reflective judgment. A sophisticated grasp of knowledge is central to reflection and the outcome at masters and AP levels (Drennan, 2010; Alfaro-LeFevre, 2013; Halpern, 2013; QAA, 2014, 2015; Rolfe, 2014). An integration of theory NMP subject and practice knowledge is identified in the literature as an enabling or disabling factor in administering, monitoring and critically evaluating the prescribed therapies (Nuttall and Rutt-Howard, 2011; Naughton et al., 2013). Applying Kane's (2006, 2013) validity argument to the assessed WRs, the study aimed to validate students' achievement at the masters and AP

level study. The results suggest that many students should not have passed without demonstrating that they can prescribe safely.

6.1.1 Summary of the SCE Results

Chapter 5 presented the Rasch 1PL measurements' findings. As an aid to the reader- Rasch 1PL is the psychometric statistics for item analysis (IA) of the SCE test items. The results showed that in that there were significant threats to validity in the SCE design, the test item criterion and constructs. Firstly, the item difficulty (DIF) parameter highlighted that the SCE was too easy. The SCEs describe the prescribing work required to require a practitioner to express and integrate multiple information from several domains (NMC, 2018). The ultimate goal of the SCE is to move mastery learning into the clinical training space, which cannot be achieved if the test items perform poorly. The psychometric item analysis application showed that the SCE items were poorly constructed. That the item criterion checklist, which should specify the measurement variables, was also problematic.

The results imply that the general nursing population members can perform positively (very well) on the SCE without difficulty. Secondly, in the Item Discrimination (ID) parameter, the psychometric term for measuring and describing a test item's effectiveness distinguishes those who demonstrate higher ability from those who exhibit lower ability levels. The ID results revealed that the SCE criteria severely lacked discriminating power, presenting negative values, sometimes extreme [i.e., -13.1539] in some test items, as illustrated in Table 4.5. The overall test information function (TIF) suggested that the test was at a level much lower than that required to represent any challenge to the N=28 individual students sampled. Further, that individual examinee of the sampled N=28 had a p-value between 0.84 (84%) and 1.0 (100%) of answering all the 42-SCE test items correctly, indicating that the SCE is too easy.

6.1.2 Re-stating the Research Problem

The NMP assessments are a design product of the NMC standards for Non-Medical Prescribing, and NMC approval is required for courses to run in higher education (HE). Comprehensive assessments are essential to mastery learning. The assessment is a

prominent feature of healthcare education and the student learning experience (Lineberry et al., 2015). The course is mapped against the NMC standards for prescribing (NMC, 2006, 2018). The standards are mandatory, and it is important to ensure the assessment and the assessed work reflect these standards.

The WRs and the SCE classed as Authentic Assessments (AA) aim to equip students with theory-practice abilities on multiple levels (Biggs and Tang, 2011). The validity of such tests is, to date, primarily uncontested in the nursing literature. However, studies are emerging in English Language testing, Psychology and Medical education concerning validity as quality assurance regarding the assessment processes (Huang, 2012; Challoube-Deville, 2016; Tavakol and Dennick, 2017). The validity concern in this study is the score assigned to the student's work. As an inferential measurement, it describes or specifies that the student has achieved an attribute or characteristic that the assessment is designed to measure. At face validity, this may be the case, but it cannot be concluded with certainty or with confidence.

In high-stakes assessments such as NMP, face validity poses a considerable problem for nurse educators and nurses in general in the wake of accountability, safeguarding legislation and safety in patient-centred care (Francis, 2013; The Kings Fund, 2017; NMC, 2018a). There is increasing attention to Kane's (2006) validity theory in the literature. In part, due to medical education and the introduction of criterion-referenced OSCE testing and proficiency testing in English language education (Tavakol and Dennick, 2017; Challoub-Deville, 2016, 2018). Also, due to the undebated and increasingly present issue of determining whether the students meet the agreed assessment objective requirement for certification and licencing purposes. It is on this basis that this research sought to address the overarching research question:-

> How well do the Assessment Processes used in the Independent Non-Medical Prescribing (NMP) course Validate Certification/ Licensing of Students to Prescribe in Practice?'

The study design reported in chapter 3 was an embedded case study. The case relied on secondary data analyses, i.e., the content analysis of the assessed WRs and criterion analysis of the SCE. For each, the measurements employed for validation was that of modern test theory (MTT). The rationale was to provide data that will support the validity

question regarding the interpretation and use of NMP test-scores as stated in the following sub-questions:-

> Sub-question [1]. Are the NMP Summative Written Reflections (WRs) a Valid Constructive Indicator of NMP Students' Proficiency and Certification/Licensing to Practice?
>
> Sub-question [2]: To What Degree are the NMP Structured Clinical Examinations (SCEs) a Valid, and Accurate Representation of the NMP Criteria they are Designed to Measure?

The content validity applied to the WRS in addressing sub-question one was theory-driven thematic analysis (TDTA), mapped against the NMC (2006, 2018) NMP standards for Nurses and Midwives. The cited theories provided supporting validity evidence of student prescribing activities achievement. Similarly, the SCE's criterion validity in sub-question 2, the cited theories provided supporting as these test items are associated with the validity evidence of students' level of ability, i.e., proficiency (or mastery) in NMP learning and achievement (QAA, 2015). Psychometrics statistics, specifically the Rasch 1PL model, was applied to perform item response (IR) and IA measurements. The statistic determined how well the SCE test items represented and measured the instructional theory and subject domain objectives against a student's ability. The rationale for MTT in both sub-question 1 and 2 was measurement as documented in the closing discussion in chapter 2. MTT more favourably matched the structure of measurement required for the NMP assessments to address the validity question.

6.1.2.1 The NMP Assessment Validation Process

Kane's (2006) validity argument formulated the assessment validation and provided a stepwise approach to the NMP assessment validation. A process of information gathering and evidence building to test assumptions about test score meaning (Sireci, 2013; Brennan, 2014). The information-gathering concerned the NMP assessments, the WRs and the SCE. The evidence-building related to the teaching and learning theories used in the analyses (see chapter 1) supports the inferences about student achievement and the link to the test scores (Kane, 2006, 2013). However, it is challenging to work with because there is a lack of specific guidelines for identifying the evidence for validating assessment scores (Sireci, 2007, 2013). The two problems in this study are (1) the

potential to bias in selecting evidence-based on convenience and (2) interpreting the evidence with plausibility to support the assessed work's validity claims. Kane's (2006) validity theory and its interpretive validity argument (IVA) provided a systematic approach for delivering evidence to show the assessed-score inferences. I.e., what the test measures, the (proficiency) claims made about the students. Whether the score use in support of NMP licensing and certification is valid.

6.1.3 The Four Theories as Evidence Measures in Kane's (2006) IVA.

In chapter 1, the cited theories emphasise the different nurse education processes and practices and are, therefore, treated as complementary rather than contradictory. Most of the literature discussion related to validly conclude with the degree to which the assessment tasks map the construct (i.e., theory) of interest. An advance in the validity debate in this research considers measurement theory for applying the evidentiary processes to justify score interpretations(Markus and Boorsboom, 2013). In this study, Kane's (2006) IVA works by building on the classic content-criterion validity theories and applies the cited theories as evidence for the logic of an argument supporting the test score claims of NMP student proficiency (OECD, 2013; NMC, 2018). Proficiency should inform academic standards, the score interpretation and the inferences made concerning student achievement, consistent with the view of masters education, the complex use of theory, and in-depth rather than superficial knowledge (OECD, 2013; QAA, 2015). AAs subjected to validity may ensure that the assessments enable the student to achieve the educational goals (Price et a., 2011; Bloxham, 2012).

Kane's (2006) IVA model is more suitable for validating (with measurement) the students NMP- Knowledge, Skills, Performance (KSP) and cognitive learning to provide evidence for the warrants 1 to 4 presented in chapter 1. In doing so, answer the stated research question(s). Consequently, the identified theories served several purposes, (a) they defined the observed performance domains in the WRs, and (b) they provided an effective way of interpreting the WR content. (c) In the SCE, the theories explicitly laid out the markers for performance measurements, the coherence and plausibility of the assessed scores assumptions about the student and the measurement of their clinical practices.

6.1.4 Chapter Overview

Chapter 6 completes the thesis, presenting a discussion of the assessment results in two segments. Part 1 addresses the content validity of the summative written reflections (WR). The evidence supporting the validity claim using the four theories of Ryle, Mezirow, Benner, and Bloom provides alternative and plausible evidence of the NMP students' reflective writing proficiency. Specifically, their demonstrated achievement of the multiple theoretical perspectives in their WRs. Part 2 addresses the criterion validity of the SCE test items. Of particular interest are the item analysis results from chapter 5, the test item measurements' anomalies, the deviations from the expected parameter values. This thesis ends section 6.4, the conclusion, the recommendations of this study, the limitations and indications for future educational research.

6.2 Discussion Part I:

6.2.1 Professional Education and the Accountability of the Profession

The licencing and certification process in nursing has several goals: Enhancing professionalism, individual practitioner accountability, and protecting the public from unsafe and ineffective care (RCN, 2018; NMC, 2018). As a note on the terminology, the concept of professionalism distinguishes and specifies one occupation from another on acquiring specialised knowledge-skill and completing formal qualifications and examinations (Tuning Project, 2011; QAA, 2014, 2015). Students undertaking these programmes are subject to several theory-practice standards. In the nursing context, professionalism is safe and clinically effective (NMC Code, 2015). These standards outline the profession's expectations, translated as the ability to make practical (theory-driven) intelligent and ethically responsive judgments in clinical circumstances (RCN, 2016, 2018; NMC, 2018). The practitioner is accountable to explain and justify decisions and actions when required (Eraut, 2004; NMC Code, 2015). In nurse education, accountability means monitoring individual nurses' quality of care through legislated registration standards (NMC Code, 2015, NMC, 2016, 2018).

The 2011 post-registration education and practice (PREP) identified multiple prescriptive outcomes to ensure practitioner accountability. PREP safeguarded that the practitioners meet the present and future healthcare challenges and facilitated decisions that lead to a

recordable qualification onto the professional register. The recent revalidation legislation reinforces accountability in the re-registration of qualifications (NMC 2016). The purpose and aim of NMP assessment in this context are to collect, measure and interpret (validate) the information related to the student's response to the process of NMP educational instructions. Neary (2000) and Gotwals et al., (2012) argued the rationale for assessment, reasoning that assessments ascertain the student's theoretical knowledge and practical clinical skills. The RCN (2016) and the NMC's (2018a) drive toward standards of excellence feature in the NMP assessment. Congruent to the professions' legislation requirement to deliver high-quality nursing care, strengthening the clinical performance profile and practice.

6.2.1.1 Authentic Assessment in Professional Education

The nursing professions standards impact the type of assessment used in healthcare and authentic assessment (AA), such as the WRs and SCEs (or clinical simulation) frequently used in nurse education to monitor student s' performance. The thinking underlying clinical decisions made is deemed essential to the patient's outcome. AAs are defined in the literature by some authors as context-dependent, i.e., focused on the care actions taken within a clinical situation (Torrance, 1995; Fischer and King, 1995). The AA's attraction is, that it amplifies and situated cognition in performance-based practice. The NMP assessments, the written reflection, and the SCE are performance related. As a learning theory, situated cognition emphasises a situation-dependent context (thinking, information interpretation and decision-making). Situated cognition requires cognitive absorption and interaction in the knowledge construction process (Lave and Wenger, 1991; Goel et al., 2010).

AA is related to accountability, which suits the stakeholders and those involved in workforce development. The AA has credence in nurse education, defended because authenticity is synonymous with actual clinical events. Aligned with the same skills needed in the real world of clinical practice, AA prompts higher-order thinking skills (Gulikers et al., 2004; Boud, 2010). The ideal assessment yields scores based firmly on the student's cognitive abilities. Assessed work should provide evidence of the individual (or group) abilities to understand multiple and abstract theories. Theory applied in reasoning to support more than one care practice strategy (Boud, 2010; Biggs and Tan, 2011). The AAs are deemed more valid than conventional assessments, providing more

valuable and specific details of students' success in learning (Gulikers et al., 2004; Bosco and Ferns, 2014).

Moreover, AAs such as the WRs provide students with skills to bridge theory and practice. According to Gielen et al., (2003), the fundamental reasons for using WRs are their construct validity and their consequential validity, i.e., the impact on student learning for future practice. However, the published research evidence to support this claim is conspicuously sparse. Informed by this study's results, the theory-practice gap remains. The results further suggest that students' knowledge is compartmentalised within the competency model. The students demonstrated knowledge that was predominantly declarative or procedural and not the expected constructed, interrelated and interdependent knowledge, synthesised (strategic) knowledge at master's level study. Such synthesised knowledge used for thinking relies on a sound understanding of the interplay underlying declarative and procedural knowledge structures (Anderson et al., 2004; Anderson, 2007).

6.2.2 Knowledge, Professionalism and Professional Practice

Drummond and Standish (2007) present the idea that knowledge and professionalism in practice are 'inextricably linked'. Professionalism distinguishes occupations such as medicine and law from manufacturing and mercantile trades (Drummond and Standish 2007, p.1). However, healthcare professionals are under pressure to make practice explicit through NMC standards, performance measures, and benchmarking for healthcare quality improvement. The culture of performance standards suggests that knowledge has taken a backseat in exchange for Lyotard's 'performativity' (Ettorchi-Tardy, 2012; NMC, 2018). Performativity also represents global transferable qualifications that control what constitutes knowledge and what knowledge is important and valued in education (Alexander, 2010; Ball, 2011; 2012; Maguire et al., 2012EHEA, 2018).

Drummond and Standish (2007) eloquently described performativity as a practical science designed to develop practitioners' to become 'technical operatives'. Who do not necessarily need to understand the tasks they are carrying out (p.9). The performativity culture is the unrivalled philosophy of nurse education. A threatening picture, replicated across the current healthcare education sector, capitulates performance measurement in preference to knowledge. Performance measures are the dominant model for judging

'professionalism' rather than educating students to be critical thinkers (McGaghie, 2015). Critical thinking is indispensable regarding the nature and purpose of healthcare practice, care management and delivery (Price, 2016; Vitale and Exley, 2016). The nursing profession has a long-standing political quest for professional recognition and the significance of advancing knowledge to legitimise the occupation. The political pursuit aimed to articulate meaning into what nurses do, what is important about what nurses do, conferring status and legislative control (Eraut, 2004; Colins and Hewer, 2014). Knowledge and knowledge acquisition demonstrated in students assessed work no longer legitimises the profession, and performativity has reconfigured what counts as knowledge in the profession, in this, nurse education is complicit. There is a hidden danger that the nursing profession may be hitting competency targets and minimum benchmark standards but missing the (professional) point.

As early as 2005, Barnett and Coate's work shrewdly pointed out that university education and knowledge had lost ground to operationalism and performance. Declarative and procedural knowledge is the knowledge function of the modern university (p.47). Barnett's and Coate's descriptors are consistent with this study's results. A worrying observation in this study results is an intellectual malaise instead of intellectual rigour (Rafferty, 2002; Baehr, 2011). The latter is the core remit of mastery, scholarship, reflexive thinking and knowledge application to connect theory and practice (Bradbury et al., 2010; QAA, 2012, 2015). Little in terms of compliance with the masters level characteristics of knowledge necessary for cognitive 'critical' thinking is achieved among this cohort of NMP students are visible.

Further, what was also prominent was Drummond and Standish's (2007) skilled performance' reminiscent of the nurse training before the nurse education reforms of the early 2000s. The academic level, established by the NMC, was at a minimum certificate and maximum of diploma. Since the dissolution of the English National Boards (ENB) governance, nurses' education is committed to performance and an outcomes-driven competency framework. Which no longer serves the nurse education requirement for proficiency in today's highly sophisticated and clinically challenging practice environment (NMC, 2018). There is little research except for Banning (2006) and Offredy et al., (2008), which pays attention to NMP students' knowledge acquisition for critical thinking and clinical reasoning about prescribing. A large proportion of the literature focuses on the patient or practitioner experience and satisfaction, the barriers to prescribing (Hobson et

al., 2010; Latter et al., 2011; Ross et al., 2012; Mulholland, 2014; Noblet et al., 2017). Alternatively, on the baseline below which the NMP standard must not fall (Gumber et al., 2011; Weeks et al., 2016). However, the skills that equip the future nurse with the necessary knowledge acquisition for critical thinking and problem-solving are essential. If nurses are to function at the advanced level required for current contemporary nursing practice and professional quality care delivery (Aiken, 2013, Aiken et al., 2014, 2017).

The Bennett et al., (2006) critique of the competency framework suggests a pre-formed plan, underpinned by a linear set of objectives, leading to singularly measurable outcomes. Based on this framework, the resultant curriculum is, conversely, market-driven service delivery (Barton and Davies, 2015). Care provision as a service without praxis, wherein praxis is concerned with theory enacted in a clinical context. The competency framework removes the assessment away from Dewey's how we think (2004) and professionalism, debatably the original philosophy of Higher Education (HE). More importantly, the raison d'etre for nurse education entering HE's spheres (Rafferty, 2002). The significance of which is pointed out by Freire's (1985, p.96) praxis, reiterated by Benade (2016) as the requirement of theory (knowledge) and critical thinking to understand and illuminate practice. Witten and Kvernbekk (2011) contend that students understand praxis only through the educational theory process. Therefore, professionals need theory, critical thinking skills, and skills to positively decide on actions to enhance the patients' clinical outcomes (Rolf, 2014; Vitale and Exley, 2016). An interpretation is that teaching students for competency promotes generalised (unprofessional) performance and compromises safe, effective praxis. Leading to the conclusion that nurse educators, in their adherence to the competency framework, are complicit in de-professionalising the profession (McGaghie and Kristopatitis, 2015). How well students engage with the praxis learning process is mainly ignored as a validity measurement in nurse education research.

Health inequalities are deep-rooted, and patients' clinical needs change in the face of new treatments, interventions and more treatment options (Durand et al., 2014). The quality-of-care demand is an educational issue (RCN, 2016; NMC, 2018). It logically follows that nurses require a solid knowledge base, work on the often-ill-defined clinical problem, and achieve this outcome. In assessed work, practitioners must apply knowledge to contextualise practice and practice decisions (Hall and Ritchie, 2013; Hughes and Quinn, 2013). Knowledge deficits can potentially lead to low standards and care quality (Willis

Commission, 2012). Validity is essential to contest current assessment practices or reduce the errors and near-miss care issues practitioners' encounter in the post-Francis era. The integrity of the competency framework and its uncritical acceptance is open to question and is arduous to contest. In the current research literature on curriculum design, the measurement of knowledge, skills and behaviours in any given healthcare situation is focused on optimum or ideal competent performance. However, this study's results suggest a compromise in the assessment (measurement) in the observed achievement of knowledge in nurse education. In this context, effective and safe care is unlikely. As is the development of thinking skills needed to transform the standards of nurse practice and, at the same time, address the legislated patent safety issues not addressed by the NMP assessment.

When knowledge is used interrelatedly, i.e., strategically, it directly impacts the practitioner's ability to make coherent and reasoned practice decisions to ensure positive patient outcomes (Facione and Facione, 1996; Chang et al., 2011). As such, it provides the means for articulating NMP action (accountability) or rationalising the consequences (patient safety) in their assessed work. The work's quality and arguments are advanced and elevated by contextualised clinical NMP practice interpretation when present in the students' written text. Measuring how students write and contextualise information to demonstrate their knowledge acquisition and professional knowledge achievement is significant; the professions legislators demand proficient practice standards (NMC, 2018). The former being an essential element of nurse education, professionalism, patient safety and the accountability debate. Experts who have studied problem-solving note that the individual's ability to solve problems depends on a high degree of subject and domain knowledge, including a comprehensive framework of theory attributed to depth and breadth (Ennis, 1987, Linn et al., 1991; Barnett, 2012).

In this study, the findings are that the students' knowledge acquisition achievement was at low levels. In applying Kane's (2006) Interpretive Validity Argument (IVA), the results do not support that the WRs facilitate the development of knowledge that enables reasoned thinking. Or independent problem-solving to develop the type of (proficient) NMP practitioner desired. Envisaged decades ago, by Peach or more recently argued by Willis (2012) and Francis (2013), as necessary for safe, accountable and professional practice. Benner et al., (2009) suggest that while students graduate with ample factual knowledge, there is a limited sense of application or integration to the real world of NMP, disenabling

nurses' preparation for practice. Benner's work supports this study's findings regarding knowledge acquisition, the recent

6.2.3 The Rhetoric of Reflection

In the 2000 nurse educational reforms and the consequential move to HE, it soon became evident that the reform disaggregated theory from practice. The attraction to reflection, viewed by many nurse educators as timely, was in its promised feature of addressing the need to shift the HE focuses of theory-learning to encompass practice (Bradley et al., 2010). The ideology of reflection is accepted readily as an essential device for developing the practitioner to challenge, question practice and the hidden assumptions about the existing clinical work forms (Ghaye and Lilleyman, 2010). The review of the reflective theorist's work revealed consensus in their suggestions that practitioners develop professionally in their work by thinking 'critically' about their clinical workplace actions (Bradley et al., 2010; Boud and Soler, 2016). Critical thinking is a crucial element of Mezirow's (1990, 1991) theory of reflection, which speaks of rational thinking about the subject domain (e.g., NMP). The objective of Mezirow's reflection is to address and expose theory-practice inconsistencies and clinical incongruences. Critical reflection draws on critical theory and defines a myriad of cognitive skills that contribute to sophisticated thinking. For example, analysis, reasoning, argument, problem-solving and decision making, described in most critical theory models (Bloom, 1959; Kitchener and King, 1990: Anderson, 2003, 2007; Schraw and Robinson, 2011).

In Mezirow's (1990) theory, 'critical ' refers to these same cognitive structures defined in critical theory, culminating in advanced judgemental thinking. Advanced judgemental thinking resides at the top end of Mezirow's conceptualised reflectivity hierarchy. Knowledge applied and theory debated are used to arrive at a new awareness of the best reasonable solution. Ennis (1987) and Halpern (2014) concede Mezirow's attributes in his definition of reflection as a cognitive and goal-directed activity. The activity is to gather and evaluate information relevant to that goal. Halpern elaborates on Dewey's (2004) published work about thinking and the deductive and inductive processes. The inductive approach allows what is known (theory) to be used to make credible inferences about the unknown. The deductive process facilitates the use of experiential (practice) evidence. The higher-level thinking is essential in Mezirow's (1981) concept of reflectivity,

achieved within levels five to seven of his hierarchy. This current study result suggests that the cognitive skills that contribute to sophisticated critical thinking in reflection are organised and demonstrated by students primarily at level 1 reflectivity. Described by Mezirow as non-reflective and found in n=21/N=28 with the WR content at 70 % attributed to this first level.

In the published guided reflection frameworks, the reflective theory is subsumed by an individualistic version of reflection. The focus is on what they did, thought and felt. Therefore, there is an overemphasis on the importance of emotions in learning to the detriment of the philosophy of reflection, the critical theory, which should underpin reflection and reflective practice (Bradbury et al., 2010; Rolfe, 2014a). The guided reflection frameworks intonation on cognition is underplayed (Fook et al., 2006). Therefore, King and Kitchener's in-depth effortful knowledge and intentional representation of knowledge and theory processing required for transforming practice through reflection are missing (1994, 2004). Moreover, while reflection is, by many authors, argued as espousing this critical theory. In light of the limited research into the levels of reflective practice demonstrated by students, or the effect of reflectivity to practice, this development of 'critical' thinking through reflection is not guaranteed.

A concern from this study's results was the small number of students [n=5/ N=28] who demonstrated the critical reflectivity level according to Mezirow's theory and its criteria. Contrary to achievement expectations at a masters level education and in conjunction with years of clinical practice exposure and experience, the student's pedagogic achievement is far from critical (Fook et al., 2006; QAA, 2015). This current study's findings are consistent with the small number of published research engaged with measuring student reflectivity levels, who also see this absence of 'critical' reflection as problematic(Chadha, 2017). Following decades of promulgating reflection in nurse education, academe and practice, there seems to be a failure to move toward a critical reflection approach (Rolfe, 2014a). The study's results suggest an impetus for measuring reflection levels because of the low levels of (non) reflectivity demonstrated across the N=28 sampled WRs. Crucial for understanding if and how individuals participate in reflexively unearth assumptions about the student's reflective processes.

Academics and researchers such as Fook et al., (2006), Boud et al., (2006), Moon (2007, 2008) and Gulwaldi, (2009) in their studies also found the students tendency toward

reflection in the educational context to be superficial and procedural in their application. Fook et al., urge the rehabilitation of the *critical* aspect of reflection, cementing her theoretical position on reflection by developing critical reflection with professionals on a master's course. Boud et al., propose that the previous potential for reflection be recaptured, suggesting that some applications of these ideas in education courses distort their original intentions. Moon refers to the need to return to the 'backbone philosophies' of reflection derived from the work of Habermas (1974) and Dewey (2004). Moon's debate favours critical thinking as a constituent part of the reflection encompassing a more analytical, critical reasoning necessary for clinical practice (Moon, 2007, p.11). Gulwaldi, in her study in the results of 226, sampled WRs a high number (97.3%) yielded low levels of reflection, congruent with this current study results.

The discussion of students reflective-thinking skills in the literature are prolific across educational and healthcare disciplines, and there are multiple publications in this subject area. However, by various methods, fewer researchers measure students' written content to challenge the firmly held beliefs and assumptions about students' reflective ability levels. The Wallman et al., (2008) study measured reflection among pharmacy students using Kember's categorisation scheme; their study found that none of the students reached the highest level of reflection. A low mean level of reflection was prominent, consistent with this study's results. Conversely, in 2010, McNeill et al., used a modified Delphi technique to judge both the quality and quantity of the reflection of 30 medical students' text from a clinical event with the scoring system Level 1, most reflective Level 2, Intermediate Level 3 the least reflective. Much of the written content scored level 2 or 3, with just n=16 out of 163 (10%) scoring level 1 (non) reflective of the entire text analysed

Fundamental to reflectivity is the student's critical thinking to solve problems. Experts who have studied problem-solving argue that the process depends on both theory and practice factors. Plus, the amount of domain-specific knowledge at the individual's disposal, in conjunction with deliberate (experiential) practice. Therefore, the 'critical' within nurse education attend to measurement of critical in assessed work conditions. Moreover, there is the assertion that critical thinking is an essential element for maintaining patient safety (the assessment outcome). Reflection as an AA, when correctly executed, gives a more transparent view of the theory-practice achievement. The students critically reflect on what is 'real' and in the clinical sense regarding issues of care needing immediate attention in the work situation to maintain patient safety (Cioffi, 1998;

Benner, 2001; Moon, 2007). Engaging nurses in deliberate learning through AAs such as the WR is a critical pedagogy model of active learning. Designed to prepare students for employment challenges, AAs are concerned with what students need to know and understand (Coles, 2014). There is mass access to medical and health-related information for nurses practising in a technically demanding socio-political context. The emphasis should be placed on enhancing knowledge to apply critical thinking to health-related information to inform practice.

Kane's (2006) IVA provides evidence that the WRs content does not relate to master's level study. What students can do in terms of reflectivity has relevance in subject-domain knowledge theory. An essential factor is that students can render critically reflective judgement about practices. In generating evidence for the IVA, the current study results do not support the assessed scores global assumption or interpretation that students' achieve the assessment goal (Guskey et al., 2013a). Reflectivity as the critical reflective element is visible in n=5 of the WRs N=28 sampled. The IVA points to the reflections being little more than a recipe for compliantly conforming to the legislator's competency surveillance. Instead of facilitating the application of multiple sources of knowledge and evidence required for critical reflection, the evaluative thinking and the questioning and challenging necessary for managing care in complex clinical situations. The critical thinking and reasoning skills displayed in the summative WR content may be appropriate to the competence and the technical procedures mapped out by protocols. However, critical thinking and reasoning are the outcomes of a mastery learning program (McGaghie and Kristopaitis, 2015; QAA, 2014, 2015). In the Advanced Practice (AP) context, the absence of critical thinking and pedagogic attainment, there is the potential for protocols to be rendered ineffective (Levett-Jones et al., 2010; Lewis, 2012; Rolfe, 2014). The implication is significant increases in risk to patients.

6.2.4 Practice and Formal Academic Standards

6.2.4.1 The Practice Level for Advanced Practice

The 2006 Modernising Nursing Careers Report saw advanced nurse practise (AP) as a nursing professional evolution product. During the formative years, the RCN established the Advanced Practice (AP) baseline competencies and nationally. Universities developed

programmes for nurse AP education. The policy drivers stated masters-level education as the requisite to the AP role, professionalising the role with the public and legitimising the AP with employers and the medical staff (DH, 2010; RCN, 2012: Health Education England, (HEE, 2017). In many countries worldwide, a qualification at a master's level is now the accepted norm (Pulcini et al., 2011). Within this context, the program presumes to produce individuals who are for-purpose and practice. While the core AP nursing practice requirement is constant, the roles undertaken are determined by the employer. For example, paediatrics and enteral feeding services (NHS, 2019). There are bespoke elements of the role, one of which is NMP, vital to alleviate human resources shortage. There are a clear education and regulatory process (NMC, 2006, 2018; RPS, 2016).

Consistent with the UK, European and international literature, the agreed expectation is that APs demonstrate expertise in a core set of competencies, significantly above that of a newly qualified nurse or an experienced nurse (Bagley et al., 2010; Jokiniemi et al., 2012; Gardener et al., 2013). The NMPs are clinical specialists in various areas, e.g., emergency care, oncology and subsequent patient care. The course entry requires that the NMPs work in the clinical field they intend to prescribe and prescribe in areas within their clinical practice scope (NMC, 2018). The clinical specialist areas in which NMPs practice provide the experience of illness, medications and patient care responsibility wherein they exercise judgement, specialist knowledge and skill (NMC Code, 2015). The fundamental nature of the AP philosophy in education is that the practitioner has already attained a skill set of practice knowledge and experience within their specialist discipline to judge. Whenever possible, the clinical and health-related risk calculation, decisions made on which aspect of healthcare needs exploration, and the action taken to minimise the risk (Rolfe and Fulbrook, 2001: DeNisco and Barker, 2015).

These already attained practice knowledge and abilities are ‘advanced’ under a master’s education programme (RCN, 2012). For several decades researchers have studied the association between nurses’ (expert) education level and the quality care provided. Nursing education argues that nursing education (a degree) is related to decreased patient mortality. The 2013 and 2017 RN4cast reports showed that hospital quality, patient safety, work environment workload problems are common in all countries. Research in the UK and internationally found that hospitals and departments with proportionately fewer qualified professionals delivered care to patients at a secondary quality of care level (Jarman et al., 2009; Keogh, 2013; Aiken et al., 2012; Griffiths et al.,

2016). Decades of published research studied the correlation between (expert) nurses' level of education and the quality of care provided, presenting the argument that nursing education (a degree) is related to a decrease in patients' mortality (Needham et al., 2002; Kane et al., 2007; Zhu et al., 2012; Schubert et al., 2012; Aiken et al., 2014). However, these studies focus on first-level entry nurse registrants, degree-level qualification and learning. The information is transferable to masers and AP nursing as it concerns staff capabilities on patient outcomes.

The data in these studies were aggregated at the hospital level and combined with graduate nurses' surveys to estimate the population's qualification level. The researchers' are inconsistent in their findings and the correlation of patient safety results. The implication is that suboptimal nursing care is attributed to limited educational preparation. However, the defining mechanism for the association between higher nurse education and better patient outcomes, to date, remains unclear (Kuteney-Lee et al.,2013; Aiken et al., 2017). Accentuating the importance of educational preparation at master's level for the AP role, the NMP programme content mirrored that criteria embedded within the AP program in its requirement of advanced KPS. Including a comprehensive physical examination, history-taking and clinical decision-making, evidence-based practice, accountability – including legal and ethical issues (RCN, 2012 p.11). Therefore, for the NMP student wishing to practice at an advanced level, in a bespoke NMP element of the role, the assessed performance must be judged proficient rather than the competent standard of initial registration.

Benner's (2001) Novice to Expert trajectory applied in this study identified whether the legislated level of NMP practice, i.e., proficiency (or indeed expert), is achieved. In categorising and measuring the level (standard) of practice articulated by students, the results revealed a significant proportion of the NMP students function at level 1-Novice with n= 22/N=28 of the WR content between 50 -100 %. Alternatively, advanced beginner (level 2) was found in n=21/N=28 WRs. The content lay between 10% for a minority of the N=28 sampled students, with the majority at 30% - 45 % among many TDTA codes. The concept of 'novice' level practice in an NMP or master's program only makes sense in context to these results. If, as described by Benner, it means students are starting from the beginning at something new. Rather than building on and advancing, the already established KPS familiar to all registered nurses.

Further, AB does not make sense in students on this programme with an entry requirement exceeding more than twelve months post-qualifying and years of practice knowledge, often embellished by specialised and high-levels of advancing practice experience from which to draw on and transfer into NMP actions. The expert nurses are the product of their education level, whose practices testify to advanced learning, i.e., Scientia, derived from the word science, logic, (critical) thought and reasoning; and the Latin origin of knowledge (Oxford Dictionary, 2013). In a review of Benner's work, she defines the expert's practice as 'intuitive', disputably a semantic disaster. This terminology detracts from what she has analysed and described in detail as a high-level knowledge and clinically experienced specialist (Benner, 2001).

The findings in much of the N=28 sampled WRs the content does not evidence a culture of advanced education, but one of 'anti-intellectualism' (Rafferty, 2002, p 4).' A denial of a need for considerable education or knowledge base. Instead, the focus was on the accuracy of a performed task against procedural protocol. Despite the profession's legislator's promulgation of master's education and its critical significance, it supports innovative, flexible clinical practice (NMC, 2018; HEE, 2017). The NMP role aligns with proficiency. The notion of Benner's expert and Rafferty's intellectualism is the optimal result of clinical and academic progression (Traynor, 2013 p.44). Without this, the NMP specialisation will continue to grow based on performativity and lack in-depth and breadth of anything, including the espoused proficiency in its specialist subject-domain.

Benner's (2001) identification of the novice to expert model lends itself to the content analysis of assessed WRs. Kane's (2006) IVA substantiated the way nurses demonstrate and articulate how they practice at each level of the model's continuum. According to Benner's theory, the level of practice performance shown in the WRs demonstrated in the students' understanding of the relationship between NMP theory for practice was inadequate at novice to advanced beginner level. Barton (2012) and Rolfe's (2014) publications about the AP nurses' ability to effectively fulfil their roles express concerns the AP nurse do not demonstrate advanced expert capability as the distinguishing feature. The IVA, when evidenced against Benner's models' characteristics, unearthed the implausibility of the proficiency claim. Mirroring Barton and Rolfe's concerns, proficiency does not hold relative to the small number of proficiency achievements demonstrated. I.e., the n=3/N=28 sampled WRs.

6.2.4.2 The Academic Level for Advanced Practice

In nurse education, students must display specific behaviours, i.e., scholastic learning achievements broadly agreed governance by the QAA (2010, 2014) and the NMC (2012). For nurse educators, the challenge is to ensure that focus is not merely on the technical and performance of procedure but involves significant intellectual development articulated by educational achievements. The summatively assessed must work to portray scholastic aptitude beyond making simple inferences (QAA, 2012). The QAA is clear about students' responsibility for independent study and demonstrating masters level achievement summatively. Concerning the ability of APs to fulfil their role, this relies on the level of critical thinking, clinical judgement, that is advanced (Rolfe, 2014).

In this researcher's university in the graduate programs, Bloom's Taxonomy (BT) of educational objectives, the cognitive domain, is used in the assessment. Classifying and identify the expected thinking levels students should employ as they respond to the coursework requirement. BT is suitable as a framework for conceptualising and measuring individual academic accomplishment, for delineating the boundary between higher-order thinking (HoT) and non-thinking attainment or achievement (Gipps, 2012). Bloom's Lexile, not explicitly defined as HoT within the BT philosophy. However, Bloom's work set the landmark for HoT classification. In BT, the HoT refers to the analysis, synthesis and evaluation. Although all authors and scholars do not conceptualise HoT as Bloom did, many retain Bloom's process in assessments designed to test HoT (Schraw and Robinson, 2011).

Alexander et al., (2003) distinguished several types of knowledge identified in studies on thinking, including domain-specific knowledge. Postulating further, if domain-specific knowledge is taught, a strong case is made that will inevitably entail high order thinking (HoT). Mastery of most subjects requires applying (use), manipulating, and interpreting knowledge in the subject area. Resnick's earlier view is that HoT skills involve "elaborating the given material, making inferences explicitly presented, building adequate representations, analysing and constructing relationships" (1987, p.45). HoT thereby develops in the student 'mastery' of subjects, concepts and problem resolution of the more complex clinical problems. Ennis (1987) earlier work elaborated that HoT includes judging the evidence's credibility, argument, reasoning, and making value judgements.

Schraw and Robinson (2011) similarly assert logic and debate are essential in learning for mastery.

The available research on assessment focuses on the effectiveness, best teaching practice, designing and assessing courses, and not HoT skills (Oermann et al., 2007; Wang and Farmer, 2008; Wehlburg, 2011; Williams, 2017). Where studies exist, they offered only dissatisfaction when considering BT in the measurement of HoT. For example, in a dated study by Hopson et al., (2002), the development of HoT was explored as defined by BT through the students' use of word-processing, databases and spreadsheets. The methodology was unclear; however, these authors concluded that the technology-enriched classroom and the successful use of the software were indicative of the development of HoT. Rupani and Bhutto (2011) sampled N=30 ninth and tenth-grade public school students using interviews and focus groups. The authors disclosed that the development of analysis, synthesis and evaluation skills remained untouched. In a qualitative study in 2014, Kantar used BT to categorise the content of a series of written assessments, i.e., case presentations, reflections, reports and summary care plans. These authors did not report the BT categorised content; however, they concluded that teaching to pass the test could impede knowledge transfer and students' HoT development.

A considerable problem of note in this study's WR results was an overwhelming tendency to low-order thinking. Coverage of a broad (superficial) range of NMP experiences rather than demonstrating an in-depth understanding. Or critical debate and recognition of NMP topics with substantial knowledge of the relationships and complexities necessary to confront important NMP issues (Linn et al., 1991; Granello, 2010). Of concern was the small number of students n=2/N=28 (7.1%) who demonstrated/achieved the HoT level ‘analysis’ in their WRs. For the IVA, if the HoT is absent in theory, the assessed setting, it is plausible that it is also missing in the clinical setting. However, passing the assessment, the general assumption is that students are using HoT skills. Schraw and Robinson state that students are expected to display the incremental gains and changes in how they think about the nature of knowledge (or metacognition) as it applies and the programme's end (2011, p.71). However, most of the NMP students at this summative stage were unable to do so. Within the confines of professional legislation, it is unsafe to assume that students have achieved HoT to develop appropriate strategies to meet today’s healthcare challenges.

6.3. DISCUSSION PART 2:

6.3.1 Introduction

According to Bloomfield et al., (2010), in the (objective) Structured Clinical Exams, examiners usually assess performance related to the acronym KMAS (Knowledge, Motor, Attitude, Structure). Directly relative to the performed skills, such as understanding and the technical aspect of the skill, the affective and attitude associated with that performance and the students' systematic, technical- organisational approach. In this research report, Bloomfield et al., KMAS, clearly describe the SCE performance, a task-based or behaviourist activity. Eraut (2004) seminal work vehemently questioned as representative of competence. There is universal agreement in the international literature on OSCE that clear criteria are necessary for a valid assessment. Validity is expressive of what a student 'actually' does clinically and safely to ensure the legislated professional standard (Norcini, 2003; Sackett et al., 2008; McGaghie and Kristopaitis, 2015).

6.3.2 The SCE Criterion Quality and Constructs

The success of an SCE is partly dependent on its scoring and its ability to discriminate between levels of expertise. Which is particularly relevant for standard-setting in performance and when assessing postgraduate students. However, to say, a student has reached a certain standard. The SCE criteria need to be designed well and accordingly weighted to confirm that students have a deep understanding of the concept. Therefore, this part 2 of the discussion addresses the following:-

> *Sub-question [2]: To What Degree are Structured Clinical Examinations (SCEs) a Valid and Accurate Reflections of the Criteria they are Designed to Measure?*

The approaches to marking the SCE performance are the checklist. The clinical version divides discrete and measurable tasks checked off on a list regarding whether they are (or are not) completed. Typically, checklists contain an agreed-on (by clinical experts) selection of items that an examinee should address in a clinical encounter. Checklists, if written well, are ideal in assessing skills that require the completion of several steps, the same way every time. For example, those within the SCE where the test-taker greets the patient introduces self and role (see Appendix 5a). However, there are measurable

differences between novice, acceptable average performers, and excellent (experts) for other questions. The information is difficult to capture with the current SCE or standard checklist (Hughes and Quinn, 2013 p.396). Table 6.3.2 contains a sample of the SCE content to illustrate the current assessment practice and the checklists' complex performance type.

Table 6.3.2 illustrates a sample of the Targeted Domains for Consultation and History Taking

Student Name.....		Student Number			
***The test items are mandatory- A fail score in these areas = overall SCE fail or potential recovery within viva**					
The Consultation skills and history taking Criteria	Stated	Incomplete answer	Not stated or incorrect	Recovered within VIVA	PASS/FAIL
* Has considered the relevance of all other treatments/medication the patient has already taken or is taking					
*Has considered the relevance of the patient's medical history					
*Has considered possible cautions / contra-indications or previous ADR's					

Table 6.3.2 contains an excerpt from the SCE examination, used here to highlight the content and criteria scoring. For complete Table see Appendix 5.

The criteria in Table 6.3.2 are compound questions, which seem to be single questions containing multiple responses, which make scoring difficult (Downing and Haladyna, 2006; Hardern et al., 2016). Moreover, the four variables in the scoring criteria in the Table can obfuscate the scoring. Confuse results when evaluating an examinee's score, and the items lack of specificity compromised its usefulness in assessing performance. For example, with scoring criteria such as 'Incomplete answer,' there is a universe of responses to capture. Increasing with each of the test item questions, making scoring difficult and the SCE even more inconsistent in the criterion it is designed to measure.

Consider the sampled SCE content in Table 6.3.2. Ponder, for example, the first question, *"Has considered the relevance of all other treatments/medication the patient has already taken or is taking"*. Two students evaluate a patient whose presenting complaint is

chronic urticaria[38] , and the responses vary. Student *A* first asks about dietary changes, soap used for laundry, or products used in hygiene habits. In the patient's presentation, some information has prompted the consideration of local allergic response. The safety concern is ubiquitous here. Student *A* incorporates an understanding of common causes of skin rashes into history. Eliminating these as a cause for the presenting skin complaint then focuses on the cause in the presence of the undelaying illness before attending to medical history ADRs. Comparatively, student *B* asks about skin changes using a review of systems used on every patient regardless of their clinical presentation. However, as the current scoring stands, both students will get credit for these on an item's checklist. This study's IA measurement revealed the checklists might favour Benner's (2001) novice learners, who often adopt a stepwise 'checklist' approach to problem-solving as in the student B illustration.

The current SCE checklist in Table 6.2.3 does not separate students performing at different levels. As highlighted by the IRT in the IA results, the ID measurements confirmed the SCE design's unsuitability due to its inability to differentiate between high and low ability performers. Therefore, these differences need to be the basis for SCE assessment to advance students from acceptable and standard proficient to expert practitioners. Under Benner's model, student *A* and student *B* could score as functioning at two very different levels. Table 6.3.2.1 Illustrates the proposed revised content and scoring criteria. This modification defines the incremental levels to assist assessors in deciding whether it is an expert achievement

Table 6.3.2.1 Illustrates the Proposed SCE Scoring using Benner's Model.

***The test items are mandatory- majority scores in the novice area = overall SCE fail with potential for recovery within the viva.**

The Consultation Skills and History Taking Criteria.	**Level 5: Expert**	**Level 4: Proficient**	**Level 3: Competent**	**Level 2: Advanced Beginner**	**Level 1. Novice**	Pass/Fail
*Has considered the relevance of all other treatments/medication the patient has already taken or is taking.						
*Has considered the relevance of the patient's medical history						
*Has considered possible cautions / contra-indications or previous ADR's						

Table 6.3.2.1 incorporates Benner's model, allowing the scoring to indicate against each test item where a specific practice level is achieved.

Further, a subjective statement such as ‘recovered within a viva’ should be removed. The ‘recovery viva’ item may still constitute a mediocre performance and unscored as such. The inclusion of Benner's model in the scoring system may allow for more precise judgements about the students' practice level and for this information to be appropriately scored against the test-takers response.

The literature on the OSCE indicates that checklists are not a good discriminator of expertise, resulting in criticism of the assessments' approach (McKinley et al., 2008; Brannick et al., 2011). Moreover, empirical evidence showed that checklists penalised experienced clinicians by omitting the capture of higher-level practice (Hodges et al. 1999; Hodges and McIlroy, 2003; McKinley et al., 2008a). The research literature encourages and recommends supplementing the checklist with other measures to include items that help discriminate between learners who understand the subject matter and those who do not. I.e., an essential features approach (Daniels et al., 2014; Ilgen et al., 2014). The illustration in Table 6.3.2.1 draws on Benner's model continuum in this context.

Hodges et al., (1999), in their comparative study of the standard checklist and global rating scores (GRS), argued that the scores in the checklist declined with the practitioners increasing level of expertise. Further, the checklist works best where questions are in stereotypical order irrespective of whether they are relevant to the case. Experienced practitioners collect, interrogate and integrate many forms of clinal data, and ask fewer focused questions to arrive at a hypothesis, diagnosis and management (p.1133). Therefore, experts do not score points because their process is not stepwise and suited to the task, which is what the checklist is presumed to capture. Expanding on previous studies, Hodges and McIlroy (2003) confirmed that checklist ratings neglected the higher levels of ability and knowledge organisation, favouring simple fact accumulation and demonstration of the task. The McKinley et al., (2008) systematic review concurs with these other and other studies. The review further stated that checklists do not explicitly assess critical competencies such as ‘safety’. Given recent high-profile healthcare scandals, much hangs on the SCE as an AA. Good clinical practice is becoming dependent upon good assessment design and assessment practices (Hodges and Quinn, 2013). A factor that should concern all who teach and assess healthcare professionals.

6.3.3 Asking the Right Questions in the Right Way

The literature reports extensively on instruments used in OSCE's and simulation to measure undergraduate student learning in the affective and psychomotor domains. However, few studies address higher-order thinking (HoT) competencies by a credible scoring method. Non-medical prescribers encounter diverse, complex medical conditions in practice, requiring theory-practice knowledge. The SCE should be equally complex and varied for NMP authenticity and should assess the high-order thinking and knowledge skills needed for safe NMP practice (Kroenzen et al., 2011). At face value, the SCE appear to do this. However, the item analysis in chapter 5 identified the performance-based (binary pass-fail score) checklists design of the NMP simulation (the SCE) was problematic. The SCE presented no real cognitive challenge.

The literature indicates that performance-based checklist scoring systems are fraught with difficulty. The assessors will agree on the observed performance but disagree over their interpretation of the assessment's focus (i.e., the construct). Alternatively, on the response format, i.e., the meaning of the scoring criteria, assessors may indiscriminately positively score test-takers (Crossley, 2010). The high order thinking (HoT) skills are attributes cited as required for safe nursing practice. Therefore, the SCEs content should include the HoT criterion (Benner, 2009; Beckwith, 2018). The problem of assessing HoT and measuring what matters most is pressing in postgraduate NMP education. The SCE assessment methods must be robust to ensure practitioners are fit-for-purpose for the complex clinical practice environment in which they work (McGaghie, 2014, 2015).

An emphasis on healthcare education is in fitness-for-purpose. In the context of this study, the IRT and subsequent IA served several functions in test item development to address this concept. Firstly, identifying flaws in the test items, pinpointing items that poorly measure an attribute of interest and those that put test takers at an advantage or disadvantage (Downing and Haladyna, 2006). Secondly, in the fitness-for-purpose perspective, referring to the test's quality and its representation of the standard of practice, it aims to measure students' assessment. Therefore, the assessment design should facilitate demonstrating the knowledge and skills intended by the NMC (2018) proficiency and QAA (2014, 2015) standards. Finally, including items with a range of difficulty enabling students at all levels of ability (traits) to demonstrate their level (Bhola et al., 2003; Hardern et al., 2016). A trait is explained as a disposition to behave or perform

in response to stimuli in various circumstances. A typical example of a trait is a higher-order thinking skill, which has generated substantial cognitive psychology research. BT as a framework provides insights into how learners reason, think through problems, and become experts (Facione and Facione, 1996; Ericsson, 2004; Benner, 2001, 2009).

The literature on test design is direct in its statements regarding the measurement of 'traits' determined by the questions' format. Format consideration should include more than the formatted response, but how the problem presents to the candidate to determine what is captured (Crossley, 2008; Mackillop et al., 2011). The SCE items in Table 6.3.3 yielded the most extreme negative DIF values of -13. The item discrimination (ID) parameters were incalculable. The item questions listed in the Table are primarily non-compound performance items designed to investigate a single value. Rated independently as a 'performance', an activity that post-registration nurses could quickly achieve with precision due to either rote learning or recall of prescription writing errors from their extensive clinical exposure to written prescriptions. The SCE scoring criteria are ambiguous. For example, an affirmative score is achieved easily by less able students who do well on recall tests and mere memorising of the content (including errors). Or through familiarity with the prescription documentation structure, i.e., what, where, and by whom (Hodges, 2006).

In reviewing the SCE criterion following this study's results, the IA highlighted several explanations for the high level of pass scores, easily attributed to proficiency, explaining the test score results. However, the item analysis suggests that this can also be an erroneous interpretation of the score results. The SCE design needs attention. The reasons are (1) the ambiguous scoring criteria, (2) student oversensitivity to the test items. (3), the item criterion does not correlate with any other items on the test or (4) poorly written items; point 4 should be rewritten or discarded (Downing and Haladyna, 2006).

Table 6.3.3 The Current SCE Checklist Criteria for Legal Prescription Writing

Item Criteria	Completed	Incomplete or not completed
Name		
Address		
Age		
Date of Birth- only legally required if child under 12 years		
Name of drug/product		
Appropriate drug/product		
Quantity		
Dose		
Frequency		
Duration of treatment (or as required)		
Directions for use		
Signature		
Date		
Blank Space underscored		
*The test items are mandatory- fail in one these areas = overall SCE fail		

Table 6.3.3. Illustrates the criteria for the performance of writing a prescription. Each item criterion is awarded 1 mark. Students can achieve in total 15 marks on this activity. For complete Table see Appendix 5.

The above Tabled SCE questions presented no contest to the test takers intellect. The answers are easily achieved through rote learning. Therefore, the logical extension and improvement in the SCE are to look at the trait approach (Secolsky and Denison, 2012). The extent to which the SCE included assessment of reasoning and cognitive skills around prescription writing (PW) is significant to the purpose of this assessment as evidence for the IVA. The SCE is, to a high degree, simulated to the job of an NMP. Therefore, it has immense relevance to critical features such as cognisant reasoning.

The SCE item criterion and the design should capture the intellectual discourse and other cognitive traits of prescription writing. For example, the significance of treatment duration, i.e., days, weeks or months, relates to cognitive traits, such as critical thinking and reasoning skills, for deciding on treatment duration (Nuttall and Rutt-Howard, 2011). Based on the patient's presenting condition and governance guidelines or the literature and formulating questions (hypothesising), searching for relevant patient information. Making informed judgments is not new to advanced clinical nursing. However, in the current SCE design, the Prescription writing checklists' limits the evaluation of more subtle, clinically relevant nuances than those listed (Cunnington et al., 1999; Tavakol and Dennick, 2011; Schuwirth and van der Vleuten, 2014). A Global Rating Scale (GRS)

application may allow for a scored recognition of discreet achievement levels. The GRS may permit examiners to reward students who excel in cognitive skills or identify students who require further development in specified domains.

For the OSCE's in medical education and simulation in nursing, the literature contends that the supplementation with global rating scales has improved the testing of traits and other learning areas of interest. A more comprehensive framework than the frequently used checklist is recommended (Kogan, 2009; Schuwirth and van der Vleuten, 2011; Prion et al., 2017). The literature often cites two types of GRS scales, holistic and analytic. Holistic rating scales, more so than a detailed checklist, provide a comprehensive score on overall performance. Analytical rating scales offer information on several dimensions. High scores are assigned individually to each characteristic assessed, identifying specific components the examinee possesses or needs to develop (van der Vleuten and Schuwrith, 2005).

Other researchers have employed GRS and applied conceptual frameworks specific to the skills and processes they wished the student to engage with and which they wanted to measure. Kim et al., (2009) used the GRS to assess crisis resuscitation management in postgraduate year 1 and 3 residents in a simulated emergency/ acute care setting. Essential for any acute care speciality and in the absence of a suitable assessment tool, the Kim et al. designed provided a seven-point anchored ordinal scale for performance in five categories and overall performance score. In their assessment of the socio-cultural aspects of patient safety in nursing and medical students, Ginsburg et al., (2014) designed and used a single 5-point holistic global rating scale ranging from clearly deficient to excellent. Ginsburg et al. aligned the GRS to oral questions, allowing students to demonstrate 'safety' knowledge that is not elicited during their simulated interaction. The Turner et al., (2014) GRS assessed professional behaviour in physical therapists; a five-item scale included professional conduct, communication, respect, compassion, and insurance of patient dignity. The authors concluded that the GSR was more suited to the assessment of advanced students than the checklist. Table 6.3.3.1 illustrates a proposed design for an analytical rating scale for the SCE. The analytical GRS gives a clear and intriguing indication that a students' cognitive skills could be determined, adding a meaningful signal to testing and measurements, other than performance competence (Cohen et al., 1991; Regehr et al., 1998; Hodges et al., 1999; Munoz et al., 2005).

Consistent with the cited research, the GRS for the SCE Bloom's taxonomy (BT) overlays the scoring to identify the cognitive levels and the desired characteristics against the selected NMP domains to which they belong. The test design literature recommends that the GRS, where possible, are construct-driven rather than task-driven. The proposed SCE in Table 6.3.3.1 is construct driven, designed to reflect the overall construct measured and delineate BT assessed via the specified NMP task (Klein et al., 1998). In that way, the overall effect is a descriptor of the individual's cognitive characteristics, collated in a meaningful way to arrive at an overall score (Epstein et al., 2004). BT is applied in this context to represent and measure specific cognitive goals and analytic scoring of the students' cognitive standing (Baxter and Glaser, 1998; Hodges and McIlroy, 2003).

In Table 6.3.3.1, the example has several critical thinking criteria in the colour zones. Based on the global score weighting and the representative level, the overall score will better indicate the students' achievement. See, for example, Bloom's Taxonomy overlay in Table 6.3.3.1 for the criterion 'right time.' In this example, the student has scored in level 2 understanding in BTs criteria, in the orange zone where assessor prompting was required. The 'right time' is essential in prescribing practice. The students must be critically aware of drug times, i.e., twice per day (BD). Three times per day (TDS), or QDS (four-time per day), all have implications for the therapeutic drug effect. The correct time is significant in antibiotic treatment to maintain the optimum plasma concentration and antibiotic effect (Kizior, 2016). Scores in the red area, as in the example, are cause for serious concern. Alternatively, other colour zones indicate a range of achievements from performed exceptionally (the green zone) to major faults (the red zone). Indicating the student has not or has not achieved at the required level.

The 5- global rating scale highlights the student's expectations of the SCE in marking criteria. The information may aid in study preparation for the assessment. In the example, the student scored on a test item in the orange zone, which indicates that further study is needed. The GRS scoring will also help the assessor give the student constructive feedback on the 'level' of cognitive achievement within each BT characteristic, identifying student trajectory or areas for further development. Further, by including Benner's model, Novice to Expert levels I to 5 to the GRS would add the practice dimension, showing whether the student has novice, proficiency, or expert capability.

Table 6.3.3.1 An illustration of the Proposed Global Rating Scoring, using the Six Rs for Safe Drug Administration. Abridged from the NMC (2007) Standards for Medicines Management, the NMC Essential Skills Circular 07/2007 and the RCN (2007, amended Jan 2019.

Criteria	**Analytical Rating Scale- Blooms Taxonomy of Educational Objectives- Cognitive Domain**																										**Guidance**
Task-legal Prescription Writing	Level 6: Evaluation					Level 5 synthesis					Level 4: Analysis					Level 3: Application					level 2: Understanding					Level 1 : Knowledge				
	Theory/practice Criteria clearly Satisfied, Performed Very Well					Most Theory/practice Criteria are Satisfied Performed Well					A significant number of Theory/practice criteria met Performed Reasonably Well					Sufficient Theory/ practice Criteria met Acceptable Performance					Theory/practice criteria met in performance, prompting required					Only minimum (Major faults)				
The student can engage in:	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	
• Right patient											X																				SEE NOTES A
• Right Drug										X																					
• Right Dose											X																				
• Right Time																									X						
• Right Route											X																				
• Right Authorisation										X																					
NOTES A:	Each X denotes 0.5 = across the levels: total mark = 3																										

Right Patient-Rationalises the requirement for Patients Name, date of Birth and Hospital ID number- context- legal, ethical/other i.e., protentional need to contact patient regarding prescription.
Right Drug- Rationalises the need for a clear and legible named drug on the prescription. Identifies the generic or approved official name, and not the trade/brand name, should be used (unless appropriate). Explains the significance of Any allergies, the need for this to be highlighted on the wristband as well as on the drug chart. Rationalises specific type of the drug, e.g., enteric coated (EC), Elixir (Elix), Slow release (SR) 'why' should also be clearly indicated.
Right Dose- Explains why dose must be present and clearly written/indicated e.g., dose size numerically 25, 250 and dose units i.e., Micrograms (mcg or µg), milligrams (mg), Millilitres (ml). administered. Critically identifies the significance of - The name of drug , safety checks and states the maximum dosage , any SAE/ADRs of medication.
Right Time- The significance of the time the drug is to be administered i.e., 08:00, 14:00 . i.e., must be clearly indicated/documented ensure efficacy. Student explains -A drug needs to be at the appropriate time(s) relative to effective pharmacokinetic outcomes (antibiotics, for example, significance of maintaining plasma half-life).
Right Route-Critically Discusses: The route the drug is to be administered must be clearly and unambiguously stated/documented e.g., Intravenous (IV) oral (O) per rectum (PR) subcutaneous (SC) sublingual (SL) *Caution as SL can be misread/mistaken for SC if poorly written. Rationalises /explains why some drugs cannot be administered by the oral route (GTN or insulin, for example). Others must be administered IV for 100% bioavailability.
Right Authorisation- Reasons (context- legal, ethical/other) the significance of Illegible, incomplete, incorrectly written prescriptions-not used to dispense medication to patients. Explains implications of prescriptions where the prescriber's signature is illegible or omitted, Medications are not dispensed to the patient.

Table 6.3.3.1 illustrates and analytical global rating scale, displays how the activity of writing a prescription can be enhanced include critical-thinking to represent a more cognitively challenging SCE assessment. The form can assist the assessor by removing ambiguity in the scoring and support the students' preparation for the SCE assessment.

In this study, in the SCE's there is an overemphasis on technical performance, which may lead to hidden incompetence. I.e., the performance behaviour is at the expense of high-order thinking within the domain-specific knowledge. Which excludes knowledge integration and synthesis to which safe practice is bound (Eraut, 2004; Brooks, 2010; Carracio and Burke, 2010). A more meaningful approach would be to change the scoring format; from a task-oriented, technical performance to one more cognitively challenging (Mackillop et al., 2011). In addition to the format change, altering the test delivery method is recommended, such as a Viva-Voce instead of a practical prescription writing exercise. To meet Kane's IVA's conditions, the SCE will require the test criteria to address students' cognitive engagement with the underpinning NMP knowledge base and experience. The IVA results support assessment validity to accurately and appropriately address the levels necessary for maintaining advanced academic and practice standards.

6.3.4 Psychometric Item Analysis (IA): Evidencing the Test Score Inferences

A central role of AA's is to prepare students for professional practice, where knowledge application and other cognitive processes are required in clinical practice. The IRT and IA were critical in ensuring that the SCE items are designed well and structurally defined for the intended purpose. Suppose the test-item criteria are flawed and do not represent the exact processes they aimed to address. In that case, the test is not authentic (Gorin, 2006). Items without cognitive response formats do not provide opportunities to observe the KSP or the target domains of interest. If SCE's are to provide meaningful information about student ability, cognition must be assimilated into the test development process (Downing and Haladyna, 2006).

The IA function is to measure unobservable constructs, e.g. (cognitive) traits. The underlying assumption of psychometric theory (PT) is that the construct's obligatory variation is reflected in item criteria and student responses (Anastasi and Urbina, 1997). Whenever a PT measure is interpreted as a measurement of a construct, the interest is in understanding the (cognitive) mechanisms through which the item responses come about (Borsboom and Mellenbergh, 2004). The database search revealed few studies where IRT and IA measure SCE assessment/test item criteria. Where studies exist, the psychometrics validated the test instrument, and not the test-items function or attended to assessment in

nursing education (Sechrist and Berlin, 2006; Walsh et al., 2010; Fenizia et al., 2019). An element of validation involves using a psychometric test for a given purpose. In this study, the IA measurement was critical in ensuring that the SCE assessments are well constructed and that the test item criterion was defined as intended. Specifically, the inferences made about the student achievements expressed in terms of item difficulty (DIF) are so named because the DIF of test items depicts the test item's cognitive processes. Equally, item discrimination (ID) and its conditional relationship to DIF identifies test items bias toward students with a low ability (Anastasi and Urbina, 1997).

The cognitive criteria were found at low levels across the N=28 sampled SCEs, primarily because the assessment contained faulty test items which measured only novice performance and knowledge concepts. The danger is that these flaws in the design can lead to features of student performances, i.e., proficiency, being unrecognised by the scoring method; or captured by the SCE item criteria to be subsequently inferred in the passing score as achieved. In the AP assessment context, the test should measure cognitive traits, stated in the professions legislative and standards of practice documents as examples of the requisite professional qualities (NMC, 2007, 2010; RCN, 2012, 2014). The SCE structure, when examined at the IR and IA level, did not show an ideal SCE or AA picture. Moreover, a mismatch between the cognitive aspects required for practice resulted in a less informative assessment regarding the student's achievement, leading to sub-standard rote performance and, therefore, unsafe practice (Messick, 1980; Norcini, 1994; Sackett et al., 2008).

Ensuring that healthcare professionals are competent protects the institution and the service user. The analysed information organised into an argument through Kane's IVA allowed the identification of what is relevant in the assessment design. The IA results support a counterclaim about test scores, the inferences about the students' proficiency achievement, and the test scores use. Raising further alarms for the professions legislated patent safety agenda (Patient Safety Agency, 2007, 2007a; Kendall-Gallagher et al., 2011; The King's Fund, 2017).

6.4 Conclusion

This research aimed to establish the validity of the NMP assessments scores related to developing sustainable student knowledge, theory and practice- characteristics for safe, effective care. The following conclusions presented are based on this study's results.

1. The qualitative results showed an unsophisticated level of knowledge and subject-domain theory among most students WR content. Primary authors such as Barnett (1999), Eraut (2004), Barnett and Coate (2005), and Halpern (2014) denote knowledge as the acquisition of a (specialised) body of knowledge, a distinguishing feature of the profession's disciplines presumed on completion of the examination. However, this study's results imply that the level of knowledge, according to Ryle's (and others) theory, was a 'technical performance' less than expected at postgraduate level education.

2. Based on the results using Mezirow's framework for analysis, the level of reflective performance demonstrated failed to meet the intended learning outcomes. The results provide an evidentiary counterclaim to the professional educators and professional bodies assumption that graduate practitioners develop in clinical decision-making through reflection. Other studies measuring students' reflection levels revealed that the students' reflective levels and content are less than ideal (Moon, 2007; Kember et al., 2008; Gulwaldi, 2009). These authors work to support the findings in this current study

3. In applying Benner's nursing expertise model to the WR content, the results indicated that these APs were not moving toward expert-level practice. Many of the N=28 students sampled showed at best secondary level practice on Benner's trajectory. In the SCE, the performance level is unmeasured. The IRT and IA results suggest that the SCE can measure only low-levels of practice. The implication is that the WR and SCE assessments' level is commensurate with their initial registration; most of the students had not clinically advanced as a condition of their certification (Rolfe, 2014). Congruent with the increased concerns in the literature about the quality and standards of patient care, correlated to the staff's competence levels (Francis, 2013; Keogh, 2013; The Kings Fund, 2017).

4. This study's findings suggest that the academic learning achievements, characterised by BTs cognitive domain, are limited, primarily at low-level BT's cognition WRs across the four teaching and learning theories. The findings indicate little evidence of students understanding the general principles in NMP theory, subject-domain, and cognitive methods and strategies for problem-solving in this context. Barnett and Coate (2005) and Halpern's (2014) work suggest that students' demonstrate the ability to argue criticality and engage in high-order thinking through coverage of several knowledge domains. This specified engagement was not evident in many of the sampled WRs. Similarly, in the SCE, there is an expectation of criterion validity, the knowledge and cognition expectation for safe practice. The item criterion mediates the relationship between learning in professional education and practice. The NMP SCE is an authentic assessment indicative of student capability in prescribing practice. Consequently, its capacity to measure the desired higher-order thinking is limited. Therefore, it is essential to understand why a test does not relate to the criterion of interest (Gulikers, 2004; Christian et al., 2010. This study's IA results suggest that criterion situations are not addressed appropriately due to the NMP SCE rating, criteria, and design flaws.

5. Data contrary to these conclusions is found in n=8 of the N=28 sampled WR assessments. These students consistently demonstrated higher achievement levels in some aspects of theory at varying degrees across the four teaching and learning theories. Conversely, the SCEs item analysis (IA) showed no variation in these n=8 students' assessments in comparison to the remaining n=20, confirming the IA findings regarding the quality of the test items criteria. What is noteworthy in these n=8/N=28 WRs was the evidence of cognitive skills such as problem-solving and critical thinking. Also present was their metacognitive reflection and knowledge in use, resembling the desired professional and advanced practice. The results suggest that these n=8 students have moved beyond Lyotard's performativity, which characterises the novice practitioner, a prominent finding in this study's results.

6.4.1 Recommendations

Healthcare educators should not be naive about the assessments and their presumed alignment to higher-order skills instead of rote knowledge reproduction and performativity. Assessments should induce changes in the education systems at a curricular and

instructional level that nurture the test's cognitive traits, thereby changing students' learning strategy (Frederiksen and Collins, 1989; Linn et al., 1991; Miller et al., 2013). The following recommendations consider the conclusions presented to match numerically.

6.4.1.1 Recommendation 1:

Conclusions 1 and 2: Clarity about the Purpose of the Assessment.

The problem here lies in that reflection is presumed to happen on access to a reflective framework. Many frameworks are procedural and do not overtly incorporate the philosophical elements to support the students' in linking the relevance of prior knowledge and clinical experiences to written content (Fook et al., 2004, Fook, 2006: Bradbury et al., 2010). Researchers have noted a general lack of reference to strategic education and tuition; students do not act on the assignment instruction or planning in the written work. In this study, strategic education refers to reflection and the instruction to address theory and subject domains in varied clinical settings (Graham et al., 2014; Gillespie et al., 2014). This study's recommendation is a fresh conceptualisation, a transformative change teaching reflection and reflective writing within nursing education. Billett (2001) and Newton et al., (2007) propose the need to resituate reflection and focus on understanding learning rather than learning as an activity. One method to achieve this is the Ogle (1986) K-W-L cognitive model. Ogle (1986) offers three clear and decisive cognitive steps. K- what we know; W- what we want to find out; L- what we learned and still need to learn. KWL, if overlapped onto the chosen framework of refection, the opening step K has levels for students to access and question prior (theory-practice) knowledge. Students explicitly combine knowledge, critical thinking and practice awareness, thereby linking the reflective activity to the theory-practice demands. The KWL is such that the WRs content would better represent the synergy of Mezirow's cognition and knowledge for theory and practice.

6.4.1.2 Recommendation 2:

Conclusions 3 and 4: Measuring Levels of Practice and Academic Achievement.

In chapter 1 raised concern about the limits of competence. There are problems with attempts to unravel the notion of competence. However, in section 1.1.4.1, proficiency represents an integral part of a complicated relationship between cognition and applying theory to practice (Eraut, 2004; Hoffmann et al., 2014). This study results suggest mastery

in the context of Benner and Bloom's theories is better revealed and used in this research to report, explain and measure performance. In terms of the principles of theory and practice that respect the philosophy of AP. Therefore, this study recommends the dual approach of combining these above theories in assessment measurement and design and specifying the academic and practice behaviours' measurement methods known to lead to positive patient outcomes (Keutney-Lee, 2012; Audet et al., 2018).

6.4.1.3 Recommendation 3:

Conclusions *1 to 4: The Validity Issues: Reaching the Target but Missing the Mark*

The assessed score inference makes assumptions about the student's ability and performance. The danger is that the student achieves competent performance but miss the legislated NMP proficiency target and the patient safety mark. Excavating beyond the assessed score to reveal the inference (truth) behind the score takes effort. It can act as a deterrent due to its conceptual difficulty (Borsboom and Mellenbergh, 2004). Tidy methodologies for validity are wishful thinking. However, a recommendation is that VT and MTT for measurement and analysis interpret and explain the nursing education's assessed score. A further recommendation is Kane's IVA as a coherent method for interpreting the analysed results, for exploring the NMP assessment functional worth, and interpreting the clinical consequences of their use.

6.4.2 The Study limitations

6.4.2.1 An Alternative Design to the Case Study

The alternative consideration was Evaluation Research, and this design could establish the NMP program's value and inform future program improvements. The design approach would focus on evaluating outcome data (assessment) to indicate the program's success. However, how well the outcome is achieved, i.e., what students know, and have learned, is not the focus of interest in Evaluation Research. Further, the practical considerations such as economics, i.e., financial, human, and resources, the stakeholder and participant cooperation in the methodological procedures excluded this design option. Therefore, the researcher chose a case study design. A common concern is that case study designs do not follow systematised processes and lack rigour (Houghton et al., 2013; Keutel et al.,

2014). Chapter 3 addresses rigour in the methods employed in the measurement and includes the conceptual and theoretical-methodological considerations applied to the study design. The researcher acknowledges and discusses the notable limitations in the following sections.

6.4.2.2 Secondary (Post-Hoc) Data Use

Assessments intend to facilitate defensible decisions about the person assessed. In this study, the assessment data were repurposed for new use. Chapter 3 reports the assessment process regulations and how they support the data's appropriateness for answering the research question. The fundamental limitation is that the NMP assessment data were collected for a purpose different from the validity and may provide an imperfect insight into the research area of interest. The interpretation of the evidence presented for Kane's (2006) IVA may have resulted in a linear and patchy sense of student achievement. See results chapters 4 and 5. Situational awareness is a core factor influencing performance, defined as a cognitive state reliant on information correctly perceived, understood and interpreted (Cooper et al., 2016; Orique and Despins, 2018). Situational awareness, therefore, impacts how NMP students articulate the developing situation, their decision-making and the ability to anticipate prescribing problems (Stubbings et al., 2012; Orique and Despins, 2018).

Supplementing the secondary data with a focus group representative of the NMP target group could address the secondary data limitations, facilitating follow-up information and unforeseen discovery. Focus group directed 'what if' and 'what next' questions may explore students' situational awareness. Situational awareness concerns student NMP learning, and perceptions of the responsibilities inherent in resolving clinical problems may have provided insight into its impact on assessed performance. The effect of situational awareness has merit. A lack thereof leads to poor clinical decisions, omissions, or treating NMP procedures as routine, apparent in the analysed assessed work (Benner, 2009; Flin and Patterson-Brown, 2016). The focus group information may explain the study's results and enhance the rigour concerning the conclusions drawn in this study. However, this study validity argument focus is the assessed content and criterion, and it is not expected that the supplementary information will affect or change the study's findings.

6.4.2.3 Bracketing- Refocusing on the Data

The literature identifies three phases of bracketing (a) research praxis (b), abstract formulation, and (c) reintegration (Gearing, 2004). Chapter 3 addressed research praxis. The study's limitations include the above points (b) and (c). Abstract formulation bracketing focuses on the phenomenon concerning the student experience (within the bracket) and the phenomenon investigated, i.e., the NMP assessment (Gearing, 2004; Bickman and Rog, 2009). The abstract formulation allows the researcher to explore the students' assessed experience and investigate how students perceive their NMP assessment, including the critical aspects such as professional NMP development (Eden, 2014; Holmes, 2015). For example, the focus group may provide data on whether the NMP assessment develops students to operate effectively in practice (Lave and Wenger, 1999; O'Keefe and Linnenbrink-Garcia, 2014). The students' perceptions of the assessment rather than the actual assessment features may impact how they deliver the content and respond to the NMP assignments' criteria (Lizzo and Wilson, 2013; Oermann and Gaberson, 2014). The focus group data can provide the researcher with information concerning the assessments ability to allow the students to apply and integrate key curricular concepts to NMP practice. This information may help diminish the secondary data methodology's limitation by adding leverage to understand those studied. However, this study's purpose was to provide evidence of the NMP assessment content and criterion validity.

6.4.2.3.1 Contextualisation

The researcher acknowledges that the lack of contextualisation may be a study limitation concerning this study's validity questions. Students are not always clear on what is asked in the assessment (Oermann and Gaberson, 2014). The context in which the task is presented affects performance. The student may perform NMP tasks well in practice but not under test conditions. Interview or focus group exploration of the student's response to the test and their perceptions of what was expected may explain the study's findings. The limitation with secondary data is that significant contextual facets, e.g., the students' assessed behaviour and factors influencing students assessed performance, are excluded. This study aimed to ensure NMP test validity. In the methods, it was necessary to focus on the NMP assessment's content and criterion measurement and generate evidence to support the assessments fitness for purpose.

6.4.2.4 Transferability

Lack of generalisability is inherent to the case study design, particularly in this research. The study aimed not to make generalisations or transferable findings but to present a unique case with intrinsic value to a specific program (Merriam and Tisdell, 2016). Concerning educational measurement and the function of validity, the study specificity is potentially transferable to other programs (Kane, 2006, 2013). Although the NMP assessment design is harmonious with prescribing standards (NMC, 2018), it is not presumed that the study's findings are transferrable to a similar NMP context or situation. These results are specific to an NMP student cohort. However, there is transferability in a limited sense concerning measurement theory and applying the cited theories in the content and criterion analysis. This study outlines the method for using Kane's (2006) interpretive validity argument as the conceptual framework for investigating a validity problem. As a framework, the IVA has transferability to nursing and healthcare education assessments, potentially more valuable than the reported findings.

6.4.5 Contribution to Education and Suggestion for Future Research

This research highlights the importance of validity and measurement to improve educators' understanding of assessing student achievement. The application of VT and MTT provides a functional theory for educators interested in assessment measurement, giving some guidance in this respect. Kane's (2006) IVA offers an approach to synthesising analysed results making assessment and validity measurement more accessible. Suggestions for future research include replicating this research on a broader scale to aid the generalisation of the method and the results. The findings revealed that students' engagement with reflection is limited to the lower levels of Mezirow's (1981) seven concepts. A significant discovery considering the expectation from the professions registering authority that practitioners develop and practice this necessary skill (NMC, 2016). There is an impetus to change the educational strategies used for teaching reflective practice in NMP and nursing courses. A study applying and investigating the Ogle (1986) KWL model to student tuition may yield better reflective content in their assessed work.

Concerning the SCE, in its current format, it is not fit for its purpose. Adding global rating scales (GRS), including dimensions for assessing proficiency (Benner, 2001) and mastery (Bloom, 1956), may improve the test items difficulty and discrimination parameter functions. The rating scale may also prove advantageous to student learning by specifying elements and performance levels not shown on the current NMP checklist format. Further research is required to determine if the GRS modification impacts the SCE, whether it enhances the NMP students' standard of prescribing achievement, i.e., proficiency. Future studies may reveal students' understanding of what constitutes novice and expert performance (Benner, 2001) and, therefore, what the examinee must achieve for proficiency and higher. Similarly, rating scales containing the idea and procedure for mastery (Bloom, 1956).

6.4.5.1 Closing Remarks

Medical education increasingly recognises that quantitative test scores fail to capture the essential features of an examinees' performance and ability (Sullivan, 2011; Selim et al., 2012; Cook et al., 2016; Shepard, 2016). Studies in nursing education are warranted, as the findings in this study have suggested that assessment validity is not guaranteed. Chapters 4 and 5 describe the data analysis methods, show their application and how they are used to validate the NMP assessment and the assessed work. Chapter 6 discussed the findings, and under Kane's (2006) interpretive validity argument (IVA), the results indicate paying careful attention to the linkages between the curriculum objectives and students assessed work products. This study highlights a need for continuous discussion about collecting, interpreting, and using assessment data to reach valid decisions about students' abilities and performance. Kane's IVA details how nurse educators can validate their assessments to establish performance levels that map to real-world NMP and mastery. The IVA introduces to nurse education a concept for best practices in high-stakes assessment.

ENDNOTES:

1. Allergy: A hypersensitive reaction to any substance, e.g., food, pollen, medication or other, following initial sensitising contact.

(Blackwell's Dictionary of Nursing, 2002).

2. ALERT ™ (2002) is the Acute Life-threatening Events Recognition and Treatment framework, an application of a structured, prioritised system of patient assessment and management to ensure safety standards in unfamiliar clinical circumstances.

(Smith et al. 2002).

2. Red Flags: The six-stepped flag' sepsis criteria; an algorithm to identify the patients at high risk of sepsis-related death.

Dillenger et al. (2013).*Journal of Critical Care Medicine, 41 [2]*, 580-637.

3. Thrombosis: Denoting blood clots or the clotting of blood.

(Blackwell's Dictionary of Nursing, 2002).

4. Cardiovascular Complications: Minor or major functional problems related to the heart, blood vessels, or circulatory system by which oxygen and nutrients are circulated through the body's organs and tissues.

6. Procollagen type 1 N-propeptide (P1NP) is a metabolic marker for bone formation used for monitoring fracture risk in patients with osteoporosis. The test use is for monitoring treatment efficacy and compliance in patients diagnosed with osteoporosis.

(McCance and Huether, 2006).

7. Bisphosphonates: This slows the rate of bone degeneration in osteoporosis [8], reduces the risk of a broken bone by maintaining bone density.

(The British National Formulary (BNF, 2017).

8. Osteoporosis: a disease in which the bone becomes porous, weakened, prone to fracture and slow to heal.

(Blackwell's Dictionary of Nursing, 2002).

9. Myeloma: Malignant tumour of the skeleton. Tumours may be solitary or multifocal- known as multiple myeloma (MM).

(McCance and Huether, 2006).

10. Cognitive Behavioural Therapy (CBT) Work with the patients using various techniques to help patients understand their thoughts and thinking processes to modify their emotions, symptoms, ill-health experience, and behaviour.

(Boron and Bouleap, 2005)

11. Psychosis: A major mental illness of organic or emotional origin, marked by abnormal behaviour and loss of contact with reality. Can include hallucinations, delusions, diminished control of pure desires and impulses.

(Blackwell's Dictionary of Nursing, 2002).

12. Polypharmacology: Excessive or multiple pharmaceuticals use- multiple medications are in many incidences more harmful than helpful, presenting higher risk than benefit to patients.

(Beckwith and Franklin, 2011).

13. Multiple Pathologies: Describes the phenomenon of polypathology to identify individuals living with numerous chronic diseases/disorders.

(McCance and Huether, 2006).

14. Acute Kidney Injury (AKI): the preferred medical terminology for Acute Renal Failure (ARF), Associated with urine output of less than 30ml per hour (oliguria) and characterised by an abrupt reduction in renal function commonly occurring within hours. Clinical markers are reduced GFR [18] and altered serum creatinine [51] because of hypo-perfusion (low circulating blood volumes and subsequent oxygen delivery) secondary to heart failure.

(Marieb, 2012).

15. Heart Failure (HF) is an abnormality in the heart's structure and function, leading to failure to pump blood efficiently and deliver oxygen at the required rate/levels to the kidneys and other organs in the body. HF is not an actual disease but a manifestation of the disease.

(Morton et al., 2005).

16. Fluid overload: Describes excessive circulating volume; the Aetiology can be iatrogenic, i.e., resulting from intravenous fluids or as a result of organ (heart, kidney) dysfunction.

(Morton et al., 2005).

17. Serum Creatinine: Serum Creatinine Serum creatinine is a by-product of muscle activity removed from the body as waste by the kidneys. In the presence of reduced kidney function, creatinine levels rise. Serum creatinine levels are used to calculate the GFR.

(McCance and Huether (2006).

18. Estimated Glomerular Filtration Rate (eGFR) Blood filtration through the glomerulus in the Bowman's capsule in the kidney via hydrostatic pressure, which needs to remain constant. In the presence of disease, e.g. AKI, filtration is impeded.

(McCance and Huether, 2006)

19. Debridement: a thorough cleansing of a wound with the removal of all foreign matter, including injured or infected tissue. May include gentle scrubbing for exfoliation.

(Blackwell's Dictionary of Nursing, 2002).

20. Anaesthetised, Drug-induced insensitivity to pain; can be achieved by a local anaesthetic, e.g. Lignocaine or general anaesthetic, by combining drugs to induce unconsciousness.

Blackwell's Dictionary of Nursing (2002).

21. Sepsis: The result of a complex interaction between inflammation, coagulation and impaired fibrinolysis (blood clotting to prevent excessive bleeding) responds to an immune trigger usually activated by toxins released from bacteria. Sepsis is a life-threatening syndrome, the continuum beginning with sepsis, severe sepsis and then septic shock.

(McCance and Huether, 2006).

22. Fluid Resuscitation: Medical Colloquialism to treat excess fluid loss, e.g., bleeding (hypovolaemia, circulatory shock), vomiting or diarrhoea causing severe dehydration.

(Morton et al., 2005).

23. Normal Saline: Isotonic solution used intravenously to correct the fluid imbalance, e.g., dehydration or maintain hydration. The topical application includes wound, eye or nasal irrigation.

(Blackwell's Dictionary of Nursing, 2002

24. Serum Chloride: Important electrolyte, maintains the body's metabolism keeping the acid balance under control.

(McCance and Huether, 2006).

25. Metabolic Acidosis: Chemical imbalance- altered blood pH making it 'acidic' caused by elevated levels of circulating hydrogen ions. Origins are either respiratory or excrete carbon dioxide ($C0^{2}$) via the lungs; Other causes are elevated bicarbonate levels or diabetes-insulin deficiency.
(Moore et al., 2009).

26. Purulent: Containing or producing pus, a thick opaque fluid formed in certain infections.
(Blackwell's Dictionary of Nursing, 2002).

27. Erythema: An abnormal redness of the skin or mucous membranes due to capillary congestion because of trauma or infection.
(Blackwell's Dictionary of Nursing, 2002).

28. Inverted Ankle: Commonly called ankle sprain- foot turns in/outward to an abnormal degree.

(Blackwell's Dictionary of Nursing, 2002).

29. The Anterior Talofibular Ligament (ATFL): Connects the front of the talus bone to the fibula, or shin bone.

(Moore et al., 2009).

30. The Calcaneofibular Ligament (CFL): Connects the calcaneus, or heel bone, to the fibula
(Moore et al., 2009).

31. Chronic Obstructive Pulmonary Disease (COPD): A respiratory disease, one of a group, i.e., chronic bronchitis emphysema. An irreversible multi-system disorder characterised by airway obstruction.

(Oxford Handbook of Adult Nursing, 2009 p.114).

32. Broncho-constriction: An inflammatory obstruction of the small airways associated with infection other terms used, i.e., bronchiolitis.
(McCance and Huether, 2006).

33. Corticosteroids: More commonly known as steroids, work by suppressing inflammatory conditions such as asthma and arthritis.
(The British National Formulary (BNF, 2017).

34. Theophylline: Anti-inflammatory effect for treatment of asthma and COPD [31] with severe disease symptoms. It opens the airways to improve breathing by relaxing the muscles in the airways.

(The British National Formulary (BNF, 2017).

35. Monoamine Oxidase Inhibitors (MAOI): Treatment for phobias and depression or hysteria where the sedative effect is required.
(The British National Formulary (BNF, 2017).

36. Tyramine: Produced from tyrosine, a naturally occurring amino acid compound that acts as a regulator of blood pressure.
(Boron and Bouleap, 2005).

37. Diazepam: Also known as Valium, used to treat various conditions- anxiety disorders, alcohol withdrawal, sometimes with other medications to treat seizures.

(The British National Formulary (BNF, 2017).

38. Pathophysiologic: The physiology of disordered function (or functions) as distinguished by structural defect because of disease.

(Blackwell's Dictionary of Nursing, 2002).

39. Chronic Urticaria: A skin eruption characterised by smooth itchy raised weal's that are either redder or paler than the surrounding skin due to sensitised exposure to allergic substances- topical or ingested.

(Blackwell's Dictionary of Nursing, 2002)

40. Continuous Positive Airway Pressure (CPAP): A respiratory ventilation procedure used for patients with respiratory failure. The basic principle is to deliver a designated oxygen volume with each breath dependent on patient lung compliance and patient-ventilator resistance.

(Morton et al., 2005, p542).

41. Vasopressin/Vasopressor: an antidiuretic hormone (ADH) known for its role in restricting (vasoconstriction) in the small ends of the artery (arterioles), in the capillaries (tiny blood vessels). Synthetic Vasopressin is used to improve organ/tissue perfusion, which is disrupted in shock states. Septic shock is associated with vasopressin deficiency; Septic shock is treated with vasopressin [40], the second line of treatment for patients in cardiac, haemorrhagic, neurogenic and septic shock states [41].

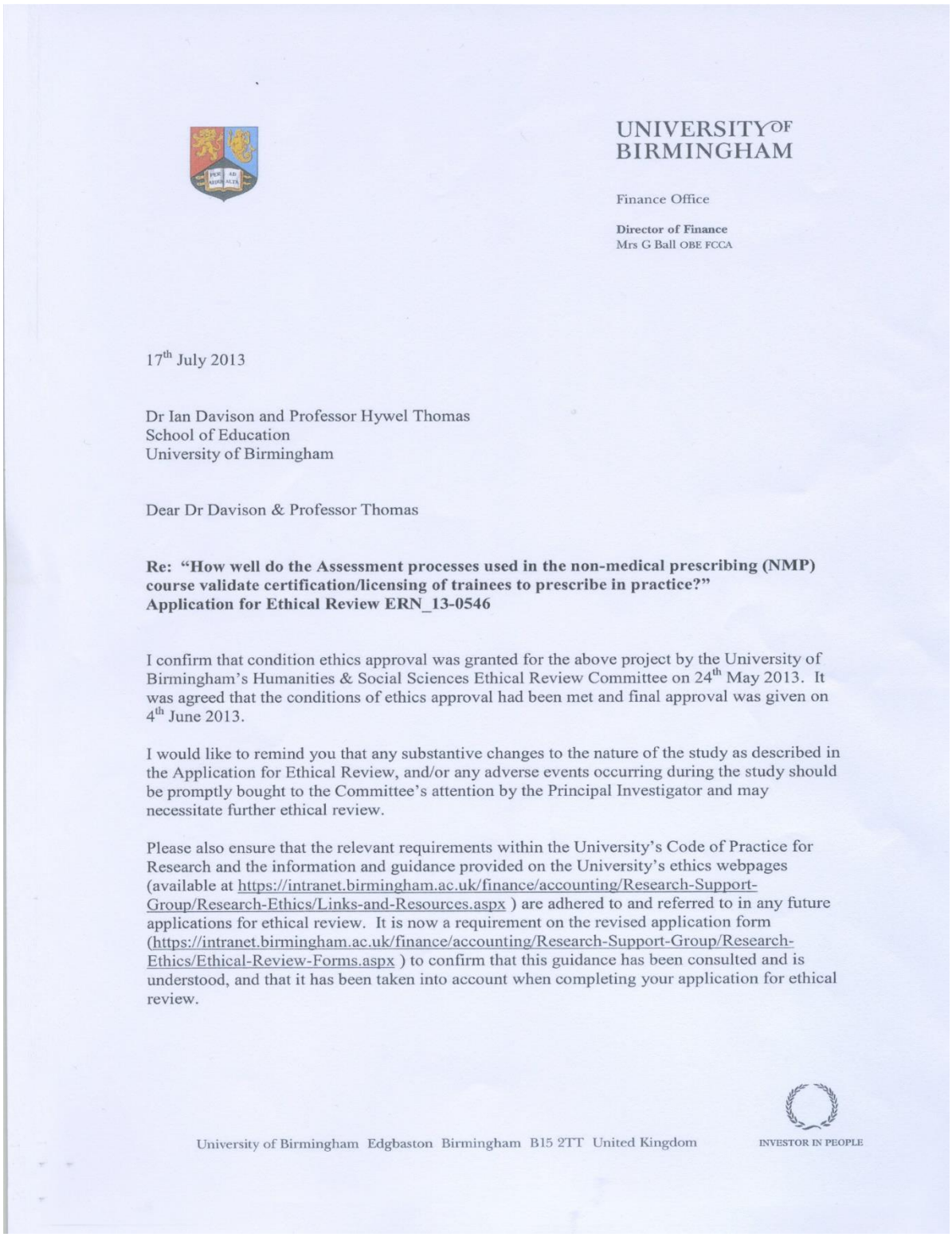
(McCance and Huether, 2006, p 1635).

42. Shocked States: Refers to an abnormal pathophysiological condition where the body's Oxygen (0^2) demands are increasing. There is a rapid consumption of 0^2. Compensatory mechanisms are initiated to meet the increased demands and restore the 0^2 and blood supply to the cells.

(McCance and Huether, 2006, p 1625).

Appendices

Appendix 1: Ethics Approval



UNIVERSITY OF BIRMINGHAM

Finance Office

Director of Finance
Mrs G Ball OBE FCCA

17th July 2013

Dr Ian Davison and Professor Hywel Thomas
School of Education
University of Birmingham

Dear Dr Davison & Professor Thomas

Re: "How well do the Assessment processes used in the non-medical prescribing (NMP) course validate certification/licensing of trainees to prescribe in practice?"
Application for Ethical Review ERN_13-0546

I confirm that condition ethics approval was granted for the above project by the University of Birmingham's Humanities & Social Sciences Ethical Review Committee on 24th May 2013. It was agreed that the conditions of ethics approval had been met and final approval was given on 4th June 2013.

I would like to remind you that any substantive changes to the nature of the study as described in the Application for Ethical Review, and/or any adverse events occurring during the study should be promptly bought to the Committee's attention by the Principal Investigator and may necessitate further ethical review.

Please also ensure that the relevant requirements within the University's Code of Practice for Research and the information and guidance provided on the University's ethics webpages (available at https://intranet.birmingham.ac.uk/finance/accounting/Research-Support-Group/Research-Ethics/Links-and-Resources.aspx) are adhered to and referred to in any future applications for ethical review. It is now a requirement on the revised application form (https://intranet.birmingham.ac.uk/finance/accounting/Research-Support-Group/Research-Ethics/Ethical-Review-Forms.aspx) to confirm that this guidance has been consulted and is understood, and that it has been taken into account when completing your application for ethical review.

University of Birmingham Edgbaston Birmingham B15 2TT United Kingdom

INVESTOR IN PEOPLE

Please be aware that whilst Health and Safety (H&S) issues may be considered during the ethical review process, you are still required to follow the University's guidance on H&S and to ensure that H&S risk assessments have been carried out as appropriate. For further information about this, please contact your School H&S representative or the University's H&S Unit at healthandsafety@contacts.bham.ac.uk.

Yours sincerely

Mrs Susan Cottam
Research Ethics Officer
Research Support Group

cc Ms Susan Blyden

University

Professor Jane Coad
Chair of Ethics Committee

27 November 2013

Dear Sir/Madam

Re: Ethical Approval - P14159

I am writing to confirm that on 5 June 2013, Susan Blyden has received ethical approval for the research project and supporting document for: *How well do the Assessment Processes used in the Non-Medical Prescribing [NMP] Course Validate Certification/ Licencing of Trainees to Prescribe in Practice?*

The research project has addressed the main ethical issues appropriately and has been approved by a member of the ethics committee at ..... University.

If you have any further queries please do not hesitate to contact me.

Yours sincerely

Professor Jane Coad

Appendix 2: Consent Form

School of Education
University of Birmingham

Direct Dial +44 (0)121 414 4866
Email: education@bham.ac.uk
Fax: +44(0)121414 4865

UNIVERSITY OF BIRMINGHAM

Date: 31st June 2013

The consent form and the right to withdraw

Study Title: How well do the Assessment Processes used in the Non-Medical Prescribing [NMP] Course Validate Certification/ Licencing of Trainees to Prescribe in Practice?

Name of Researcher: Sue A Blyden

Please tick box

1] I confirm that I have read and understood the information sheet dated 31st June 2013 for the above study ☐

2] I understand that access to my assessment documents from the NMP course and used for the research will remain anonymous and confidential, and only the results of the study will be made public. ☐

3] I understand that my agreement to have my assessment documents from the NMP course is voluntary, and that I am ***free to withdraw from the study*** at any time without giving any reasons. Students will not be penalised or subject to loss of access to current or future university courses. ☐

4] I agree that the investigators can read and use my assessment documents for the above study. ☐

…… …… ……

Student identification Number Signature ***initials only*** Date

Please return to:
Miss Sarah Wixon
Administration Office, RCXXX
RXXXXX CXXXXXX Building
JXXXXX WXXXXX
CXXXX University CXX 1XX

Researcher's Supervisor Contact Details
Dr Ian Davison. PhD
School of Education
University of Birmingham
Edgbaston, Birmingham B15 5TT
Email: i.w.davison@bham.ac.uk
0121 4144808

Appendix 3 Study Information

School of Education
University of Birmingham

Direct Dial +44 (0)121 414 4866
Email: education@bham.ac.uk
Fax: +44(0)121414 4865

UNIVERSITY OF BIRMINGHAM

Date: 31st June 2013

Study Information sheet

Dear Student

Study Title: *How well do the Assessment Processes used in the Non-Medical Prescribing Course Validate Certification/ Licensing of Trainees to Prescribe in Practice?*

To date, there has been a significant number of NMP students trained with 54 000 nurse and midwife prescribers across the UK and over 19,000 nurse independent and supplementary prescribers (Royal College of Nursing (RCN, 2012). In medicine and healthcare, assessment and exam scores are considered good grounds for practitioners' certification and licensing. However, very little attention is paid to validating examinations and assessment criteria and the subsequent scores used to certify and license healthcare professionals. The validity of assessment rests on whether the tests/assessment measure what they intend to measure. Whether the interpretation of tests or assessment scores is an accurate or a true representation of student proficiency and whether the decisions resulting from the test scores' interpretation are appropriate. Validity tests attract much attention in medical and clinical research. However, such tests are rarely applied to healthcare courses assessment scores.

What is the purpose of the Study?

This study is a University of Birmingham research project and will form part of the researcher's Doctoral thesis. This study aims to find out how well the assessment processes used in the Non-Medical Prescribing (NMP) course validate the Certification/ Licensing of trainees to prescribe in practice. The study aims to carefully look at the test/assessment score data to reveal the practical value, merits and limitations of the tests/assessments used in the NMP program. The study further aims to look at how well the NMP course meets the specified education objectives.

What does the study involve?

Your NMP course assessment documents will be viewed to collect information for the study. *Firstly,* the Multiple-Choice Questions and Short Answer Question exams for testing numeracy and pharmacology. *Secondly*, records of your patient health assessment and history assessment performance will include the scenario-based assessment and the practice assessment documents. *Finally, your work samples*, such as your reflective essays; plus, copies of your written prescriptions from your practice assessment and portfolios, which you have handed in as part of the NMP assessment process.

What are the risks of taking part?

The risks are low; you will not be contacted by the researcher for information about your course. Your assessment documents will be identified from the university database using your unique identifying number (SiD); your identity will remain confidential. This means that the researcher will not know your details, such as your name or any other information used to identify you. The information about your assessment or any information in your assessment documents will not be given to anyone, including your employers and will be used only for this study. Paper copies of your assessment documents will be accessed through the administrative staff from archives using your SiD given on the consent form. Any information taken from your assessment documents will be coded for use in the study so that the information is not traced back to you. You may refuse to participate or withdraw participation during the study without penalty or prejudice, including losing access to future university courses.

Are there benefits to taking part in the study?

Education is recognised as an integral component in improving standards of care. Participating in the study will enable the university to provide Research evidence to support NHS investment in our courses. I.e., value for money about NMP course. The purchase and potential evidence that the NMP course is 'fit for purpose' in providing post-qualification staff to deliver safe, quality patient care.

What happens to the information once the study is completed?

Once the study is completed, the information collected from your assessment documents and used for the research will be destroyed. You can have a copy of a summary report about the study and what was found if you wish, but you must contact the researcher in writing for the report using the contact details below.

What do I need to do if I want to participate in the study?

For entry into the study, you will need to read the consent form and return it to the address at the end of the form. You must use your student Identification [SId] number on the consent form. ***DO NOT*** use your name as your assessment documents can only be identified using your SId. This is to ensure that your details, name, address etc., remain strictly confidential, that the Data Protection Act uses the information in your assessment records (1998) and the results of the study stored as required by the Privacy and Electronic Communications [EC Directive] Regulations (2003).

What do I do if I change my mind about being in the study?

You are within your rights to change your mind about allowing your assessment documents to be used and withdraw from the study even after signing and returning the form. If you choose to withdraw, you will not be refused the right to apply for further study at this university, nor will withdrawal from the study affect your course now or in the future. If may withdraw, but you must contact the researcher in writing within 14 days using the information below.

You must use your SId in the letter to the researcher, ***DO NOT*** use your name. You can only be identified using your SId. If you withdraw any information taken from your assessment documents and used in the results, it will be destroyed.

If I want more information about the study?

If you require further information about the study, please contact the researcher:-

Miss Sue Blyden
Senior Lecturer in Adult nursing
Faculty of Health and Life Sciences
---------- University
----------. XV1 XFB

Researcher's Supervisor Contact Details

Dr Ian Davison PhD
School of Education
University of Birmingham
Edgbaston
Birmingham. B15 2TT

Appendix 4: The Written Reflection, Assignment Brief

M42CPD Non-Medical Prescribing (Masters 1) Summative

Coursework 3: A 2500-word critical reflective essay.

The critically reflective essay will relate to the student's area of practice and will demonstrate that they have learnt and can apply the principles of prescribing within their role. Students are expected to utilise a model of reflection as a framework for the structure of their essays.

If you are unclear as to the area of prescribing practice you wish to consider, please discuss this with the module team.

The essay is required to demonstrate **learning outcomes 1-8** and should particularly focus on the following

- Consultation, decision making and therapy, including referral
- Prescribing decisions and medications choices, including pharmacology
- Influences on and psychology of prescribing
- Evidence-based practice, clinical governance in relation to non-medical prescribing

Rationale

The purpose of the reflective essay is to allow the demonstration of a range of knowledge gained which underpin safe, effective prescribing implicit in the case you have chosen

A model of reflection should be chosen, e.g. the Gibbs model, and used as a framework for writing up the reflective case. You are **not** expected to discuss the reflection, the model selected, or why you have chosen it since the wordage allocated is limited. However, it is important that you use theory against which to reflect and analyse your experiences. The written structure must be coherent and cohesive. See separate guidelines on reflection on the Moodle platform

Front sheet

Please include this, guidance is found in the **student handbook**

Introduction

- Briefly state the purpose of the essay and what prescribing area is explored
- The context of your prescribing should be holistic in nature, taking into account the context of the patient (social and wider) as appropriate.
- It should demonstrate a systematic approach to consultation and diagnosis of the patients' condition and any requirement for onward referral.

The Body of the Assignment

- Using an identified reflective model, critically analyse the prescribing assessment carried out in practice
- The following questions may help. What went well, and why? How did your knowledge base influence the process?
- The scope of your role in the prescribing assessment must be clear, and evidence of role development

The Conclusion

A clear conclusion must summarise your main points and indicate how this module has developed your practice. No new material must be introduced in this part of the essay

Teaching, Learning and Assessment

Intended Learning Outcomes: On completion of this programme, the student must be able to demonstrate achievement in the following areas

1. Undertake a holistic assessment and consultation with patients/clients, clients parents and carers

2. Conduct and evaluate a thorough medications history, including current medications (over the counter, alternative, complementary health therapies) to inform diagnosis

3. Undertake an appropriate clinical history, utilising clinical assessment skills and make an appropriate decision based on that assessment to either diagnose or refer, having considered the legal, cognitive, emotional and physical differences-related to children or adults

4. Critically examine, evaluate and apply relevant research and legislation to the practice of nurse/midwife/AHP prescribing, critically appraise and use sources of information/advice and decision support systems in prescribing practice

5. Compare, analyse and evaluate the influences that can affect the prescribing practice, and demonstrate an understanding of managing a prescribing practice in an ethical way

6. Appraise and apply knowledge of drug actions in prescribing practice

7. Evaluate and compare the roles and relationships of others involved in prescribing, supplying and administrating medicines safely, appropriately and cost-effectively

8. Justify and apply the framework of professional accountability and responsibility

Appendix 5: Structured Clinical Assessment, Assessment Brief

Guidance for Objective Structured Clinical Assessments (OSCE)

The OSCE is a means of assessing clinical competency based on testing through direct observation. It is a multipurpose evaluative tool examining areas most critical to performance, e.g., the ability to obtain/interpret data and problem solve. The purpose of the examination is to determine clinical competence and safe prescribing. The standard for non-medical prescribing is Safe, Appropriate, Economical and Effective (The Royal Pharmaceutical Society (RPS, 2016).

Marking Guidelines

The marking Table guideline is shown on page 34 of the document. It demonstrates competencies you must consider relevant to the clinical situation. The Table is the actual assessment form and shows how the student in the clinical situation must demonstrate these competencies.

Assessment

The student will be given a clinical scenario related to your specialist practice area. The student may be asked about only certain parts of the process. You are expected to complete a whole episode of care for the patient from diagnosis to treatment and follow-up either independently or in conjunction with other healthcare professionals. The examiner may intervene to offer reassurance where there is hesitation for long periods, and any prompts will be non-directive.

To obtain an overall pass, the student must achieve a tick in each section's “stated” column. If this does not occur, the student can recover specific skills at the end of the examination. The student MUST prescribe/write a sample prescription as part of the clinical examination.

If the student does not achieve the required result, a fail result is given. If this occurs, please inform the module leader. The student may be issued a period of 6 weeks before retaking the clinical exam. If a further assessment is required, it will be in the same manner as the first attempt. **The student must achieve a Pass result on the second attempt**.

Student initials:..... Student Number:.....

Consultation skills and history taking Criteria	Stated	Incomplete answer	Not stated or incorrect	Recovered within VIVA	PASS/FAIL
Greets patient					
Introduces self and role					
Identifies patient's reason for attendance					
Gains the patient/client's consent					
Identifies patient's concerns and expectations					
Uses appropriate verbal and non-verbal communication skills to elicit information					
Seeks clarification appropriately and checks patients understanding					
Demonstrates respect for patient/client views					
Demonstrates an efficient and logical approach to information gathering					
Uses skills of negotiation in order to achieve concordance					
Demonstrates an awareness of the patient's ability to concord with the treatment plan					
Closes consultation appropriately					

Decision Making and Rationale Criteria	Stated	Incomplete answer	Not stated or incorrect	Recovered within VIVA	PASS/FAIL
Has reached a decision, including consideration of any potential differential diagnosis.					
Recognises the need for further examination or investigation					
Has considered the practical aspects of the treatment plan whether this is suitable for the patient					
Identifies further information required for holistic care					
Has identified the possible cause/causes					
Makes a treatment choice based on patient preference or in agreement with the patient					
Recognises the limits of personal competence and makes					

recommendations for referral where appropriate					
Has considered the relevance of all other treatments/medication the patient has already taken or is taking					
Has considered the relevance of the patient's medical history					
Has considered possible cautions / contra-indications or previous ADR's					
Makes a treatment choice based on cost and clinical effectiveness					
Explains the potential side effects and provides information to the patient					
Uses sources of information available appropriately (NPF / BNF)					
Identifies that the patient has a central role in decision-making					
Has recognised the need to ensure the patient understands the information given.					

Criteria	Completed	Incomplete or Not completed
Name		
Address		
Contact Details		
Age		
Date of birth – only legally required if a child under 12 years		
Name of drug/product		
Appropriate drug/product		
Contact details		
Quantity		
Dose		
Frequency		
Duration of treatment (or as required)		
Directions for use		
Prescribers contact Details		
Signature		
Date		
Blank space underscored		

Appendix 5a: Structured Clinical Assessment, Work Sample

NON-MEDICAL PRESCRIBING

OBJECTIVE STRUCTURED CLINICAL EXAMINATION (OSCE) & VIVA-VOCE

Student Name ...M.....K.....

Student Number.....

Markers Name.....H.....R.....

Internal moderator.....S.....C.....

PASS/FAIL
PASS

The OSCE is PASS/FAIL. The criteria highlighted in grey are essential requirements to pass the OSCE.
However, if a registrant fails to answer correctly any question that may result in direct harm to a patient/client they must be referred [NMC, 2006].

Students have the opportunity to recover any areas they feel they have not fully explored within the VIVA-VOCE

	PASS/FAIL
Consultation skills and history taking	P
Decision Making and Rationale	P
Prescription Writing	P

Comments: As a community matron I would expect a more holistic approach to the care of this patient. Checking compliance and concordance, follow up need to be improved. INR check

Student Name: M..... K..... Student Number:.....

Consultation skills and history taking Criteria	Stated	Incomplete answer	Not stated or incorrect	Recovered within VIVA	PASS/FAIL
Greets patient	✓				
Introduces self and role	✓				
Identifies patient's reason for attendance	✓				
Gains the patient/client's consent	✓				
Identifies patient's concerns and expectations	✓				
Uses appropriate verbal and non-verbal communication skills to elicit information	✓				
Seeks clarification appropriately and checks patients understanding		✓			
Demonstrates respect for patient/client views		✓			
Demonstrates efficient and logical approach to information gathering	✓				
Uses skills of negotiation in order to achieve concordance		✓			
Demonstrates an awareness of the patient's ability to concord with treatment plan		✓			
Closes consultation appropriately	✓				

Decision Making and Rationale Criteria	Stated	Incomplete answer	Not stated or incorrect	Recovered within VIVA	PASS/FAIL
Has reached a decision including consideration of any potential differential diagnosis.	✓				
Recognises the need for further examination or investigation	✓				
Has considered the practical aspects of the treatment plan whether this is suitable for the patient	✓				
Identifies further information required for holistic care		✓			
Has identified the possible cause/causes	✓				
Makes a treatment choice based on patient preference or in agreement with the patient		✓			
Recognises the limits of personal competence and makes recommendations for referral where appropriate	✓				
Has considered the relevance of all other treatments / medication the patient has already taken or is taking	✓				
Has considered the relevance of the patients medical history	✓				
Has considered possible cautions / contra-indications or previous ADR's	✓				
Makes a treatment choice based on cost and clinical effectiveness	✓				
Explains the potential side effects and provides information to the patient		✓			
Uses sources of information available appropriately (NPF / BNF)	✓				
Identifies that the patient has a central role in decision-making		✓			
Has recognised the need to ensure the patient understands the information given.		✓			

Sample of what?

Student Name: M[illegible] K[illegible] Student Number: .....

Now that you have made a prescribing decision please write a legal prescription for your patient/client from your given scenario:

Criteria	Completed	Incomplete or Not completed
Name	✓	
Address	✓	
Age	✓	
Date of birth – only legally required if child under 12 years	N/A = 8	
Name of drug/product	✓	
Appropriate drug/product	✓	
Contact details	✓	
Quantity	✓	
Dose	✓	
Frequency	✓	
Duration of treatment (or as required)	✓	
Directions for use	✓	
Signature	✓	
Date	✓	
Blank space underscored	✓	

Appendix 6: TDTA Codes, NMP Themes and Categories

TDTA Core (theme) Code	TDTA Code Categories	RQDA CODES
01 – The Consultation	1. Allergies	1_
	2. Adverse Drug Reaction [ARD] Reporting	2_
	3. Clinical Effectiveness	3_
	4. Consent	4_
	5. Consultation	5_
	6. Cost-Effectiveness	6_
	7. Disease Severity Assessment	7_
	8. Drug Interactions	8_
	9. Identify Desirable/Undesirable Outcomes	9_
	10. Knowledge of NPF Formulary	10_
	11. Knowledge of Product	11_
	12. Medical Conditions- natural progress	12_
	13. Medical History	13_
	14. Medications History	14_
	15. Medicine Altered by, e.g., age etc.	15_
	16. Medicines Licencing, monitoring, supply	16_
	17. Misuse Potential of Drug	17_
	18. Over the Counter [OtC] Medications	18_
	19. Pharmacokinetics of Medicines	19_
	20. Precautions- Contra indications	20_
	21. Prescribing Safely	21_
	22. Presenting Condition	22_
	23. Principles of Evidence-Based Practice	23
	24. Public Health Medicines Use	24_
	25. Drug Mode of action	99_

TDTA Core (theme) Code	TDTA Code Categories	RQDA CODES
02 Establishing Options		
	1. Physical Examination	25_
	2. Differential Diagnosis	26_
	3. Effects of Multiple Pathologies	27_
	4. Medications, Contraindications on Treatment Options	28_
	5. Establishes and Maintains Plan for Reviewing	29_
	6. Holistic Needs Assessment	30_
	7. Monitors Effectiveness of Treatment	31_
	8. Most Appropriate Treatment	32_
	9. Working Diagnosis	33_
	10. Requests, Interprets Relevant Diagnostic Tests	34_

TDTA Core (theme) Code	TDTA Code Categories	RQDA CODES
	1. Adapts Consultation Style	36_
	2. Checks Patient Understanding of Treatment	37_

03 Communicating with Patients	3. Clear Instructions	38_
	4. Communicates with patient	39_
	5. Communication Skills	40_
	6. Cultural, religious Implications of Prescribing	41_
	7. Deals Sensitively with Patients	42_
	8. Existing Medications, Contraindication on Treatment Options	43_
	9. Explains Nature of Patient Condition, Treatment	44_
	10. Helps Patient- informed Choices	45
	11. Listens to Patients Beliefs, Expectations	46_
	12. Negotiates Outcome, Patient-Prescriber	47
	13. Potential Risk-Benefit of Treatment	48_
	14. Referral- Preventive Measures	49
	15. Non- Drug Treatment Option	50_

TDTA Core (theme) Code	TDTA Code Categories	RQDA CODES
04- Prescribing Safely		
	1. Knows Limits of Knowledge	51_
	2. Checks Dose and Side Effects	52_
	3. Checks Dose Calculations	100
	4. Knowledge of Drug Actions	53_
	5. Knows Common Medications Errors	54_
	6. Accurate, Timely Record Keeping	55_
	7. Knows when to Refer	56
	8. Maintains Competence, Confidence	57_
	9. Prescriptions meet Legal Requirement	58_
	10. Seeks Guidance	59_
	11. Writes Legible Clear, Complete Prescriptions	60

TDTA Core (theme) Code	TDTA Code Categories	RQDA CODES
05- Prescribes Professionally		
	1. Accepts Personal Responsibility in Prescribing	61_
	2. Understand legal Implications	62_
	3. Understands Current Legislation, prescribing	63_
	4. Makes prescribing Decisions, Patient	64
	5. Safety Concerns Related to Prescribing	65_
	6. Safety of Prescription Pads	66_
	7. Prescribes within Current Professional Code	67_
	8. Up-to-date with Advances in Practice	68

TDTA Core (theme) Code	TDTA Code Categories	RQDA CODES
06- Improving Prescribing Practice		
	1. Establishes Professional Links	69_
	2. Learns & Changes Practice	70_
	3. Participates in Clinical Supervision	71_
	4. Reviews, Reports Prescribing Errors	72_
	5. Understands how to Improve Prescribing	73_
	6. Willing to Share, Debate prescribing Practice	74_

TDTA Core (theme) Code	**TDTA Code Categories**	**RQDA CODES**
07- Information in Context		
	1. Understands Advantages/Limitations of Information Sources	75_
	2. Uses Up-to-date Information- written, verbal, electronic	76_
	3. Critically appraises the Validity of Information-research, reports etc	77_
	4. Applies Information in Clinical Context	78_
	5. Uses Relevant Record Systems	79_
	6. Uses relevant Prescribing and Information Systems	80_
	7. Clinical EBP Application	81_

TDTA Core (theme) Code	TDTA Code Categories	RQDA CODES
08- The NHS in Context		
	1. Local/ National Drug Budgetary Constraints,	82_
	2. Works with NHS Organisation, Code of Conduct	83_
	3. NHS in Context	84_
	4. Uses Appropriate formularies	85_
	5. Understands NHS Frameworks, Medicines Use- NICE, NSF,	86_
	6. Medicines Management and Clinical Governance	87
	7. Works within Frameworks for Medicine	88_

TDTA Core (theme) Code	TDTA Code Categories	RQDA CODES
09 -The Team and Individual Context		
	1. Appropriate Support Network	89_
	2. Ensures Continuity of Care, not compromised	90_
	3. Establishes Relationships with Colleagues	91_
	4. Improves practice	92_
	5. Negotiates appropriate Support Levels for Role	93_
	6. Not Compromise Continuity of Care	94_
	7. Prescribes professionally	95_
	8. Provides support, Advice- others	96_
	9. Recognises and Deals with Pressures, Prescribing	97_
	10. Understands, Respects, other Roles	98_
	11. Works with Colleagues	101_

Appendix 6a: Ryle's (1949) TDTA Codes: Knowledge Theory

Knowledge Type		Knowledge Acquisition, Interpretation, and Outcome	RQDA CODES
Declarative (Know-That)	Knowledge; information recall, deals with specifics, terminology/facts	I. The cognitive process concerning memory, factors that influence performance. II. Helps the individual in making the transition from one problem to another; contains sequence actions/performance that are domain-specific. III. The individual uses explicit and defensible (know-that) knowledge to provide actioned sequenced solutions (know-how).	**DK**
Procedural (Know-How)	Ability to translate, interpret and explore	I. Task knowledge - the psychological requirements of learning tasks; retrieval of relevant rules and information application. II. Prior (task) knowledge	**PK**
Strategic (Know-Why)	Know-why presupposes the existence of knowledge at lower levels I-II.	I. To have such a theory is to have the terminology without the need to labour the importance of proper and known language/terminology. II. Not seeing the theory in isolation, but about situations dependent upon factors [I] already identified. a) the prior existence of a more elementary type of theory b) cause and effect relationship; a prerequisite for situation producing theory III. Significant for practical activity in the light of theory, assessment, construction and refinement of theory. Research that explores the validity of theory or research geared to stimulate thinking toward theory.	**SK**

Appendix 6b: Mezirow's (1981) TDTA Codes: Seven Concepts of Reflection

Seven Concepts of Reflection		RQDA Code
Me-1-Reflectivity	Identified as our ordinary consciousness is described as recognition of our habits, how we think or act, a specific behaviour.	114_
Me-2-Affective Reflectivity	Being aware of how we feel about the way we are thinking or acting and our habits. An ability to assess the efficacy of thought, actions habits and ways of doing things. Ability to identify causes for thought and actions and what our relationships are in the situation.	115_
Me-3-Discriminant Reflectivity	Assessing the efficacy of our perceptions, thoughts actions and habits; of ways doing things and functioning, identifying relationships in the situation	116_
Me-4-Conceptual Reflectivity	Self-reflection which may (may not) lead to critical questioning and intellectual argument/debate.	117_
Me-5-Judgemental Reflectivity	Making value judgements about our thoughts, actions and habits in terms of them being positive or negative liked or disliked, beautiful or ugly is considered. Here Mezirow (1981), also ascertains another form of reflectivity '*Conceptual*' where the ability to critique and question the constructs (political, cultural, social, economic) are used.	118_
Me-6-Psychic Reflection	Leads to the ability to recognise in oneself the habit of making judgments on limited information.	119_
Me-7-*Theoretical Reflection*	Identified as an awareness of the reason for the habit of making judgements or assumptions. Or for conceptual inadequacies – taken for granted assumptions- (central to transformative learning).	120_

Appendix 6c: Benner's (2001) TDTA Codes: Novice to Expert Model

Levels	Model Descriptor	RQDA CODES
Be 01- Nov= Novice Applies template of rules, minimal knowledge, limited or no information processing	Learns to recognise various objective facts and features relevant to the clinical skill. Dependent on rules for determining clinical action based on those facts, which are sometimes reorganised and applied context-free, without reference to the overall situation in which they occur.	102_
Be_02-Ad-Beg= Advanced Beginner Good working knowledge of rules, Limited information processing, understanding of subject and area of practice	Sometimes recognise and uses various objective facts and features relevant to the clinical practice. It is primarily context-free, some deliberate conscious application of rules, accountable only to those rules, not the results.	103_
Be_03-Compt= Competent Applies rules with some limited breadth and depth of information processing moderate understanding of discipline and area of practice	Context-driven application of rules; a sense of what is important is missing. Sees a complex situation as a set of facts, concludes, problem-solving based on set rules. Accountable to those rules, choice of action based on those rules, recognition of the importance of the results without conscious thought.	104_
Be_04-Profic= Proficient Information processing and knowledge application based on rules. Know-how is a situational, i.e. recall of similar situations	Applies rules for decision making; choice and deliberation are contextual. Detached decision–making derived from the rule-like calculations, anticipates and solves problems based on previous similar events.	105_
Be_05- Expr= Expert Authoritative knowledge of rules and discipline, deep tacit understanding of the subject, autonomous across most areas of practice	Internalises rules, applied holistically to view and grasp of complex situations, context-dependent moves between intuitive and analytical approaches with ease. Critically reflective, takes responsibility for going beyond existing rules /standards, interpretive evidence-based knowledge application, derives alternative approaches to complex practice	106_
Be_06- MA= Master Autonomous across areas of practice can instantaneously process knowledge; consistent in producing appropriate perspectives and associated action plan	No rules or formulae applied recognises a potentially new perspective that better explains past events and better dictates future actions. Abstract reasoning, practice and concrete experience. Subject to deliberation, senses opportunities beyond that which can be seen.	107_

Appendix 6d: Bloom's (1959) TDTA Codes: Taxonomy of Educational Objectives

The Taxonomy		RQDA CODES
Bl_01-Knowl= Knowledge	Knowledge; information recall, ways and means of dealing with the specifics/ facts	108_
Bl_02-Comph= Comprehension	Ability to translate, interpret and explore	109_
Bl_03-Appli= Application	Apply principles, theories or other abstracts to new situations; relate principles and draw conclusions	110_
Bl_04-Anal= Analysis	Material is broken into its constituent parts and detecting how the parts are organised, related to one another in an overall structure or purpose.	111_
Bl_05-Synth= Synthesis	Reconstruction of previous knowledge, experience; Derivation of abstract, new material and experience into an integrated theoretical, methodological /framework or whole	112_
Bl_06-Eval= Evaluation	Make judgements about the value of ideas, theoretical/ methodological framework, solutions; Use criteria and standards for apprising quality, effect or accuracy.	113_

Appendix 6e. The Written Reflection, Assessed Coded Sample.

St 12 Created by RQDA

The purpose of th s ass gnment s to cr t ca y ref ect on a pract ca prescr b ng assessment was asked to rev ew a deter orat ng pat ent and make suggest ons for an ongo ng p an of care By /<< 1-05-Consultation | ref ect ng on the consu tat on | 1-05-Consultation >> | assessment d agnos s therapy referra and fo ow-up p an w demonstrate pr nc p es of prescr b ng have been earnt and app ed Prescr b ng dec s ons and cho ce of med cat on nc ud ng pharmaco ogy w be ana ysed ref ect ve y Cons derat on w be g ven to nf uences on and psycho ogy of prescr b ng together w th ev dence-based pract ce n re at on to non-med ca prescr b ng

| << DK, know-that | w use John (1995) mode of ref ect on wh ch s based on f ve cue quest ons enab ng me to break down my exper ence and ref ect on the process and outcomes | Me-1-Reflectivity >> | work as an Advanced Nurse Pract t oner w th n a sma D str ct Genera Hosp ta as pa t of the Cr t ca Care Outreach Team Part of my ro e enta s /<< 9-11-Works-ic-colleagues | tak ng referra s from nurs ng staff and med ca co eagues us ng a track and tr gger system ca ed Mod f ed Ear y Warn ng Score (MEWS) | 9-11-Works-ic-colleagues >> | as descr bed by Harr son and Da y (2006) regard ng deter orat ng pat ents n an acute sett ng had been asked to rev ew a pat ent who had a h gh MEWS /<< 7-07-clin-EBP-applic | Fo ow ng recommendat ons n the Comprehens ve Cr t ca Care Report the Department of Hea th (DoH) (2000) ntroduced the concept of us ng Ear y Warn ng Scores DK, know-that Th s report had out ned the concept for serv ce de very to pat ents based on need rather than oca ty - cr t ca care w thout wa s | 7-07-clin-EBP-applic >> | The a m of Outreach be ng to extend cr t ca care serv ces outs de the usua phys ca m ts of an ntens ve care un t and act as a serv ce and educat ona pa tnersh p between ntens ve care and genera wards (eary and R d ey 2003) |<< 7-07-clin-EBP-applic | | DK-know that>> | The Ear y Warn ng Score (DoH 2000) s the ca cu at on of an aggregate tr gger score based on phys o og ca abnorma t es of heart rate b ood pressure resp ratory rate temperature ur ne output and eve of consc ousness t s ntended to support ob ect ve dec s on mak ng to he p staff dent fy deter orat ng pat ents | 7-07-clin-EBP-applic >> | Be-1-Novice >> | PK, Know-how>> | (Carberry 2014)

Before assessed the pat ent checked the pat ent s MEWS score to see the r atest observat ons and the r trends B ck ey (2003) suggests that nterv ew ng pat ents to obta n a hea th h story requ res p ann ng w th the necess ty to rev ew any med ca records and set goa s for the nterv ew By rev ew ng any records and cha ts | 1-13-Medical-History >> | we can gather nformat on to deve op deas about what to exp ore w th the pat ent De Mar Doust and G asz ous (2006) are of the op n on that /<< Be-1-Novice | nurses soon d scover that more nformat on s der ved from the h story tak ng than the exam nat on wou d agree n fact the greater part of my ob s often taken up by read ng | Be-1-Novice >> | through the notes and tak ng the h story | << 2-01-Physical-Exam | Th s ed to a fa r y short succ nct focussed phys ca exam nat on | 2-01-Physical-Exam >> | wh ch was a that was necessary for me |<< 2-02-Diff-diagnisis | to arr ve at the d fferent a d agnos s | 2-02-Diff-diagnisis >> | and deta a p an of care Rhoades (2006) appears to concur when stat ng met cu ous and comprehens ve do not necessar y mean engthy The sk ed c n c an str ves to record a essent a c n ca data conc se y and to document the c n ca dec s on-mak ng that under es d agnost c and treatment dec s ons | DK-know that>> | (Genera Med ca Counc 2006)

DK-know that>> | The ob ect ves are to prov de effect ve commun cat on to a careg vers to ensure cont nu ty of h gh qua ty care for the pat ent to m n m ze ega vu nerab ty for the prov der | 3-05-communication-skills >> | Prescr pt ons are a way of commun cat ng the pat ent s persona deta s and the spec f c treatments prescr bed by an author sed reg stered profess ona (Joyce and Carrera 2011) |<<DK, know-that |Med c nes can do a great dea of good but a so have the ab ty to cause harm |<< 4-05-Common-drug-errors | Med cat on errors are one of the most common causes of pat ent harm and prescr b ng accounts for a arge proport on of med cat on errors | 4-05-Common-drug-errors >> | (The Hea th Foundat on 2012) Predom nant y the mode s of assessment used w th n Cr t ca Care are the med ca mode (Barnum and Kerfoot 1995) and the ABC approach to hea th assessment (Resusc tat on Counc UK 2011) | 1-07-Disease-Severity-Ass >> | Some w argue that |<< 8-07-work-in-medical-Fwk | the med ca mode w th ts focus on contro ng the hea th care system and ts b as towards curat ve med c ne can be nadequate and expens ve | 8-07-work-in-medical-Fwk >> | (H ggs and Edwards 1999) /<< 8-07-work-in-medical-Fwk | | DK-know that>> | The med ca or d sease mode tse f s the very

process of reduc ng a pat ents and a nesses regard ess of the nature of the nesses to the d agnost c names and treatments of the d sordered pa ts | 8-07-work-in-medical-Fwk >>| We become enmeshed n our own mode even when t does not work Barbour (1997) states that the |<< 8-07-work-in-medical-Fwk| mode s abso ute y essent a for the organ c d sease to wh ch a sc ent f c approach does app y | 8-07-work-in-medical-Fwk >>| and t he ps us to understand and re eve some of the |<< 8-07-work-in-medical-Fwk| symptoms of funct ona d sorders but when t s extended as a str ct y b omed ca approach | 8-07-work-in-medical-Fwk >>| to the many nesses caused predom nant y by persona s tuat ons the core prob em s bypassed n the process of reduct on

The prob em cou d e not so much w th the mode as w th the c n c ans who use t |<< 1-13-Medical-History| The pat ent s h story s not ust about symptoms and d sease but nc udes a persona and soc a survey and emot ona and psycho og ca prob ems are not exc uded | 1-13-Medical-History >>| There s no ru e aga nst understand ng the human s tuat ons that cause so much ness and the med ca mode does not proh b t warm persona sens t ve nteract ons between c n c ans and pat ents | 3-07-sensitive-with-patient >>| However when the med ca mode s app ed n actua pract ce t can be far more restr ct ve w th the c n c an be ng overwhe med w th the demands of modern care |<< DK-know that>>| | w th tt e t me to know the pat ents as persons Ref ect ng back on the pat ent assessment be ng d scussed | 2-06-Holistic-Needs-Ass >>| |<< 1-07-Disease-Severity-Ass| noted th s pat ent had severa symptoms | 1-07-Disease-Severity-Ass >>| and |<< 1-20-precautions-contra-ind| to prescr be and treat one symptom cou d be detr menta to the prescr pt on of drugs for another symptom |<< Me-1-Reflectivity| n part cu ar th s pat ent has Heart Fa ure and may requ re Frusem de to be prescr bed to off- oadf u ds but a so has an Acute K dney n ury and therefore needs f u ds prescr b ng | Me-1-Reflectivity >>| These prescr pt ve treatments wou d appear to conf ct | 1-20-precautions-contra-ind >>| and th s wou d appear to support the theory that the c n c an needs to know the nd v dua pat ent n order to suggest the best treatment opt ons

|<< 2-01-Physical-Exam| The use of the ABC approach n pat ent assessment s w despread on resusc tat on courses (Resusc tat on Counc UK 2011) |<< Be-2-Ad-Beg| The rat ona e s that prob ems that more qu ck y ead to pat ent death are dea t w th and stab sed f rst | 2-01-Physical-Exam >>| Be-2-Ad-Beg >>| The fact that the abbrev at on for the approach cons sts of the f rst three etters of the a phabet s fo tu tous |<< Me-1-Ref ect v ty |<< B -2-Comph | C ass ca med ca teach ng promotes the use of the med ca mode when assess ng pat ents Th s mode works we when estab sh ng that |<< 2-09-Work ng-D agnos s | the correct d agnos s s the ma or prob em and when there s no urgency n start ng treatment | Me-1-Reflectivity >>| Bl-2-Comph >>| t works ess we dea ng w th cr t ca y pat ents n the f rst nstance | 2-09-Working-Diagnosis >>| | |<< 1-13-Medical-History |<< Bl-2-Comph| My pat ent had n t a y presented to hosp ta w th a ke y chest nfect on or commun ty acqu red pneumon a two days prev ous y They had presented w th a h gh temperature tachycard a h gh resp ratory rate and product ve cough | 1-13-Medical-History >>| Bl-2-Comph >>| |<< 1-22-presenting-condition| nd cat ng they were Seps s pos t ve | 1-22-presenting-condition >>| (Surv v ng Seps s Campa gn 2012) |<< 1-23-Principles-EBP| The Surv v ng Seps s Campa gn (2012) deta s a gu de ne ca ed the Seps s S x pathway wh ch s the recommended treatment for th s group of pat ents Th s pathway had been n t ated wh ch cons sts of g v ng a f u d cha enge | 1-23-Principles-EBP >>| << 8-02-NHS-code-conduct| Ant b ot cs had been prescr bed as per

|<< Me-1-Reflectivity| Trust gu de nes wh ch were C ar thromyc n 500 m grams tw ce da y ntravenous y and Coamox c av 1 2 grams three t mes da y ntravenous y | 8-02-NHS-code-conduct >>| |<< 1-03-Clinical-Effectiveness| G v ng f u d s part of the Seps s pathway (Surv v ng Seps s Campa gn 2012) | 1-03-Clinical-Effectiveness >>| however |<< 2-10-Requests-interprests-tests| on check ng the b ood resu ts that were taken that morn ng the pat ent had ra sed Urea and Creat n ne wh ch nd cated an Acute K dney n ury (AK) (The Br t sh Rena Assoc at on (Davenport and Stevens 2008) | 2-10-Requests-interprests-tests >>| << 2-08-Most-Appropiate-Tx| F u d adm n strat on s frequent y necessary to stab ze the pat ent w th seps s | 2-08-Most-Appropiate-Tx >>| moreover seps s has been reported to account for approx mate y 50% of |<< 2-09-Working-Diagnosis| pat ents w th AK (Sh geh ko and Ke um 2005) R vers Haystad and Ress er (2001) state the t m ng of the f u d adm n strat on and the phys o og c end po nts to be mon tored have emerged as Ear y Goa - D rected Therapy |<< Be-2-Ad-Beg| The management of AK s therefore to g ve f u ds Schr er | 2-09-Working-Diagnosis >>| (2010) states that guarantee ng adequate rena perfus on and ntravascu ar vo ume s mportant

n the prevent on and therapy of AK | *Be-2-Ad-Beg >>*| |*<< 1-13-Medical-History*| On check ng the pat ent s past med ca h story t was noted that the pat ent suffered from Heart Fa ure (HF) | *1-13-Medical-History >>*| |*<< 1-20-precautions-contra-ind*| When treat ng HF t s necessary to prescr be f u ds caut ous y so as not to over oad the pat ent wh ch cou d resu t n pu monary oedema | *1-20-precautions-contra-ind >>*| (Burton and udw g 2011) My d emma therefore was that /*<< 1-22-presenting-condition*| needed to g ve f u ds to treat the AK but a so needed to off oad f u ds to treat the poss b e pu monary oedema |*<< Be-2-Ad-Beg* /*<< Bl-2-Comph*| AK s character sed by sudden deter orat on n k dney funct on man fested by an ncrease n serum creat n ne eve w th or w thout reduced ur ne output (Rahman Shad and Sm th 2012) | *1-22-presenting-condition >>*| *Be-3-Comp >>*| *Bl-2-Comph >>*| Recogn t on of|*<< 1-20-precautions-contra-ind*| r sk factors e g o der age seps s hypovo em a/shock card ac surgery nfus on of contrast agents d abetes me tus pre-ex st ng chron c k dney d sease heart fa ure and ver fa ure) s mportant because n some cases the treatment of the r sk factor can have a negat ve effect on the AK n th s part cu ar nc dence my pat ent had severa of the r sk factors e o der age seps s and hea t fa ure | *1-20-precautions-contra-ind >>*| |*<< Bl-01-Knowl*|

One th rd of pat ents hosp ta sed for HF deve op AK assoc ated w th ncreased n-hosp ta morb d ty and mo ta ty (Abo-Sa em Sher f and Prabhakar 2013) Sambandam (2014) states that pat ents w th congest ve HF often suffer frequent adm ss ons for the management |*Bl-01-Knowl >>*| of f u d over oad |*<< 1-19-pharmacokinetics-Medicines*| F u d remova and re ef of congest on are centra to treatment of HF w th d uret cs be ng the decongest ve ma nstay | *1-19-Pharmacokinetics-Medicines >>*| (Kr shnamoo thy and Fe ker 2014) Sambandam (2014) further suggests |*<< 4-02-checks-dose-SE*| that a though oop d uret cs are p vota n the management of th s common c n ca prob em hav ng been n c n ca use s nce the 1960s we st do not understand how to opt ma y adm n ster these drugs |*<< 1-25.Drug-mode-action*| *4-02-checks-dose-SE >>*| t s unknown why some decompensated HF pat ents exh b t mprovements n rena funct on w th d ures s whereas others d sp ay rena funct on deter orat on m t ng atta nment of euvo em a | *1-25.Drug-mode-action >>*| eto Aspromonte and Feo a (2014) argue that a though d uret c therapy s w de y used and strong y recommended by most recent c n ca gu de nes /*<< 1-23-Principles-EBP*| prospect ve stud es and random sed c n ca tr a s are ack ng and so re ab e ev dence s m ss ng about the best therapy n terms of doses and methods of adm n strat on /*<< Bl-3-Appli*|

n add t on c n ca eff cacy and safety outcomes are often affected by the presence of contrast ng ev dence | *1-23-Principles-EBP >>*| They further found that ana yses of the pub shed stud es d d not prov de conc us ve data about the effects on c n ca outcomes (morta ty rate of hosp ta readm ss ons ength of hosp ta stay and adverse events) | *Bl-3-Appli >>*| Furthermore |*<< 1-11-Knowl-of-product*| recent stud es focus the r attent on on a ternat ve strateg es of f u d remova such as vasopress n antagon sts adenos ne antagon sts and u traf trat on | *1-11-Knowl-of-product >>*| but ava ab e data are often nconc us ve |*<< Bl-01-Knowl*| *<< Be-1-Novice* /Fo ow ng a terature search and a rev ew of the resu ts found that the use of a oop d uret c such as Frusem de was the common approach to treat ng f u d over oad n HF | *Bl-01-Knowl >>*| *Be-1-Novice >>*| However a ot of the terature d scusses a ternat ve therap es such as |*<< Bl-2-Comph*| extracorporea u traf trat on be ng used to remove f u d from d uret c-refractory hypervo em c pat ents Recent c n ca stud es us ng user-fr end y mach nes have suggested that u traf trat on may be h gh y effect ve for decongest ng pat ents w th HF (| *Bl-2-Comph >>*| Freda S awsky Ma d and Braden 2011) Rego st and F accador (2013) suggest /*<< 1-23-Principles-EBP*| u traf trat on e f u d remova from b ood by hemoconcentrat on obta ned w th ded cated mach nes be ng a more rap d and effect ve a ternat ve to the convent ona therapeut c approach | *1-23-Principles-EBP >>*| |*<< Bl-3-Appli*| | *Bl-3-Appli >>*| Fo ow ng the r research they conc uded that u traf trat on shou d be reserved to very se ected pat ents w th advanced HF and true d uret c res stance as part of a more comp ex strategy a m ng at an adequate contro of f u d retent on | *Bl-3-Appli >>*| |*<< 5-04-Makes-Rx-decisions-pt*| n rea ty w th n the Trust that work am ook ng for a qu ck response to th s deter orat ng pat ent and therefore Frusem de as a oop d uret c s the drug of cho ce | *5-04-Makes-Rx-decisions-pt >>*| |*<< 1-01-Allergies*| f rst y checked f the pat ent had any known a erg es | *1-01-Allergies >>*| and |*<< 5-04-Makes-Rx-decisions-pt*| then prescr bed 40mg Frusem de to be g ven mmed ate y and | *5-04-Makes-Rx-decisions-pt >>*| |*<< 2-05-Est-Maintains-plan-review*|p anned to assess n two hours for a response f the response further aggravated the AK wou d prescr be a bo us of V f u ds and a so mon tor that response | *2-05-Est-Maintains-plan-review >>*| had |*<< 2-10-Requests-interprests-tests*| asked for a further set of b oods to be taken n the morn ng so that urea and creat n ne cou d be assessed thereby mon tor ng the AK | *2-10-Requests-interprests-tests >>*| |*<< 9-05-Appr-support-for-role*| Th s treatment p an was d scussed w th my C n ca Superv sor n

depth who agreed w th my recommendat ons | *9-05-Appr-support-for-role >>*|

|*<< Me-2-Affect-Reflect*| Hav ng assessed th s pat ent /*<< 2-03-Effects-multiple-pathologies*| am consc ous that they have many co-morb d t es resu t ng n the pat ent tak ng an assortment of drugs wh ch have been prescr bed for each of the d sorders |*2-03-Effects-multiple-pathologies >>*|*Me-2-Affect-Reflect >>*| /*<< 2-04-Meds-contrainds-Tx-options*| Mu t p e prescr pt ons or po ypharmacy as t s somet mes referred to can be prob emat ca n that some drugs can nteract w th another drug and cause the pat ent to exper ence add t ona s de effects |*2-04-Meds-contrainds-Tx-options >>*| The Br t sh Nat ona Formu a y (BNF) (2014) nc udes an append x of over one hundred pages wh ch sts how one drug can nteract w th another |*<< 2-04-Meds-contrainds-Tx-options*| nteract ons between drugs can e ther ncrease or decrease the act v ty of one of the drugs (McFadden 2013) However t cou d be /*<< 5-05-safety-conserns-Rx*| the nteract on of the drugs s of tt e or no c n ca re evance because e ther one drug w on y marg na y decrease the act v ty of another drug |*5-05-safety-conserns-Rx >>*| or does not produce any measurab e change n the therapeut c effect veness These drugs wou d perhaps not cause ma or concern but t wou d be mportant to be aware of the nteract ons |*DK know-that >>*|

There are drugs where a sma change n concentrat on can ead to ser ous nteract ons |*2-04-Meds-contrainds-Tx-options >>*| |*<< 1-15-Meds-Altered-Age-etc*| The pat ent that have rev ewed s e der y and age s a factor that must be cons dered when assess ng c n ca y Two o der peop e can vary mass ve y w th regard to the r hea th Th s d vers ty w have evo ved due to the festy e d et soc a c ass and whether the person smoked Regard ess of the o der person s festy e t s poss b e that the effect veness of the pat ent s ver and k dneys have dec ned due to age |*1-15-Meds-Altered-Age-etc >>*| |*<< 1-19-Pharmacokinetics-Medicines*| ver mass and ts b ood f ow decreases w th o d age C earance of some drugs can a so be affected by a dec ne n act v ty of var ous P450 enzymes O d age a so affects rena funct on resu t ng n drugs be ng metabo sed and excreted more s ow y |*1-19-Pharmacokinetics-Medicines >>*| |*<< 6-02- earns-changes-pract ce*| t was a usefu exerc se ref ect ng back upon both the pos t ve and negat ve factors regard ng th s exper ence wh ch has deve oped and mproved my sk s n hea th assessment |*6-02-learns-changes-practice >>*| had been keen to manage the care and prescr be drugs appropr ate y for a pat ent w th th s part cu ar group of symptoms |*<< Me-2-Affect-Reflect*| am regu ar y faced w th th s d emma and pr or to th s had been apprehens ve about prescr b ng frusem de to treat the HF and f u d over oad thus exacerbat ng the AK |*Me-2-Affect-Reflect >>*| /*<< Me-3-Discr-Refelect*| |*BI-3-Appli >>*| S mu taneous y was a so concerned about prescr b ng f u ds to treat the AK wh ch cou d have a negat ve effect on the HF w th further f u d over oad |*Me-3-Discr-Refelect >>*| |*<< Be-3-Compt*| have earnt that t rea y s a pat ent by pat ent approach each be ng nd v dua and a though s m ar symptoms may requ re d fferent treatment t s a case of assess treat mon tor and further assess |*Be-3-Compt >>*| Hav ng the ab ty to prescr be w enhance my pat ent s care because there w be no t me de ay n assessment d agnos s and treatment

As a non-med ca prescr ber am on y ab e to prescr be once a comprehens ve h story assessment and d agnos s has been made (8eckw th & Frank n 2011) |*<< Be-3-Compt*| Th s s a ma or part of my ro e and fee comfortab e and competent w th th s aspect of pat ent care and fee that t w be a great advantage to both myse f and pat ents when am ab e to prescr be ndependent y |*Be-3-Compt >>*| /*<< 4-01-knows-limits-knowledge*| am aware of my m tat ons and w work w th n the NMC Code of Conduct (2006 2015) a s ng w th others and referr ng pat ents on but u t mate y fee that team work ng w be enhanced w th my ab ty to prescr be ndependent y |*4-01-knows-limits-knowledge >>*| |*<< BI-01-Knowl*| As a Non Med ca Prescr ber w be respons b e for prescr b ng for pat ents by mak ng safe effect ve and cost effect ve prescr b ng dec s ons n order to de ver the best treatment and care ava ab e |*BI-01-Knowl >>*| (Beckw ck and Frank n 2011) |*<< BI-2-Comph*| |*<< Me-1-Reflectivity*|

n conc us on th s ref ect on has been a worthwh e exerc se for ana ys ng my anx et es and a so ook ng at my strengths The above has exp ored prescr b ng dec s ons and med cat on cho ces based on pharmaco ogy and ev dence based pract ce n re at on to non-med ca prescr b ng |*BI-2-Comph >>*| |*Me-1-Reflectivity >>*| n the future when f nd myse f w th a pat ent w th s m ar symptoms /*<< 4-08-maint-confidence-competence*| w be more conf dent n recommend ng a treatment p an |*4-08-maint-confidence-competence >>*| and when am ab e to prescr be there w be no de ay n commenc ng treatment Hav ng worked through th s med ca prob em and

undertak ng a terature rev ew now apprec ate that t s acceptab e to |<< *9-04-improves-practice*| treat th s cond t on a certa n way reassess and then treat d fferent y depend ng upon the react on to the treatment| *9-04-improves-practice >>*|

Appendix 7: The 'R' ***itm*** Command Codes for the Rasch Analysis

```
load(file.choose())
    load(file.choose())
    pQdmand<-read.csv(file.choose(),header=T)
    names(pQdmand)
    names(Qdmand)
    names(Qd)
    library(ltm)
    ## Non-Mandatory Questions:
    m1<-rasch(Qd,constraint=cbind(length(Qd)+1,1))
    m2<-rasch(Qd)
    m3<-ltm(Qd~z1)
    m4<-tpm(Qd,type="rasch",max.guessing=1)
    anova(m1,m2)
    anova(m2,m3)
    anova(m2,m4)
    GoF.rasch(m2,B=199)
    coef(m2,prob=T,order=T)
    Command = plot(m2, legend = TRUE, cx = "bottomright", lwd = 3,
    cex.main = 1.5, cex.lab = 1.3, cex = 0.9)
    par(mfrow=c(1,2))
    plot(m2, type = "IIC", annot = FALSE, lwd = 3, cex.main = 1.5,
    cex.lab = 1.3)
```

```
plot(m2, type = "IIC", items = 0, lwd = 3, cex.main = 1.5,
cex.lab = 1.3)
information(m2,c(-4,0))
information(m2,c(0,4))
## Mandatory Questions:
m1<-rasch(Qdmand,constraint=cbind(length(Qdmand)+1,1))
m2<-rasch(Qdmand)
m3<-ltm(Qdmand~z1)
m4<-tpm(Qdmand,type="rasch",max.guessing=1)
anova(m1,m2)
anova(m2,m3)
anova(m2,m4)
GoF.rasch(m2,B=199)
coef(m2,prob=T,order=T)
Command = plot(m2, legend = TRUE, cx = "bottomright", lwd = 3,
cex.main = 1.5, cex.lab = 1.3, cex = 1.4)
plot(m2, legend = TRUE, cx = "bottomright", lwd = 3,
```

- cex.main = 1.5, cex.lab = 1.3, cex = 1.4,items=c(1,2,3,4,5,6,8,9))

```
par(mfrow=c(1,2))
plot(m2, type = "IIC", annot = FALSE, lwd = 3, cex.main = 1.5,
cex.lab = 1.3)
plot(m2, type = "IIC", items = 0, lwd = 3, cex.main = 1.5,
cex.lab = 1.3)
par(mfrow=c(1,2))
```

```
plot(m2, type = "IIC", annot = FALSE, lwd = 3, cex.main = 1.5,
cex.lab = 1.3,items=c(1,2,3,4,5,6,8,9))
plot(m2, type = "IIC", items = 0, lwd = 3, cex.main = 1.5,
cex.lab = 1.3)
information(m2,c(-4,0))
information(m2,c(0,4))
## Pharma Mandatory Questions: NOT POSSIBLE DUE TO SINGULARITIES
## TOTAL:
tot<-cbind(Qd,Qdmand,pQdmand)
m1<-rasch(tot,constraint=cbind(length(tot)+1,1))
m2<-rasch(tot)
m3<-ltm(tot~z1)
m4<-tpm(tot,type="rasch",max.guessing=1)
anova(m1,m2)
anova(m2,m3)
anova(m2,m4)
GoF.rasch(m2,B=199)
coef(m2,prob=T,order=T)
Command = plot(m2, legend = TRUE, cx = "bottomright", lwd = 3,
cex.main = 1.5, cex.lab = 1.3, cex = 0.5)
par(mfrow=c(1,2))
plot(m2, type = "IIC", annot = FALSE, lwd = 3, cex.main = 1.5,
cex.lab = 1.3)
plot(m2, type = "IIC", items = 0, lwd = 3, cex.main = 1.5,
```

```
cex.lab = 1.3)

information(m2,c(-4,0))

information(m2,c(0,4))

## Groups of Questions; can only be applied to questions with variance:

tot<-cbind(Qd,Qdmand,pQdmand)

t1<-as.data.frame(colSums(tot))
```

- ```
  t1$Prop<-t1[,1]/nrow(tot)

  t1$Qus<-row.names(t1)

  t2<-t1[which(t1[,2]!=1),]

  tot2<-tot[,names(tot)%in%t2$Qus]

  tot2.1<-tot2[,1:4]

  tot2.2<-tot2[,5:10]

  tot2.3<-tot2[,11:15]

  tot2.4<-tot2[,16:20]

  tot2.5<-tot2[,21:25]

  tot2.6<-tot2[,26:30]

  ## tot2.1:

  m1<-rasch(tot2.1,constraint=cbind(length(tot2.1)+1,1))

  m2<-rasch(tot2.1)

  m3<-ltm(tot2.1~z1)

  m4<-tpm(tot2.1,type="rasch",max.guessing=1)

  anova(m1,m2)

  anova(m2,m3)

  anova(m2,m4)
  ```
  ```

```
GoF.rasch(m2,B=199)

coef(m2,prob=T,order=T)

par(mfrow=c(2,2))

plot(m2, legend = TRUE, cx = "bottomright", lwd = 3,

cex.main = 1.5, cex.lab = 1.3, cex = 1)

plot(m2, type = "IIC", annot = FALSE, lwd = 3, cex.main = 1.5,

cex.lab = 1.3)

plot(m2, type = "IIC", items = 0, lwd = 3, cex.main = 1.5,

cex.lab = 1.3)

Command = plot(0:1, 0:1, type = "n", ann = FALSE, axes = FALSE)

info1 <- information(m2, c(-4, 0))

info2 <- information(m2, c(0, 4))

text(0.5, 0.5, labels = paste("Total Information:", round(info1$InfoTotal, 3),

"\n\nInformation in (-4, 0):", round(info1$InfoRange, 3),

paste("(", round(100 * info1$PropRange, 2), "%)", sep = ""),

"\n\nInformation in (0, 4):", round(info2$InfoRange, 3),

paste("(", round(100 * info2$PropRange, 2), "%)", sep = "")), cex = 1.5)

## tot2.2:

m1<-rasch(tot2.2,constraint=cbind(length(tot2.2)+1,1))

m2<-rasch(tot2.2)

m3<-ltm(tot2.2~z1)

m4<-tpm(tot2.2,type="rasch",max.guessing=1)

anova(m1,m2)

anova(m1,m3)
```

- anova(m1,m4)

  GoF.rasch(m1,B=199)

  tt1<-as.data.frame(coef(m1,prob=T,order=T))

  write.table(tt1,"clipboard",sep="\t",row.names=T)

  par(mfrow=c(2,2))

  plot(m1, legend = TRUE, cx = "bottomright", lwd = 3,

  cex.main = 1.5, cex.lab = 1.3, cex = 1)

  plot(m1, type = "IIC", annot = FALSE, lwd = 3, cex.main = 1.5,

  cex.lab = 1.3)

  plot(m1, type = "IIC", items = 0, lwd = 3, cex.main = 1.5,

  cex.lab = 1.3)

  plot(0:1, 0:1, type = "n", ann = FALSE, axes = FALSE)

  info1 <- information(m1, c(-4, 0))

  info2 <- information(m1, c(0, 4))

  text(0.5, 0.5, labels = paste("Total Information:", round(info1$InfoTotal, 3),

  "\n\nInformation in (-4, 0):", round(info1$InfoRange, 3),

  paste("(", round(100 * info1$PropRange, 2), "%)", sep = ""),

  "\n\nInformation in (0, 4):", round(info2$InfoRange, 3),

  paste("(", round(100 * info2$PropRange, 2), "%)", sep = "")), cex = 1.5)

  ## tot2.3:

  m1<-rasch(tot2.3,constraint=cbind(length(tot2.3)+1,1))

  m2<-rasch(tot2.3)

  m3<-ltm(tot2.3~z1)

  m4<-tpm(tot2.3,type="rasch",max.guessing=1)

```
anova(m1,m2)

anova(m2,m3)

anova(m2,m4)

GoF.rasch(m2,B=199)

tt1<-as.data.frame(coef(m2,prob=T,order=T))

write.table(tt1,"clipboard",sep="\t",row.names=T)

par(mfrow=c(2,2))

plot(m2, legend = TRUE, cx = "bottomright", lwd = 3,

cex.main = 1.5, cex.lab = 1.3, cex = 1)

plot(m2, type = "IIC", annot = FALSE, lwd = 3, cex.main = 1.5,

cex.lab = 1.3)

plot(m2, type = "IIC", items = 0, lwd = 3, cex.main = 1.5,

cex.lab = 1.3)

plot(0:1, 0:1, type = "n", ann = FALSE, axes = FALSE)

info1 <- information(m2, c(-4, 0))

info2 <- information(m2, c(0, 4))

text(0.5, 0.5, labels = paste("Total Information:", round(info1$InfoTotal, 3),

"\n\nInformation in (-4, 0):", round(info1$InfoRange, 3),

paste("(", round(100 * info1$PropRange, 2), "%)", sep = ""),

"\n\nInformation in (0, 4):", round(info2$InfoRange, 3),

paste("(", round(100 * info2$PropRange, 2), "%)", sep = "")), cex = 1.5)

## tot2.4:

m1<-rasch(tot2.4,constraint=cbind(length(tot2.4)+1,1))

m2<-rasch(tot2.4)
```

```
m3<-ltm(tot2.4~z1)

m4<-tpm(tot2.4,type="rasch",max.guessing=1)

anova(m1,m2)

anova(m2,m3)

anova(m2,m4)

GoF.rasch(m2,B=199)

tt1<-as.data.frame(coef(m2,prob=T,order=T))

write.table(tt1,"clipboard",sep="\t",row.names=T)

par(mfrow=c(2,2))

plot(m2, legend = TRUE, cx = "bottomright", lwd = 3,

cex.main = 1.5, cex.lab = 1.3, cex = 1)

plot(m2, type = "IIC", annot = FALSE, lwd = 3, cex.main = 1.5,

cex.lab = 1.3)

plot(m2, type = "IIC", items = 0, lwd = 3, cex.main = 1.5,

cex.lab = 1.3)

Command = plot(0:1, 0:1, type = "n", ann = FALSE, axes = FALSE)

info1 <- information(m2, c(-4, 0))

info2 <- information(m2, c(0, 4))

text(0.5, 0.5, labels = paste("Total Information:", round(info1$InfoTotal, 3),

"\n\nInformation in (-4, 0):", round(info1$InfoRange, 3),

paste("(", round(100 * info1$PropRange, 2), "%)", sep = ""),

"\n\nInformation in (0, 4):", round(info2$InfoRange, 3),

paste("(", round(100 * info2$PropRange, 2), "%)", sep = "")), cex = 1.5)

## tot2.5:
```

```
m1<-rasch(tot2.5,constraint=cbind(length(tot2.5)+1,1))

m2<-rasch(tot2.5)

m3<-ltm(tot2.5~z1)

m4<-tpm(tot2.5,type="rasch",max.guessing=1)

anova(m1,m2)

anova(m2,m3)

anova(m2,m4)

GoF.rasch(m2,B=199)

tt1<-as.data.frame(coef(m2,prob=T,order=T))
```

- `write.table(tt1,"clipboard",sep="\t",row.names=T)`

```
par(mfrow=c(2,2))

plot(m2,legend = TRUE, cx = "bottomright", lwd = 3,

cex.main = 1.5, cex.lab = 1.3, cex = 1,items=c(1,2,3,4,5))

plot(m2, type = "IIC", annot = FALSE, lwd = 3, cex.main = 1.5,

cex.lab = 1.3)

plot(m2, type = "IIC", items = 0, lwd = 3, cex.main = 1.5,

cex.lab = 1.3)

plot(0:1, 0:1, type = "n", ann = FALSE, axes = FALSE)

info1 <- information(m2, c(-4, 0))

info2 <- information(m2, c(0, 4))

text(0.5, 0.5, labels = paste("Total Information:", round(info1$InfoTotal, 3),

"\n\nInformation in (-4, 0):", round(info1$InfoRange, 3),

paste("(", round(100 * info1$PropRange, 2), "%)", sep = ""),

"\n\nInformation in (0, 4):", round(info2$InfoRange, 3),
```

```
paste("(", round(100 * info2$PropRange, 2), "%)", sep = "")), cex = 1.5)
plot(m2,legend = TRUE, cx = "bottomright", lwd = 3,
cex.main = 1.5, cex.lab = 1.3, cex = 1,items=c(1,4,5))
plot(m2, type = "IIC", annot = FALSE, lwd = 3, cex.main = 1.5,
cex.lab = 1.3,items=c(1,4,5))
plot(m2, type = "IIC", items = 0, lwd = 3, cex.main = 1.5,
cex.lab = 1.3)
Command =  plot(0:1, 0:1, type = "n", ann = FALSE, axes = FALSE)
info1 <- information(m2, c(-4, 0))
info2 <- information(m2, c(0, 4))
text(0.5, 0.5, labels = paste("Total Information:", round(info1$InfoTotal, 3),
"\n\nInformation in (-4, 0):", round(info1$InfoRange, 3),
paste("(", round(100 * info1$PropRange, 2), "%)", sep = ""),
"\n\nInformation in (0, 4):", round(info2$InfoRange, 3),
paste("(", round(100 * info2$PropRange, 2), "%)", sep = "")), cex = 1.5)
## tot2.6:
m1<-rasch(tot2.6,constraint=cbind(length(tot2.6)+1,1))
m2<-rasch(tot2.6)
m3<-ltm(tot2.6~z1)
m4<-tpm(tot2.6,type="rasch",max.guessing=1)
anova(m1,m2)
anova(m1,m3)
anova(m1,m4)
GoF.rasch(m1,B=199)
```

```
tt1<-as.data.frame(coef(m1,prob=T,order=T))
write.table(tt1,"clipboard",sep="\t",row.names=T)
par(mfrow=c(2,2))
plot(m1, legend = TRUE, cx = "bottomright", lwd = 3,
cex.main = 1.5, cex.lab = 1.3, cex = 1)
plot(m1, type = "IIC", annot = FALSE, lwd = 3, cex.main = 1.5,
cex.lab = 1.3)
plot(m1, type = "IIC", items = 0, lwd = 3, cex.main = 1.5,
cex.lab = 1.3)
plot(0:1, 0:1, type = "n", ann = FALSE, axes = FALSE)
info1 <- information(m1, c(-10, 0))
info2 <- information(m1, c(0, 10))
text(0.5, 0.5, labels = paste("Total Information:", round(info1$InfoTotal, 3),
"\n\nInformation in (-10, 0):", round(info1$InfoRange, 3),
paste("(", round(100 * info1$PropRange, 2), "%)", sep = ""),
"\n\nInformation in (0, 10):", round(info2$InfoRange, 3),
paste("(", round(100 * info2$PropRange, 2), "%)", sep = "")), cex = 1.5)
```

Appendix 8: The Confidence Intervals (CI) and Kappa Results

Kappa Results from All

Student 1:

Subjects = 67

Raters = 2

%-agree = 87.4

Cohen's Kappa for 2 Raters (Weights: unweighted)

Kappa = 0.759

z = 11.7

p-value < 0.01

95% CI = 0.63, 0.89

Student 2:

Subjects = 67

Raters = 2

%-agree = 87.4

Cohen's Kappa for 2 Raters (Weights: unweighted)

Kappa = 0.692

z = 10.4

p-value < 0.01

95% CI = 0.53, 0.85

Student 3:

Subjects = 67

Raters = 2

%-agree = 87.4

Cohen's Kappa for 2 Raters (Weights: unweighted)

Kappa = 0.752

z = 10.7

p-value < 0.01

95% CI = 0.62, 0.89

Student 4:

Subjects = 67

Raters = 2

%-agree = 88.1

Cohen's Kappa for 2 Raters (Weights: unweighted)

Kappa = 0.78

z = 9.84

p-value < 0.01

95% CI = 0.64, 0.92

Student 5:

Subjects = 67

Raters = 2

%-agree = 87.4

Cohen's Kappa for 2 Raters (Weights: unweighted)

Kappa = 0.703

z = 9.52

p-value < 0.01

95% CI = 0.54 0.70 0.86

Student 6:

Subjects = 67

Raters = 2

%-agree = 90.8

Cohen's Kappa for 2 Raters (Weights: unweighted)

Kappa = 0.756

z = 11

p-value < 0.01

95% CI = 0.60, 0.92

Student 7:

Subjects = 67

Raters = 2

%-agree = 86.2

Cohen's Kappa for 2 Raters (Weights: unweighted)

Kappa = 0.702

z = 10.5

p-value < 0.01

95% CI = 0.55, 0.86

Student 8:

Subjects = 67

Raters = 2

%-agree = 52.2

Cohen's Kappa for 2 Raters (Weights: unweighted)

Kappa = 0.272

z = 4.01

p-value = 6.12e-05

95% CI = 0.12, 0.42

Student 9:

Subjects = 67

Raters = 2

%-agree = 81.6

Cohen's Kappa for 2 Raters (Weights: unweighted)

Kappa = 0.692

z = 10.6

p-value < 0.01

95% CI = 0.56, 0.82

Student 10:

Subjects = 67

Raters = 2

%-agree = 89.7

Cohen's Kappa for 2 Raters (Weights: unweighted)

Kappa = 0.765

z = 10.6

p-value < 0.01

95% CI = 0.63, 0.90

Student 11:

Subjects = 67

Raters = 2

%-agree = 92

Cohen's Kappa for 2 Raters (Weights: unweighted)

Kappa = 0.817

z = 10.3

p-value < 0.01

95% CI = 0.69, 0.95

Student 12:

Subjects = 67

Raters = 2

%-agree = 88.5

Cohen's Kappa for 2 Raters (Weights: unweighted)

Kappa = 0.831

z = 12.8

p-value < 0.01

95% CI = 0.74, 0.93

weighted kappa 0.96 0.96 0.96

Student 13:

Subjects = 67

Raters = 2

%-agree = 89.6

Cohen's Kappa for 2 Raters (Weights: unweighted)

Kappa = 0.803

z = 9.4

p-value < 0.01

95% CI = 0.67, 0.94

Student 14:

Subjects = 67

Raters = 2

%-agree = 88.5

Cohen's Kappa for 2 Raters (Weights: unweighted)

Kappa = 0.73

z = 9.89

p-value < 0.01

95% CI = 0.58, 0.88

Student 15:

Subjects = 67

Raters = 2

%-agree = 87.4

Cohen's Kappa for 2 Raters (Weights: unweighted)

Kappa = 0.744

z = 9.76

p-value < 0.01

95% CI = 0.61, 0.88

weighted kappa 0.95 0.95 0.95

Student 16:

Subjects = 67

Raters = 2

%-agree = 87.4

Cohen's Kappa for 2 Raters (Weights: unweighted)

Kappa = 0.769

z = 10.8

p-value < 0.01

95% CI = 0.65, 0.89

Student 17:

Subjects = 67

Raters = 2

%-agree = 82.8

Cohen's Kappa for 2 Raters (Weights: unweighted)

Kappa = 0.705

z = 10.5

p-value < 0.01

95%CI = 0.58, 0.84

Student 18:

Subjects = 67

Raters = 2

%-agree = 94

Cohen's Kappa for 2 Raters (Weights: unweighted)

Kappa = 0.872

z = 10.8

p-value < 0.01

95% CI = 0.76, 0.99

Student 19:

Subjects = 67

Raters = 2

%-agree = 83.9

Cohen's Kappa for 2 Raters (Weights: unweighted)

Kappa = 0.676

z = 9.42

p-value < 0.01

95% CI = 0.53, 0.82

Student 20:

Subjects = 67

Raters = 2

%-agree = 85.1

Cohen's Kappa for 2 Raters (Weights: unweighted)

Kappa = 0.658

z = 9.34

p-value < 0.01

95% CI = 0.50, 0.81

Student 21:

Subjects = 67

Raters = 2

%-agree = 86.2

Cohen's Kappa for 2 Raters (Weights: unweighted)

Kappa = 0.71

z = 10.5

p-value < 0.01

95% CI = 0.56, 0.86

Student 22:

Subjects = 67

Raters = 2

%-agree = 83.9

Cohen's Kappa for 2 Raters (Weights: unweighted)

Kappa = 0.668

z = 9.46

p-value < 0.01

95% CI = 0.52, 0.82

Student 23:

Subjects = 67

Raters = 2

%-agree = 88.5

Cohen's Kappa for 2 Raters (Weights: unweighted)

Kappa = 0.81

z = 11.9

p-value < 0.01

95% CI = 0.70, 0.92

Student 24:

Subjects = 67

Raters = 2

%-agree = 88.5

Cohen's Kappa for 2 Raters (Weights: unweighted)

Kappa = 0.714

z = 9.56

p-value < 0.01

95% CI = 0.56, 0.87

Student 25:

Subjects = 67

Raters = 2

%-agree = 86.2

Cohen's Kappa for 2 Raters (Weights: unweighted)

Kappa = 0.694

z = 9.2

p-value < 0.01

95% CI = 0.54, 0.85

Student 26:

Subjects = 67

Raters = 2

%-agree = 88.1

Cohen's Kappa for 2 Raters (Weights: unweighted)

Kappa = 0.766

z = 10.3

p-value < 0.01

95% CI = 0.62, 0.92

Student 27:

Subjects = 67

Raters = 2

%-agree = 89.6

Cohen's Kappa for 2 Raters (Weights: unweighted)

Kappa = 0.837

z = 10.5

p-value < 0.01

95% CI = 0.72, 0.95

Student 28:

Subjects = 67

Raters = 2

%-agree = 88.5

Cohen's Kappa for 2 Raters (Weights: unweighted)

Kappa = 0.79

z = 11.5

p-value < 0.01

95% CI = 0.67, 0.91

weighted kappa 0.97 0.97 0.97

List of References

Adamson. CW, and Bailie. J, (2012). Education Versus Learning: Restorative Practices in Higher Education. *Journal of Transformative Learning, 10, [3],* 139-156.

Adamson. KA, Gubrud. P, Sideras. S, Lasater. K, (2012a). Assessing the Reliability, Validity, and use of the Lasater Clinical Judgment Rubric: Three Approaches. *Journal of Nurse Education, 51, [2], 66–73.*

Acquaviva. K, Haskell. H, and Johnson. J, (2012) Human Cognition and the Dynamics of Failure to Rescue: The Lewis Blackman case. *Journal of Professional Nursing*, *29, [2],* 95-105

Alexander. PA, (2003). The Development of Expertise: The Journey from Acclimation to Proficiency. *Educational Researcher*, *32,[8],* 10–14.

Alfaro-LeFevre. R, (2016*). Critical Thinking, Clinical Reasoning and Clinical Judgement: A Practical Approach*, [6th Edn.], St Louis, Missouri. Elsevier Publication.

Andrade. HL, (2010). Summing up and Moving Forward: Key Challenges and Future Directions for Research and Development in Formative Assessment. In H. L. Andrade and G. J. Cizek (Eds.), *Handbook of Formative Assessment* (p. 344–351). New York: Routledge.

Anfara. VA, and Mertz. NT, [Eds.], (2015). Theoretical Frameworks in Qualitative Research [2nd Edn] London, Sage Publication.

Astin. AW, and Lising. AA, (2012*). Assessment for Excellence: The Philosophy and Practice of Assessment and Evaluation in Higher Education*. Plymouth. UK: Rowman and Littlefield Publishers.

Aiken, L., Smith, H., and Lake, E. (1994). Lower Medicare Mortality Among a Set of Hospitals Known for Good Nursing Care. *Medical Care, 32,* 771-787.

Aiken LH, Sloane DM, Bruyneel L, Van den Heede. K, Sermeus W, and the RN4CAST Consortium, (2013). Nurses' Reports of Working Conditions and Hospital Quality of Care in 12 Countries in Europe. *International Journal of Nursing Studies*, *50, [2],* 143–53.

Aiken. LH, Sloane, DM, Bruyneel. L, Van den Heede. K, Griffiths. P, Busse. R, Diomidous. M, Kinnunen. J, Kózka. M, Lesaffre. E, McHugh. MD, Moreno-Casbas. MT, Rafferty.AM, Schwendimann. R, Scott. PA, Tishelman. C, van Achterberg. T, Sermeus. W, for the RN4CAST consortium (2014). Nurse Staffing and Education and Hospital Mortality in Nine European Countries: A retrospective Observational Study. *The Lancet, 384, [9931], 1824-1830.*

Aiken. LH, Sloane. D, Griffiths. P, Rafferty. AM, Bruyneel. L, McHugh. M, Maier. CB, Moreno-

Casbas. MT, Ball. JE, Ausserhofer. D, and Sermeus. W for the RN4Cast Consortium (2017). Nursing Skill-mix in European Hospitals: Cross-sectional Study of the Association with Mortality, Patient Ratings, and Quality of Care. *BMJ Quality and Safety,* 26, [7], 559-568.

Alexander. PA, (2003). The Development of Expertise: The Journey from Acclimation to Proficiency. *Educational Researcher, 32, [8],* 10-14.

American Heart Association, (2017). Get with the guidelines-resuscitation overview. From http://www.heart.org/HEARTORG/Professional/GetWithTheGuidelines/GetWithTheGuidelines-Resuscitation/Get-With-The- Guidelines-Resuscitation- Overview_UCM_314497_Article .jsp#.WPefZ4jyv6Q. Accessed Apr 20, 2019.

American Association of Colleges of Nursing (AACN, 2011). The Essentials of Master's Education in Nursing. http://www.aacn.nche.edu/education-resources/MastersEssentials11.pdf. Accessed Apr 04, 2013.

Anastasi. A, and Urbina. S, (1997). *Psychological Testing.* Upper Saddle River, New Jersey: Prentice-Hall.

Anderson. JR, Bothell. D, Byrne. MD, Douglas. S, Lebiere. C, and Qin. Y, (2004). An Integrated Theory of Mind. *Psychological Review*, 111:1036-1060.

Anderson. JR, (2007) *How Can the Human Mind Occur in the Physical Universe*? New York, NY: Oxford University Press.

Andrews. S, Gregory. L, Cowin. LS, Eagar. SC, Hengstberger-Sims. C, and Rolley. J, (2008). Psychometric Properties of the Australian Nurse Competency 2000 Standards. *International Journal of Nursing Studies, 45*, 1512-1215.

Argyris. C, and Schon. DA. (1974). *Theory in Practice: Increasing Professional Effectiveness.* San Fransisco: Josey-Bass.

Aromataris. E, Fernandez. R, Godfrey. C, Holly . C, Kahlil. H, and Tungpunkom. P (2015) Summarizing Systematic Reviews: Methodological Development, Conduct and Reporting of an Umbrella Review Approach. *International Journal of Evidence-Based Healthcare, 13, [3],* 132-40.

Asikainen. H, and Gijbels. D, (2017). Do Students Develop Towards More Deep Approaches to Learning During studies? A Systematic Review on the Development of Students' Deep and Surface Approaches to Learning in Higher Education. *Educational Psychology Review, 29,* 205–234.

Audet. L, Bourgault. P, and Rochefort. CM, (2018). Associations between Nurse Education and Experience and the Risk of Mortality and Adverse Events in Acute Care Hospitals; A Systematic Review. *Journal of International Nursing Studies,* 80, 128-146.

Potter. A (2017). Reasoning Between the Lines: A Logic of Relational Propositions. *Dialogue and Discourse, 9, [2],* 80-110.

Avery. A, (2010). Avoidable Prescribing Errors: Communication and Monitoring. *Prescriber, 21, [6]*, 44–46.

Avery. A, (2010a). Avoidable Prescribing Errors: Incidence and Causes. *Prescriber, 2, [5]*, 52-55.

Avery. T, Barber. N, Ghaleb. M, Franklin. BD, Armstrong. S, Crowe. S, Dhillon. S, Freyer. A, Howard. R, Pezzolesi. C, Serumaga. B, Swanick. G, and Talabi. O, (2012). *Investigating the Prevalence and Causes of Prescribing Errors in General Practice: The PRACtICe Study.* A report for the General Medical Council.

Axley. L, (2008). Competency: A Concept Analysis. *Nursing Forum, 43, [4]*, 214-222.

Bagnasco. A, Totti. A, Pagnucci. N, Torre. G, Timmins. F, Aleo. G, and Sasso. L, (2015). How to Maintain Equity and Objectivity in Assessing the Communication Skills in a Large Group of Student Nurses During a Long Examination Session, using the Objective Structured Clinical Examination (OSCE). *Nurse Education Today, 38*, 54-60.

Baehr. J, (2011). *The Inquiring Mind: Intellectual Virtues and Virtue Epistemology.* Oxford. Oxford University Press.

Baid. H, (2011). The Objective Structured Clinical Examination within Intensive Care Nursing Education. *British Association of Critical Care Nurses, 16, [2]*, 99-105.

Baker. FB, and Kim. SH, (2004). *Item Response Theory. Parameter Estimations Techniques. 2nd Edn Revised and Expanded.* New York: March Dekker, Inc.

Bakeman. R, and Queraa. V, (2011). *Sequential Analysis and Observational Methods for the Behavioural Sciences.* Cambridge, Cambridge University Press

Ball. SJ, (2003). The Teacher's Soul and the Terrors of Performativity. *Journal of Education Policy, 18, [2]*, 215–228.

Ball. SJ, Maguire.M, Braun. A, and Hoskins. K, (2011). Policy Actors: Doing Policy Work in Schools. Discourse: *Studies in the Cultural Politics of Education, 32, [4],* 625-639.

Ball. SJ, (2012). *Golbal Education Inc. New Policy and the Neoliberal Imaginary.* New York, Routledge Publication.

Banning. M, (2006). Measures that Can be Used to Instill Critical Thinking Skills in nurse Prescribers. N*urse Education and Practice, 6 [2],* 98-105.

Barnett. R, (1997). *Higher education: A Critical Business*. Milton Keynes: Open University Press.

Barnett. R, (1999). *The Limits of Competence: Knowledge, Society and Higher Education.* Buckingham: Open University Press.

Bennett. D, (2012). A creative approach to exploring student identity. *IJCPS-International Journal of Creativity and Problem Solving, 22*, *[1],* 27.

Barnett. R, and Coate. K, (2005). *Engaging the Curriculum in Higher Education.* Buckingham: RHE and Open University Press.

Barnhill. D, McKillop. A, and Aspinall. C, (2012). The Impact of Postgraduate Education on Registered Nurses Working in Acute Care. *Nursing Praxis in New Zealand, 28, [2]*, 27–36.

Baron. WF, and Boulpaep. (2005). *Medical Physiology.* Philadelphia: Elsevier Saunders.

Barton. TD, Bevanet. L, and Mooney. G, (2012) Advanced Nursing 3: What Does the Future Hold for Advanced Nursing? *Nursing Times; 108, [26]*, 19-21.

Barton. A, and Davis. H, (2015). Neoliberalism, HE and Anti-politics: The Assault on the Criminological Imagination. In J. Frauley (Ed.). *C.W. Mills and the Criminological Imagination,* p. 201–218. London: Routledge Publication.

Barry. M, Bradshaw. C, and Noonan. M, (2013). Improving the Content and Face Validity of OSCE Assessment. *Nurse Education in Practice, 13*, 477e480.

Baxter. GP, and Glaser. R (1998). Investigating the Cognitive Complexity of Science Assessments. *Educational Measurement: Issues and Practice, 17, [3]*, 37-45.

Beck. AT, (2002). *Cognitive Behaviour Therapy: A Guide for the Practising Clinician: Volume 1.* Hove, East Sussex: Brunner-Routledge.

Beckstead. JW, (2013). On Measurements and their Quality: Paper 1. Reliability – History, Issues and Procedures. *International Journal of Nursing Studies, 50*, 968–973.

Beckwith. S, and Franklin. P, (2011). *Oxford Handbook of Prescribing for Nurses and Allied Health Professionals.* Oxford: Oxford University Press.

Beckwith. PT, (2018). Signature Pedagogies and the HOTSHOT Educator: A Systematic Literature Review. *E-Leader International Journal, 13, [1],* 1–20.

Bell. A, Kelton. J Donagh. N, Mladenovic. R, and Kelli. M, (2011). A Critical Evaluation of the usefulness of a Coding Scheme to Categorise Levels of Reflective Thinking. *Assessment and Evaluation in Higher Education, 36, [7]*, 797-815.

Benade. L, (2016). T*eaching Critically Reflective Practice in Freire.* In Encylopedia of Educational Philosophy and Theory, p1-6. Springer Publication.

Benner. P, (2001). *From Novice to Expert. Excellence in Power and Clinical Nursing Practice. Commemorative Edition.* New Jersey: Prentice-Hall Health.

Benner. P, (2009). *Expertise in Clinical Practice Caring, Clinical Judgement and Ethics* (2nd Edn) New York. Springer Publishing Company.

Bennett. W, Lance. CE, and Woehr. DJ, (2006). *Performance Measurement: Current Perspectives and Future Challenges.* Hove, East Sussex. Taylor and Francis Group.

Berger. R, (2015). Now I see it. Now I don't: Researcher's Position and Reflexivity in Qualitative *R*esearch. *Qualitative Research,* 15, 219–234.

Bergstein. H, Avery. GC, and Neumann. R, (2010). Kolb's Experiential Learning Model: Critique from a Modelling Perspective. *Studies in Continuing Education, 32 [1]*, 29-46.

Bernstein. B, (2000). *Pedagogy, Symbolic Control and Identity, (2nd Edn)*. Oxford: Rowman and Littlefield Publishers.

Bertucio. B, (2017). The Cartesian Heritage of Bloom Taxonomy. *Studies in Philosophy and Education, 36, 477-497.*

Bickman. L, and Rog. JR, (2009). *The SAGE Handbook of Applied Social Research Methods*. Thousand Oaks, Sage Publication.

Biggs. J, (2003). *Teaching for Quality Learning at University Buckingham*: Maidenhead. The Society for Research into Higher Education (SRHE) and Open University Press.

Biggs. J, (2014). Constructive Alignment in University Teaching. *Higher Education Research and Development Society of Australia (HERDSA) Review of Higher Education, 1*, *[5]*, 5–22.

Biggs. J, and Tang. C, (2011). Teaching for Quality Learning at University: What the Student Does. Maidenhead, Berkshire: Open University Press.

Billett. S, (2001). *Learning in the Workplace: Strategies for Effective Practice*. Sydney, Allen and Unwin.

Bhola. D, Impara. J, and Buckendahl. C, (2003). Aligning Tests with States' Content Standards: Methods and issues. *Educational Measurement: Issues and Practice*, 22 *[3],* 21– 29.

Black. P, and Plowright, D. (2010). A Multi-dimensional Model of Reflective Learning for Professional Development. *Reflective Practice, 11, [2],* 245–258.

Blackwell. (2002). *Blackwells Dictionary of Nursing.* Oxford: Blackwell Publishing.

Blegen. MA, Goode. CJ, Park. SH, Vaughn. T, Spetz. J, (2013). Baccalaureate Education in Nursing and Patient Outcomes. J*ournal of Nursing Administration* 43: 89–94

Bloomfield. J, Pegram. A, and Jones. C, (2010). *How to Pass Your OSCE: A Guide to Success in Nursing and Midwifery.* Harlow. Pearson Education.

Bloom. BS, (1956). *Taxonomy of Educational Objectives: Book 1 Cognitive Domain.* Ann Arbor: David McKay Company, Inc.

Bloxham. S, (2012). You Can See the Quality in Front of Yor Eyes: Grounding Academic Standards Between Rationality and Interpretation. *Quality in Higher Education, 18, [2],* 185-204.

Bond. TG, and Fox. CM. (2015). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences.* London: Routledge-Taylor and Francis Group.

Bondy. KN, (1983). Criterion-referenced Definitions for Rating scales in Clinical Evaluation. *Journal of Nurse Education 22*: 376–82.

Bosco. AM and Ferns. S, (2014). Embedding of Authentic Assessment in work-integrated Learning Curriculum. *Asia- Pacific Journal of Cooperative Education, 15,* 281–290.

Booth. A, (2016). Searching for Qualitative Research for Inclusion in Systematic Reviews: A Structured Methodological Review. *Systematic Reviews, 5, [74]*, 1-23.

Boron. WF and Boulpaep. EL, (2005). *Medical Physiology.* Philadelphia: Elsevier Saunders.

Borsboom. D, (2007). *Measuring the Mind. Conceptual Issues in Contemporary Psychometrics.* Cambridge: Cambridge University Press.

Borsboom. D, and Mellenbergh. GJ, (2004). The Concept of Validity. *Psychological Review, 111, [4]*, 1061–1071.

Borsboom. D, and Mellenbergh. GJ, (2013). Truth and Evidence in Validity Theory. *Journal of Educational Measurement, 50, [1]*, 110-114.

Borsboom. D, Mellenbergh. GJ and Van Heerden. J, (2002). Different Kinds of DIF: A Distinction Between Absolute and Relative forms of Measurement Invariance and Bias. *Applied Psychological Measurement, 26*, 433-450.

Boud. D, (2006). ' Creating the Space for Reflection', in D. Boud, P. Cressey, and P. Docherty (eds) Productive Reflection at Work, 158– 69, London and New York: Routledge.

Boud, D. (2010). Sustainable assessment: Rethinking Assessment for the Learning Society. *Studies in Continuing Education, 22,* 151–167.

Boud. D, Keogh. R, and Walker. D, (2002). *Reflection: Turning Experience into Learning.* London: Kogan Page.

Boud. D, and Soler. R, (2016). Sustainable Assessment Revisited. *Assessment and Evaluation in Higher Education, 41*(3), 400–413.

Bovaird. JA, and Embretson. SE, (2008). *Modern Measurement in the Social Sciences: The Sage Handbook of Social Research Methods.* London: Sage Publication.

Boyatzis. RE. (1998). *Thematic Analysis and Code Development: Transforming Qualitative Information.* London. Sage Publication.

Bradbury. H, Frost. N, Kilminster. S, and Zukas. M, (2010). Beyond Reflective Practice: New Approaches to Professional Life-long Learning. London, Routledge. Taylor and Francis Group.

Brammer. WM, Giustini. D, de Jonge. GB, Holland. L, Bekhuis. T, (2016). De-duplication of

Database Search Results for Systematic Reviews in EndNote. *Journal of the Medical Library Association, 104, [3],* 240–243.

Brannick. MT, Erol-Korkmaz. HT, and Prewett. M, (2011). A Systematic Review of the Reliability of Objective Structured Clinical Examination Scores. *Medical Education, 45 [12]*, 1181-1189.

Braune. R, and Foshay. WR, (1983). Toward a practical Model of Cognitive Information processing Task Analysis and Schema Acquisition for Complex, Problem-solving Situations. *Instructional Science, 12*, 121-145.

Brennan. RL, (2013) Commentary on 'Validating the Interpretations and Uses of Test Scores'. Journal of Educational Measurement, 50, 74–83.

Brimi. HM, (2011). Reliability of Grading High School Work in English. *Practical Assessment, Research and Evaluation, 16, (17)*.

Brier J, Carolyn M, Haverly. M, Janurio. ME, Padula. C, Tal. A, and Triosh. H, (2014). Knowing "Something is Not Right" is Beyond Intuition: Development of a Clinical Algorithm to Enhance Surveillance and Assist Nurses to Organise and Communicate Clinical Findings. *Journal of Clinical Nursing*, 24, [5–6], 832–843.

British Pharmaceutical Commission British Pharmacopoeia (1998). *Published Report on the Recommendation of the Medicines Commission Pursuant to the Medicines Act 1968.* London: The Stationery Office.

Broadbent. J. Dietrich. M, and Roberts. J, (1997). *The End of the Professions: The Restructuring of Professional Work.* London: Routledge.

Brookfield. SD, (2012). Teaching for Critical Thinking: Tools and Techniques to Help Students Question their Assumptions. San Francisco. Josey-Bass.

Brooks. MA, (2009) Medical Education and the Tyranny of Competence. *Perspectives in Biology and Medicine, 52, [1], 90-102.*

Brown. A, (2011). Lessons from Policy Failure: The Demise of a National Qualifications Framework Based Solely on Learning Outcomes in England. *Journal of Contemporary Educational Studies, 62 (5), 36–55.*

Burg. K, Blatt. B, Lopreiato. J, Julianna. J, Schaeffer. A, and Heil. DT, (2015). Standardized Patient Assessment of Medical Student Empathy: Ethnicity and Gender Effects in a Multi-institutional Study. *Academic Medicine, 90, [1],* 105-11.

Burke. JW [Ed], (1995). *Competency-Based Education and Training.* London: Falmer Press.

Burns. J, Magee. K, Cooley. H, Hensler. A, Montana. J, Shumaker. D, Snyder. J, and Polk. A (2010). "I feel your Pain": A Research Study Addressing Paranaesthesia Healthcare Providers' Knowledge and Attitudes Toward Pain. *Journal of Paranaesthesia Nursing,*

25, [1], 212-223.

Bustamante. C , and Moeller. J, (3013). The Convergence of Content, Pedagogy, and Technology in Online Professional Development for Teachers of German: An Intrinsic Case Study. *CALICO Journal, 30, [1],* 82-104.

Calligaris. L, Panzera. A, Arnoldo. L, Londero. C, Quattrin. R, Troncon. MG and Brusaferro. S, (2009). Errors and Omissions in Hospital Prescriptions: A Survey of Prescription Writing in a Hospital. *BioMedCentral-Pharmacology and Toxicology, 9, [1]*, 9-15.

Campbell. RD, Hecker. KG, Biau. DJ, and Pang. DSJ, (2014). Student Attainment of Proficiency in Clinical Skills: The Assessment of Individual Learning Curves. *PLoS ONE, 92, [2]*, e88526.

Carper. BA, (1978). Fundamental Patterns of Knowing in Nursing. *Advances in Nursing Sciences, 1,[1]*, 12-23.

Carraccio. C, Wolfsthal. SD, Englander. R, Ferentz. K and Martin. C, (2002). Shifting Paradigms: From Flexner to Competencies. *Academic Medicine, 77*, 361-367.

Carter. N, Bryant-Lukosius. D, Blythe. J, and Neville. AJ, (2014). The use of Triangulation in Qualitative Research.Methods and Meaning. *Oncology Nursing Forum, 14, [5]*, 454-457

Casey. D, Clarke. L, and Hayes. S (2011). *Study Skills for Master's Level Students: A Health and Social Care Workbook*. Exeter, Reflect Press.

Castelli. PA, (2011). An Integrated Model for Practicing Reflective Learning. *Academy of Educational Leadership Journal,* 15, Special Issue, pp. 15-30.

Castlebury. AN, Payakachat. N, Ashby. S, Nolen. A, Carle. M, Neill. KK, and Franks. AM, (2016). Qualitative Analysis of Written Reflections during a Teaching Certificate Program. *American Journal of Pharmaceutical Education, 80, [1]*, 1-7.

Cathro. V, O'Kane. P, and Gilbertson. D, (2017). Assessing Reflection Understanding Skill Development through Learning Journals. *Education and Training, 59, [4]*, 427-442.

Centre for Reviews and Dissemination (CRD, 2009). *Systematic reviews: CRD Guidance for Undertaking Reviews in Healthcare.* Centre for Reviews and Dissemination. The University of York. York Publishing Services.

Chadha. A, (2017). Comparing Student Reflectiveness in Online Discussion Forums across Modes of Instruction and Levels of Courses. *Journal of Educators Online, 14 [2],*1-19.

Chalhoub-Deville. MB, (2016). Validity theory: Reform Policies, Accountability Testing, and Consequences. *Language Testing*, *33, [4]*, 453-472.

Chalhoub-Deville. MB, (2018). Multilingual Testing Constructs: Theoretical Foundations. *Language Assessment Quarterly*, *16*(4-5), pp.472-480.

Chang. M J, Chang. YJ, Kuo. SH, Yang. YH, and Chou. FH, (2011). Relationships between Critical Thinking Ability and Nursing Competence in Clinical Nurses. *Journal of Clinical Nursing,* 20, 3224–3232.

Charlton. R, (Ed). (2007). *Learning to Consult.* Abingdon: Radcliffe Publishing Ltd.

Cheng. M, (2010). Audit Cultures and Quality Assurance Mechanisms in England: A Study of their Perceived Impact on the Work of Academics. *Teaching in Higher Education, 15,[3]*, 259–271.

Chinn. PL, and Kramer. MK, (2004). *Integrated Knowledge Development in Nursing [6th Edn.].* St. Louis. MO: John Wiley and Sons.

Cho. E, Sloane. DM, Kim. EY, Kim. S, Choi. M, Yoo. IY, Lee. HS, Aiken. LH, (2015). Effects of Nurse Staffing, Work Environments, and Education on Patient Mortality: An Observational Study. *International. Journal of Nursing Studies, 52, [2],* 535–542.

Cioffi. J, (1998). Education for Clinical Decision-making in Midwifery Practice. *Midwifery, 14, [1], 18-22.*

Cizek. GJ, and Bunch. MB, (2007). *Standard Setting: A Guide to Establishing and Evaluating Performance Standards on Test*. Thousand Oaks: Sage Publication.

Cizek. G J, (2008). Assessing Educational Measurement: Ovations, omissions, opportunities. *Educational Researcher,* 37, 96–100.

Cizek. JG, (2012). Defining and Distinguishing Validity: Interpretations of Score Meaning and Justifications of Test Use. *Psychological Methods, 17, [1]*, 31–43.

Clarke. SP and Aiken. LH, (2003). Failure to Rescue. *American Journal of Nursing, 103, [1],* 42-47

Clarke. MC, and Wilson. AL, (1991). Context and rationality in Mezirow's theory of Transformational Rationality in Mezirow's theory of Transformational Learning. *Adult Education Quarterly,* 41, 75-91.

Clarkeburn. H and Kettula. K, (2011). Fairness and using Reflective Journals in Assessment and *Teaching in Higher Education, 17, [4]*, 439-452.

Cohen. R, Rothman. AI, Poldre. P, and Ross. J, (1991). "Validity and Generalizability of Global Ratings in an Objective Structured Clinical Examination," *Academic Medicine, 66, [9],* 545–548.

Cohen-Scali. V, (2012). *Competence and Competence Development*. Opladen, Berlin and Toronto Barbara Budrich Publishers.

Coles. T, (2014). Critical Pedagogy: Schools Must Equip Students to Challenge the Status Quo.

Retrieved Dec 13, 2016, from https://www.theguardian.com/teacher-network/teacher-blog/2014/feb/25/critical-pedagogy- schools-students-challenge.

Collins. S, and Hewer. I, (2014). The Impact of the Bologna Process on Nursing in Higher Education in Europe: A Review. *International Journal of Nursing Studies, 51, [1],* 150-156.

Cook. DA, and Beckman. TJ, (2006). Current Concepts in Validity and Reliability for Psychometric Measurements. *American Medical Journal, 119, [166]*, e167–116.

Cook. DA, Kuper. A, Hatala. R, and Ginsburg. S, (2016). When Assessment Data Are Words: Validity Evidence for Qualitative Educational Assessments. *Academic Medicine*, *91, [*10], 1359-1369.

Cooper. S, Porter. J, and Peach. L, (2016). Measuring Situation Awareness in Emergency Settings: A Systematic Review of Tools and Outcomes. *Emergency Medicine, 6,* 1-7.

Corbin. J, and Strauss. AL. (1990) Cited in McKay. FH, and Dunn. M, (2015). Student Reflections in a First-Year Public Health and Health Promotion Unit. *Reflective Practice, 16, [2]*, 242-253.

Corlett. J, (2000). The Perceptions of Nurse Teacher, Student Nurses and Preceptors of the Theory-practice Gap in Nurse Education. *Nurse Education Today, 20, [6]*, 499-505.

Crooks. T (2011) Assessment for learning in the accountability era: New Zealand. *Studies in Educational Evaluation, 37*, 71–77.

Crosier. D, and Parveve. T, (2013). *Fundamentals of Education and Planning. The Bologna Process: Its Impact in Europe and Beyond*. Paris, the United Nationals Educational, Scientific and Cultural Organization.

Crozier. MS, Ting. HY, Boone. DC, O'Regan. NB, Bandrauk. N, Furey. A, Squires. C, Hapgood. J and Hogan. MP, (2015). Use of Human Patient Simulation and Validation of the Team Situation Awareness Global Assessment Technique (SAGAT): A multidisciplinary Team Assessment Tool in Trauma Education. *Journal of Surgical Education*, *72*, *[1],* 156-163.

Courtenay. M, Stenner. K and Carey. N (2010). The Views of Patients with Diabetes about Nurse Prescribing. *Diabetic Medicine,* 27: 1049-1054.

Cowan. DT, Wilson-Barnett. J and Norman. IJ, (2007). A European Survey of General Nurses' Self Assessment of Competence. *Nurse Education Today, 27*, 452-458.

CoRe 2, (2011). *Tuning Educational Structures in Europe. A Guide to Formulating Degree Programme Profiles – Including Programme Competencies and Programme Learning Outcomes*", published by the Competences in Education and Recognition Project (CoRe).

Core Capabilities Framework for Advanced Clinical Practice (Nurses) Working in General Practice / Primary Care in England (2020). Health Education England, NHS England and Skills for Health.

Council of Deans for Health (2018). *Advanced clinical practice education in England: Event Report from the 2018 Council of Deans of Health/ Health Education England Advanced Clinical Practice Education Conference,* London: Council of Deans for Health.

Craig. JV and Smyth. RL, (Eds), (2002). *The Evidence-Based Practice Manual for Nurses.* London: Churchill, Livingstone.

Critical Appraisal Skills Programme (CASP, 2018). Retrieved Nov 08, 2018, from http://joannabriggs.org/research/critical-appraisal-tools.html: https://casp-uk.net/.

Crossley. J, (2010) Vive la difference! A Recall from Knowing to Exploring. *Medical Education, 44, [10],* 946–948.

Crotty. M, (2009). *The Foundations of Social Research. Meaning and Perspective in the Research Process.* London: Sage Publication.

Crozier. G, Reay. D, and Clayton. J, (2010). The Socio-cultural and Learning Experiences of Working-class Students in Higher Education. *in* David, M. (Ed.) *Improving Learning by Widening Participation.* London: Routledge Publication.

Cunnington. JPW, Neville. AJ, and Norman. GR, (1996). The Risks of Thoroughness: Reliability and Validity of Global Ratings and Checklists in an OSCE," *Advances in Health Sciences Education, 1, [3],* 227–233.

Dalton. M, Harrison. J, Malin. A, and Leavey. C, (2018). Factors that Influence Nurses Assessment of Patient Acuity and Response to Acute Deterioration. *British Journal of Nursing, 27, [4]*, 212 - 218.

Daniels. H, (2001). *Vygotsky and Pedagogy.* London: Routledge-Falmer: Taylor and Francis Group.

Daniels. VJ, Bordage. G, Gierl. MJ, and Yudkowsky. R, (2014). Effect of Clinically Discriminating, Evidence-based Checklist Items on the Reliability of Scores from an Internal Medicine Residency OSCE. *Advances in Health Sciences Education Theory Practice,19, [4],* 497–506.

Danilli. E, and Reid. N, (2006). Cognitive Factors that Can Potentially Affect Pupils' Test Performance. *Chemistry Education Research and Practice, 7, [2]*, 64-83.

Dante. A, Valoppi. G, Saiani. L, and Palese. A, (2011). Factors Associated with Nursing Students' Academic Success or Failure: A Retrospective Italian Multicentre Study. *Nurse Education Today, 31, [1],* 59–64.

Darzi . A, (2008). *High-Quality Care For All: Next Stage Review Final Report. The House of Commons.* London: The Stationery Office Limited.

Dearden. J, Macaro. E, and Akincioglu. M, (2016). "English Medium Instruction in Universities: A Collaborative Experiment in Turkey." *Studies in English Language Teaching 4, [1],* 51-76.

deJong. T, and Ferguson-Hessler.GM, (1996). Types and Qualities of Knowledge. *Educational Psychologist, 31, [2]*, 105-113.

Delavar. MA, Salmalian. H, Faramarzi. M, and Pasha. H, (2013). Using the Objective Structured Clinical Examinations in Undergraduate Midwifery Students. *Journal of Medicine and Life, 6 [1]*, 76-79.

Dellinger. RP, Levy. MM, Rhodes. A, Djillali. A, Gerlach. H, Opal. SM, Sevransky. JE, Sprung. CL, Douglas. IS, Jaeschke. R, Osborn. TM, Nunnally. ME, Townsend. SR, Reinhart.K, Kleinpell. RM, Angus. DC, Deutschman. CS, Machado. FR, Rubenfeld. GD, Webb. SA, Beale. RJ, Vincent. JL, and Moreno. R, (2013). Surviving Sepsis Campaign: International Guidelines for Management of Severe Sepsis and Septic Shock: 2012. *Journal of Critical Care Medicine, 41 [2]*, 580-637.

DeMars. C, (2010). *Item Response Theory: Understanding Statistical Measurement.* Oxford: Oxford University Press.

DeNisco. S, and Barker. AM, (2015). *Advanced Nursing Practice (3rd Edn). Essential Knowledge for the Profession.* Massachusets, Jones and Bartlett Learning.

Denscombe. M, (2014). *The Good Research Guide for Small Scale Social Research [4th Edn].* London: McGraw-Hill Education.

Denton. D, (2011). Reflection and learning: Characteristics, obstacles, and implications. *Educational Philosophy and Theory*, *43, [*8*]*, 838-852.

Deville. C, and Chalhoub-Deville. M, (2011) Accountability-assessment under No Child Left Behind: Agenda, practice, and future." (2011): 307-32.

Dinsmore. DL, and Alexander. PA, (2012). A Critical Discussion of Deep and Surface Processing: What it Means, How it is Measured, the Role of Context, and Model specification. *Educational Psychology Review, 24, 499-567.*

Dencev. H, and Collister. R, (2010). Authentic Ways of Knowing, Authentic Ways of Being: Nurturing a Professional Community of Learning and Praxis. *Journal of Transformative Education, 8, [3],* 178-196.

Department for Innovation, Universities and Skills (2007). *World-Class Skills, Implementing the Leitch Review.* London: Department of Health.

Department of Health (DH, 1986). *Neighbourhood Nursing- A Focus for Care (Cumberlege Report).* London: Her Majestys Stationery Office (HMSO).

Department of Health (DH, 1998). *Review of Prescribing, Supply and Administration of Medicines. A Report on the Supply and Administration of Medicines. Report on the supply and administration of medicines under group protocols (Chair: Dr June Crown).* London: Department of Health.

Department of Health (DH, 2005). *Agenda for Change. NHS Terms and Conditions of Service Handbook.* London: Department of Health.

Department of Health (DH, 2005a). *Research Governance Framework for Health and Social Care.* London: Department of Health.

Department of Health (DH, 2006). *Improving patients' Access to Medicines A Guide to Implementing Nurse and Pharmacist Independent Prescribing within the NHS in England.* London: Department of Health.

Department of Health (DH, 2006a). *Modernising Nursing Careers Report, Setting the Direction.* London: Department of Health.

Department of Health (DH, 2008). *A High-Quality Workforce- Next Stage Review.* London: Department of Health.

Department of Health (DH, 2008a). *An Evaluation of Supplementary Prescribing in Nursing and Pharmacy.* London: Department of Health.

Department of Health (DH, 2010). *Advanced Level Nursing: A Position Statement.* London: Department of Health.

Department of Health (DH, 2010a). *The Caldicott Guardian Manual.* London: Department of Health.

Department of Health (DH, 2010b) *Front Line Care; Report by the Prime Ministers Commission on the Future of Nursing and Midwifery in England*. London: Department of Health.

Dessolle. L, Freour. T. Barriere. P, Jean. M, Ravel. C, Darai. E, and Biau. DJ, (2010). How Soon Can I be Proficient in Embryo Transfer? Lessons from the Cumulative Summation Test for Learning Curve (LC-CUSUM). *Human Reproduction, 25, [2],* 380–386.

DeVilles. RF, (2006). Classical Test Theory in a Multi-ethnic Society. *Medical Care, 44, [11]*, S50-559.

DeVon. HA, Block. ME, Moyle-Wright. P. Dernst. DM, Hayden. SJ, Lazzara. DJ, Savoy. SM and Kostas-Polston. E, (2007). A Psychometric Toolbox for Testing Validity and Reliability. *Journal of Nursing Scholarship, 39, [2]*, 155-164.

Dewey. J, (2004). *How We Think.* New Dehli: Cosmo Publications.

de Wit. H, and Leask. B, (2017). *Preparing Global Citizenry, Implications for Curriculum. In: GUNI, Higher Education in the World 6; Towards a Socially Responsible University: Balancing the Global with the Local*. Barcelona: GUNI-ACUP.

Dickoff. J, and James. P, (1968). A Theory in a Practice Discipline. *Nursing Research, 17, [5]*, 415-435.

Dillon. PM, (2007). *Nursing Health Assessment: A Critical Thinking Case Studies Approach.* Philadelphia: FA Davis Company.

Dinsmore. DL, and Alexander. PA, (2016). A Multidimensional Investigation of Deep-level and Surface-level Processing. *Journal of Experimental Education, 84, [2]*, 213–244.

Dinsmore. DL, (2017). Towards a Dynamic, Multidimensional Model of Strategic Processing. *Educational Psychology Review, 29, [2],* 235–268.

Dix. M, (2016). The Cognitive Spectrum of Transformative Learning. *Journal of Transformative Education, 14, [2],* 139-162.

Dornan. T. Ashcroft. D, Heathfield. H, Lewis. P. Taylor. D, Tully. M and Wass, V, (2009). *An In-depth Investigation into the Causes of Prescribing Errors by Foundation Trainees about their Medical Education. Errors Questioning Undergraduate Impact on Prescribing (EQUIP) Study.* Manchester: Manchester University.

Doucette. A, and Wolf. AW, (2009). Questioning the Measurement Precision of Psychological Research. *Psychotherapy Research, 19, [3-5]*, 374-389.

Downing. SM, (2003). Validity: On the Meaningful Interpretation of Assessment Data. *Medical Education*, 37: 830-837.

Downing. SM, and Haladyna. TM, (Eds). (2006). *Handbook of Test Development.* Mahwah, NJ: Lawrence Erlbaum Associates. Routledge -Taylor and Francis Group.

Downing. SM, and Yudkowsky. R, (2009). *Introduction to Assessment in the Health Professions. In S. M. Downing and R. Yudkowsky (Eds.), Assessment in Health Professions Education: pp. 1-20.*

Drennan. J, (2010). Critical Thinking as an Outcome of a Masters Degree Programme in Nursing *Journal of Advanced Nursing, 66, [2],* 422-432.

Dreyfus. HL, and Dreyfus. SE. (1986). *Mind Over Machine: The Power of Human Intuition and Expertise in the Era of Computer.* New York: Free Press: Davidson of McMillan Inc.

Drummond. JS, and Standish. P, (2007). The Philosophy of Nurse Education. Basingstoke. Palgrave MacMillan.

Duff. B, (2012). A Theoretically Informed Education Program Designed Specifically for Acute Surgical Nurses. *Nurse Education Today 32*, e73–e78.

DuHamel. M, Himle. C, Karvonen. C, Sayre. C, Wyant. S, Smith. N, Keener. S, Barrett. S, and Whitney. J, (2011). Enhancing Medical-Surgical Nursing Practice: Using Practice Tests and Clinical Examples to Promote Active Learning and Programme Evaluation. *Journal of Continuing Education, 42 [10]*, 457-562.

Durand. MA, Carpenter. L, Dolan. H, Bravo. P, Mann. M, Bunn. F, and Elwyn. G, (2014). Do interventions Designed to Support Shared Decision Making Reduce Health Inequalities? A Systematic Review and Meta-Analysis. PLOS-One, 9, [4], e94670.

Dwyer. CP, Hoan. MJ, and Steward. I, (2015). The Effects of Argument Mapping-infused Critical-thinking Instruction on Reflective Judgement Performance. *Thinking Skills and Creativity, 16*, 11-26.

Dwyer. CA, Gallagher. A, Levin. J, and Morley. ME, (2003). *What is Quantitative Reasoning? Defining the Construct for Assessment Purposes.* Educational Testing Service (ETS) Princeton, NJ. Research Publication Office.

Dwyer. CA, and Powers. DE, (2003a). *Toward Specifying a Construct for Reasoning.* Educational Testing Service (ETS) Princeton, NJ. Research Publication Office.

Dwyer. CP, Hogan. MJ, and Stewart. I, (2014). An Integrated Critical Thinking Framework for the 21st Century. *Critical Thinking Skills and Creativity,* 12, 43-52.

Dwyer. CP, Hoan. MJ, and Steward. I, (2015). The Effects of Argument Mapping-infused Critical Thinking Instruction on Reflective Judgement Performance. *Thinking Skills and Creativity, 16*, 11-26.

Dyment. JE, and O'Connell. TS, (2011). Assessing the Quality of Reflection in Student Journals: A Review of the Research. *Teaching in Higher Education, 16, [1]*, 81-97.

Dyson. S, (2018). *Critical Pedagogy in Nursing: Transformational Approaches to Nurse Education in A Globalised World*. London. Palgrave MacMillan.

Eemeren. FH, van Garssen. B, and Meuffels. B, (2009). Fallacies and Judgments of Reasonableness. Dordrecht: Springer Publication.

Eden. S, (2014). "Out of the Comfort Zone: Enhancing Work-Based Learning about Employability through Student Reflection on Work Placements." J*ournal of Geography in Higher Education, 38, [2],* 266–276.

Embretson. SE, and Reise. SP, (2009). *Item Response Theory for Psychologists.* London: Psychological Press-Taylor and Francis Group.

Englander. R, Cameron. T, Ballard. AJ, Dodge. J, Bull. J, and Aschenbrener. CA, (2013). Toward a Common Taxonomy of Competency Domains for the Health Professions and Competencies for Physicians. *Academic Medicine, 88, [8],* 1088-1094.

Ennis, R. (1985). A Logical Basis for Measuring Critical Thinking Skills. *Educational Leadership, 43,* 44-48.

Ennis. RH, (1987). *A Taxonomy of Critical Thinking Dispositions and Abilities.* In J. Baron and R. Sternberg (Eds.) Teaching Thinking Skills: Theory and Practice. New York. WH Freeman.

Epstein. RM, Dannefer. EF, Nofziger. AC, Hansen. JT, Schultz. SH, Jospe. N, Connard. LW,

Meldrum. SC, and Henson. LC, (2004). Comprehensive Assessment of Professional Competence: The Rochester Experiment. *Teaching and Learning in Medicine, 16, [2],* 186-196.

Epstein. RM, (2007) Assessment in Medical Education. *New England Journal of Medicine* 356. [4], 387-396.

Eraut. M, (2004). *Developing Professional Knowledge and Competence.* Philadelphia: Taylor and Francis Group.

Ericsson. KA, (2004). Deliberate Practice and the Acquisition and Maintenance of Expert Performance in Medicine and Related Domains. *Academic Medicine; 79, (Suppl. 10)*, S70–S81.

Erlandson. P, (2005). The Body Disciplined: Rewriting Teaching Competence and the Doctrine of Reflection. *Journal of Philosophy of Education,* 39, 661–670.

Ettorchi-Tardy A, Levif. M, and Michael. MD, (2012). Benchmarking: A Method for Continuous Quality Improvement. *Healthcare Policy, 7, [4],* e101-e119.

European University Association (EUA, 2010) *Trends 2010: A Decade of Change in Higher Education.* Brussels, Belgium. EUA Asbl.

European Higher Education Area (EHEA, 2015) *The Bologna Process Implementation Report.* Luxembourg: Publications Office of the European Union.

European Qualifications Framework (EQF, 2018) T*he Qualifications Framework: Supporting Learning, Work and Cross-border Mobility.* Luxembourg: Publications Office of the European Union.

Facione. N, and Facione. P, (1996). Externalising Critical Thinking in Knowledge Development and Clinical Judgment. *Nursing Outlook,* 44:129–36.

Farrell. T, (2012). Reflecting on Reflective Practice: (Re) Visiting Dewey and Schön. *TESOL Journal,* 3, 7–16.

Fredsco. N, Kentner. A, and Natt. J, (2017).The Effect of Relevance Strategies on Student Perceptions of Introductory Courses. *Communication Education, 66, [2],* 196-209

Fenizia. E, Marchetti. A, Biagioli. V, Raso. A, Gambera. A, DeMarinis. MG, and Piredda. M, (2019). Psychometric Testing of the Caring Behaviors Inventory for Nursing Students. J*ournal of Clinical Nursing,* 28 (19-20), 3567-3574.

Frenk. J, Chen. L, Bhutta. ZA, Cohen. J, Crisp. N, Evans. T. Garcia. P. Ke. Y, Kelley. P, Kistnasamy. B, Meleis. A, Naylor. D. Pablos-Mendez. A, Reddy. S, Scrimshaw. S, Supulveda. J, Serwadda. D, and Zurayk. H, (2010). Health Professionals for a New Century: Transforming Education to Strengthen Health Systems in an Interdependent World. T*he Lancet, 375, [9756]*, 1923-1958.

Fernandez. N, Dory. V, Ste-Marie. LG, Chaput. M, Charlin. B, and Boucher. A, (2012). Varying Conceptions of Competence: An Analysis of How Health Sciences Educators Define Competence. *Medical Education, 46,[4]*, 357-365.

Fink. A, (2010). *Conducting Research Literature Reviews [3rd Edn.].* London: Sage Publication.

Field. A, Miles. J, and Field, Z. (2012). *Discovering Statistics using R (pp.62.115)* London: Sage Publication.

Flin. R, and Patterson-Brown. S. (2016). *Situational Awareness.* In Flin. R, Youngson. GG, and Yule. S, (Eds). Enhancing Surgical Performance. A Primer in Non-technical Skills, p. 63–83. Florida, CRC Press, Taylor and Francis Group.

Fischer. CF, and King. RM, (1995). *Authentic Assessment: A Guide to Implementation*. Thousand Oaks. Corwin Press

Fook. J, White. S and Gardner. F, (2006). *Critical Reflection: A Review of Contemporary Literature and Understandings. In S. White, J. Fook and F. Gardner (Eds.) Critical Reflection in Health and Social Care.* Maidenhead, Berks: Open University Press.

Fook. J, (2004). *Critical Reflection and Organisational Learning and Change: A Case Study, in N. Gould, and M. Baldwin (eds) Social Work, Critical Reflection and the Learning Organisation,* 57– 74, Aldershot: Ashgate.

Ford. P, Johnston. B, Brumfit. C, Mitchell. R, and Myles. F, (2005). ' Practice Learning and the Development of Students as Critical Practitioners– Some Findings from research', *Social Work Education, 24, [4],* 391– 407.

Foucault. M, (1979). *Discipline and Punish.* New York: Vintage Books.

Fragkos. KC, (2016). Reflective Practice in Healthcare Education: An Umbrella Review. *Education Sciences, 6, [3],* 27-43.

Francis. R, (2013). *Independent Inquiry into Care Provided by Mid Staffordshire NHS Foundation Trust.* London. United Kingdom: DH.

Freire. P (1985). *Pedagogy of the Oppressed*. London. Penguin Books

Fredericksen. J, and Collins. A, (1989). A systems approach to educational testing. *Educational Researcher, 18, [9]*, 27–32.

Fry. H, Ketteridge. S, and Marshall. S, (2009). *A Handbook of Teaching and Learning in Higher Education. Enhancing Academic Practice* (3rd Edn). London, Routledge. Taylor Francis Publication.

Fry. M, and MacGregor. C, (2014). Confidence and Impact on Clinical Decision-making and Behaviour in the Emergency Department. *Australasian Emergency Nursing Journal, 17, [3],* 91–97.

Gagne. RM, (1985). *Conditions of Learning and Theory of Instruction (4th Edn).* New York: Holt, Rinehart and Winston.

Gardner. H, (1983). *Frames of Minds: The Theory of Multiple Intelligences.* New York: Basic Books.

Gardner. H, (1993). *Multiple Intelligences: The Theory in Practice.* New York: Basic Books.

Gardner. G, Chang. AM, Duffield. C, and Doubrovsky. A, (2013). Delineating the Practice Profile of Advanced Practice Nursing: A cross-sectional Survey using the Modified Strong Model of Advanced Practice. *Journal of Advanced Nursing*, *69, [9],* 1931-1942.

Garside. JR, and Nhemanchena. JZZ, (2013). A Concept Analysis of Competence and its Transition in Nursing. *Nurse Education Today, 33*, 514-545.

George. JB, (1995). *Nursing Theories, the Base for Professional Nursing Practice (4th Edn).* Norwalk, Connecticut: Appleton and Lange.

Gerrard. J, (2017). *Health Sciences Literature Review Made Easy: The Matrix Method, 5th Edn.* London: Sage Publication.

Gerring. J, (2004). What Is a Case Study and What Is It Good for? *American Political Science Review, 98, [2]*, 341- 354.

General Data Protection Regulations (GDPR, 2016). Regulation (EU) 2016/679 of the European Parliament of the Council. *Official Journal of the European Union*

Ghaye. T, and Lilleyman. S, (2010). *Reflection: Principles and Practices for Healthcare Professionals* London. Quay Books Division.

Gibbs. G, (1988). *Learning by Doing: A Guide to Teaching and Learning Methods.* Oxford: Oxford Further Education Unit.

Gibbs. E, (2007). *Analysing Qualitative Data.* London: Sage Publications. Ltd.

Gibbert. M, Ruigork. W, and Wicki. B, (2008). What Passes as A Rigorous Case Study. *Strategic Management Journal, 29, [13],* 1465-1474.

Gielen. S, Dochy. F, and Dierick. S, (2003). Evaluating the Consequential Validity of New Modes of Assessment: The Influence of Assessment on Learning, Including Pre-, Post- and True Assessment Effects. In M. Segers, F. Dochy, and E. Cascallar (Eds.) *Optimising New Modes of Assessment: In Search of Quality and Standards.* Dordrecht: Kluwer Academic Publishers.

Gillespie. A, Graham. S, Kiuhara. S, and Hebert. M, (2014). High School Teachers use of Writing to Support Students' Learning: A National Survey. *Reading and Writing*, *27, [6],* 1043- 1072.

Ginsburg. LR, Tregunno. D, Norton. RG, Smee. S, de Vries. I, Sebok. S, VanDenKerkhof. EG Luctkar-Flude. M and Medves. J, (2015). Development and Testing of an Objective Structured

Clinical Exam (OSCE) to Assess Socio-cultural Dimensions of Patient Safety Competency. *British Medical Journal of Quality and Safety, 24*, 188–194.

Girelli. L Alivernini. F, Lucidi. F, Cozzolino. M, Savarese. G. Silillo. M and Salvatore. S, (2018). Autonomy Supportive Context, Autonomous Motivation and Self Efficacy Predict Adjustment of First-year University Students. *Frontiers in Education*, 3, 1-11.

Gipp. VG, (2012). *Beyond Testing: Towards a Theory of Educational Assessment.* London. Routledge, Taylor and Francis Group.

Giroux. HA, (2005). The Terror of Neoliberalism: Rethinking the Significance of Cultural Politics. *College Literature, 32, [1]*, 1–19.

Gobbi. M, and Kaunonen. M, [Eds] (2018). *TUNING Guidelines and Reference Points for Design and Delivery of Degree Programs in Nursing*. Tuning and CALOHEE Project.

Goertz. G, and Mahoney. J, (2012). Concepts and Measurement: Ontology and Epistemology. *Social Science Information*, 519-521.

González-González. I, and Jiménez-Zarco. AI, (2015). Using Learning Methodologies and Resources in the Development of Critical Thinking Competency: An Exploratory Study in a Virtual Learning Environment. *Computers in Human Behaviour,* 51, 1359-1366.

Gormley. G, Maney. A, Layard. B, Hart. N and McCourt. C, (2013). Temporary Tattoos: A Novel OSCE Assessment Tool. *Clinical Teacher, 10, [4],* 251-257.

Graham. S, Capizzi. A, Harris. K, Hebert. M, and Morphy. P, (2014). Teaching Writing to Middle School Students: A National Survey. *Reading and Writing*, 27, *[6],* 1015-1042

Granello. DH. (2010). Cognitive Complexity Among Practicing Counselors: How Thinking Changes with Experience. *Journal of Counseling and Development, 88, [1]*, 92-100.

Grant. M, (2016). Aligning Skills Development with Labour Market Need. Ottawa: The Conference Board of Canada.

Greenhalgh. T, (2007). How to Read a Paper: The Basics of Evidence-Based Medicine [5th Edn.] London. BMJ Books.

Greeno. JG, (1978). *A study of Problem Solving. In Glaser. R [Ed,]. Advances in Instructional Psychology. Vol 1.* Hillsdale, New Jersey. Erlbaum Publication.

Greeno. JG, and Gelman. R, (1989). On the Nature of Competence: Principles for Understanding in a Domain. In Resnick. LB. [Ed], *Knowing Learning and Instruction: Essays in Honour of Robert Glaser* (pp. 125-187). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Gregory. DM, Guse. LW, Dick. DD, and Russell. CK, (2007). Patient Safety: Where is Nursing Education? *Journal of Nurse Education, 46, [2]*, 79-82.

Grugulis. I, and Stoyanova. D, (2011). Skill and Performance. *British Journal of Industrial Relations* 49, 515–36.

Guide to the General Data Protection Regulation (GDPR, 2018). International Commissioner's Office. Gov.UK.

Guisande. MA, Páramo. MF, Tinajero. C, and Almeida. LS, (2007). Field Dependence-Independence (FDI) Cognitive Style: An Analysis of Attentional Functioning. *Psychothema, 19, [4]*, 572-577.

Gulikers. JTM, Bastiaens. TJ, and Kirschner. TA, (2004). The Five-Dimensional Framework for Authentic Assessment. *Educational Technology Research Development, 52, [3], 67-86.*

Gulwaldi. BG, (2009). Using Reflective Journals in a Sustainable Design Studio. *International Journal of Sustainability in Higher Education*, 96-106.

Gumber. R, Khoosal. D, and Gajebasia. N, (2012). Non-medical Prescribing: Audit, Practice and Views. *Journal of Psychiatric and Mental Health Nursing*, *19, [*6*]*, 475-481.

Guskey. TR, (2001). High percentages are not the same as high standards. *Phi Delta Kappan, 82, [7], 534–536.*

Guskey. TR, (2010). Lessons of Mastery Learning. *Educational Leadership, 68, [2],* 52-57.

Guskey. TR, and Anderman. EM, (2012) In Search of a Useful Definition of Mastery. *Educational Leadership, 71, [4],* 18-23.

Guskey, TR, [Ed.], (2012a). *Benjamin. S. Bloom: Portraits of an Educator (2nd ed.).* Lanham, MD, Rowman and Littlefield Education.

Guskey. TR, (2013). In Search of a Useful Definition of Mastery. *Educational Leadership, 71, [4]*, 18-23.

Guskey, TR, (2013a). The case against percentage grades. Educational Leadership, 71, *[1],* 68–72

Gutierrez. X, (2016). Analysed Knowledge, Metalanguage, and Second Language Proficiency. *System,* 60: 42-54.

Habermas. J, (1972). *Knowledge and Human Interest.* London: Heinemann.

Habermas. J, (1974). Theory and Practice. London: Heinemann.

Hacking. S and Taylor. J, (2010). *An Evaluation of the Scope and Practice of Non-Medical Prescribing in the North West for the NHS North West*. The University of Central Lancashire.

Haffling. AN. Beckman. A, Pahlmblad. A, and Edgren. G, (2010). Students' Reflections in a Portfolio Pilot: Highlighting Professional Issues. *Medical Teacher, 32*, e532–e540.

Haladyna. T, (1976). *The Paradox of Criterion-Referenced Measurement. A Paper Presented at the Annual Meeting of the National Council on Measurement in Education, (pp 1-25)*. San Francisco, California: ERIC number ED126155.

Hall. C, and Ritchie. D, (2013). *What is Nursing? Exploring Theory and Practice (3rd Edn).* Thousand Oaks. Sage Publication.

Hallgren. KA, (2012). Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutorials in Quantitative Methods for Psychology*, *8, [1],* 23-34.

Halpern. DF, (2014) *Thought and Knowledge: An Introduction to Critical Thinking (5th Edn.)* London. Psychology Press. Taylor and Francis Group.

Hambleton. RK and Cook. LL. (1977). Latent Trait Models and their use in the Analysis of Educational Test Data. *Journal of Educational Measurement, 14, [2]*, 75-94.

Hambleton. RK, Swaminathan. H, and Rogers. HJ, (1991). *Fundamentals of Item Response Theory.* London: Sage Publications.

Hardern. RM, Lilley. P, and Patricio. M, (2016). *The Definitive Guide to the OSCE.* Oxford: Elsevier.

Hassan. S and Hod. R, (2017). Use of Item Analysis to Improve the Quality of Single Best Answer Multiple Choice Question in Summative Assessment of Undergraduate Medical Students in Malaysia. *Education in Medicine, 9, [3],* 33-43.

Hatton. N and Smith. D, (1995). "Reflection in Teacher Education: Towards Definition and Implementation", *Teaching and Teacher Education, 11*, 33-49.

Hawkins. RE, Welcher. CM, Holmboe. ES, Kirk. LM, Norcini. JJ, Simmons. KB and Skochelak. SE, (2015). Implementation of Competency-based Medical Education: Are we Addressing the Concerns and Challenges. *Medical Education in Review, 49 [11]*, 1086-1102.

Health Education England (HEE, 2017). *Multi-professional Framework for Advanced Clinical Practice in England. London. HEE.*

Health and Social Care Act (HSCA, 2001). Printed in the UK by The Stationery Office Limited.

Heidi. C, (2008). What does an OSCE Checklist Measure? F*amily Medicine, 40, [8]*, 589-591.

Heilmann. C, (2015). "A New Interpretation of the Representational Theory of Measurement." *Philosophy of Science*, *82, [5],* 787-797.

Helminen. K, Coco. K, Johnson. M, Turunen. H and Tassovainen. K, (2016). Summative Assessment

of Clinical Practice of Student Nurses: A Literature Review. *International Journal of Nursing Studies, 53,* 308-319.

Henderson. V, (1966). *The Nature of Nursing*. New York. The Macmillan Company.

Hettinga. AM, Denessen. E, and Postma. CT, (2010). Checking the Checklist: A Content Analysis of Expert and Evidence-based Case-Specific Checklist Items. *Medical Education, 44*, 874-883.

Hewitt-Taylor. J, (2017). *The Essential Guide to Doing a Health and Social Care Review.* Abingdon, Oxon: Routledge.

Hickerson. KA, Taylor. LA, and Terhaar. MF, (2016). The Preparation–Practice Gap: An Integrative Literature Review. *Journal of Continuing Education in Nursing, 14147, [1]*, 17-23.

Hinchliffe. S, and Rogers. R, (2008). *Competencies for Advanced Nursing Practice.* London: Royal College of Nursing (RCN).

Hobson. RJ, Scott. J, and Sutton. J, (2010). Pharmacists and Nurses as Independent Prescribers: Exploring the patient's perspective. *Family Practice, 27, [1],* 110–120.

Hockings. C, (2010). Inclusive Learning and Teaching in Higher Education: A Synthesis of Research. *Higher Education Academy Online Assets*, p.1-67.

Hodges. B, and McIlroy. HJ, (2003). Analytic Global OSCE Ratings are Sensitive to Level of Training. *Medical Education, 37, [11]*, 1012–1016.

Hodges. B, Regehr. G, McNaughton. N, Tiberius. R, and Hanson. M, (1999). OSCE Checklists do not Capture Increasing Levels of Expertise. *Academic Medicine, 74, [10],* 1129–1134.

Hodges. B, Regehr. G, Hanson. M, and McNaughton. N, (1998) Validation of an objective structured clinical examination in psychiatry. Academic Medicine, 73, [8], 910– 912.

Hoffman. RR, Ward. P, Feltovich. PJ, DiBello. L, Fiore. SM, and Andrews. DH, (2014). *Accelerated Expertise: Training for High Proficiency in a Complex World.* East Sussex. Taylor and Francis Publication.

Holland. PW, (1990). On the Sampling Theory Foundations of Item Response Theory Models. *Psychometrika, 55*, 577–601.

Holloway. I, and Todres. L, (2003). The Status of Method: Flexibility, Consistency and Coherence. *Qualitative Research, 3, [3]*, 345-357.

Holmes, L. M, (2015). "Becoming a Graduate: The Warranting of an Emergent Identity." *Education and Training, 57, [2],* 219–238.

Hooper. L, Begg. MD, and Sullivan. LM, (2014). Integrating Competencies and Learning Outcomes in Core Courses for the Masters of Public Health. *Public Health Reports*, 129, [4], 376- 181.

Hopper. MK, and Kaiser. AN, (2018). Engagement and Higher Order Skill Proficiency of Students Completing a Medical Physiology Course in Three Diverse Learning Environments. *Advanced Physiology Education, 42: 429-438.*

Hopson. M, Simms. R, and Knezek. G, (2002). 'Using a Technology-Enriched Environment to Improve Higher-Order Thinking Skills', *Journal of Research on Technology in Education*, *34,[4]*, 109 -120.

Houghton. C, Casey. D, Shaw. D, and Murphy. K, (2013). Rigour in Qualitative Case Study Research. *Nurse Researcher, 20, [4],* 12-17.

Huang. C, (2012). Discriminant and Criterion-related Validity of Achievement Goals in Predicting Academic Achievement: A meta-analysis. *Journal of Educational Psychology, 104, [*1*]*, 48–73.

Hughes. C, (2013). A Case Study of Assessment of Graduate Learning Outcome at the Program Course and Task Level. *Assessment and Evaluation in Higher Education, 38, [4],* 492-506.

Hughes. SJ, and Quinn. FM, (2013). *Quinn's Principles and Practice of Nurse Education (6th Edn).* The United Kingdom. Cengage Learning.

Hulme. PE, (2014). Bridging the Knowing-doing Gap: Know-who, Know-what, Know-why, Know-how, Know-when. J*ournal of Applied Ecology*, 51, 1131-1136.

Hulley. SB, Cummings. SR, Browner.WS, Grady .DG, and Newman. TB, (2013). Designing Clinical Research *(4thEdn).* Philadelphia, PA: Lippincott Williams and Wilkins.

Huot. B, (2002). A*rticulating Writing Assessment for Teaching and Learning*. Logan, Utah: Utah State University Press.

Hutchinson. C, and Young. M, (2011). Assessment for learning in the accountability era: Empirical evidence from Scotland. *Studies in Educational Evaluation, 37,* 62–70.

Ilgen. JS, Ma. IWY, Hatala. R and Cook. DA, (2015). A Systematic Review of Validity Evidence for Checklists Versus Global in Simulation-based Assessment. *Medical Education, 49*, 163-173.

Imison. C, Castle-Clarke.S, and Watson. R, (2016). *Workforce to Deliver the Care Patients Need.* Research Report. London, Nuffield Trust.

International Labour Office (ILO, 2010). *A Skilled Workforce for as Strong, Sustainable and Balanced Growth: A G20 Training Strategy*. Geneva, ILO.

Ionescu. T, (2012). Exploring the Nature of Cognitive Flexibility. *New Ideas in Psychology, 30, [2],* 190-200.

Iramaneerat. C, Smith, Jr. EV, and Smith. ER, (2008). *An Introduction to Rasch Measurement* Chapter 4. *Cited in Osborn. JW, (Ed) Best Practices in Quantitative Research Methods* (p. 50-71).Thousand Oaks. Sage Publications, Inc.

Ironside. PM, and McNelis. AM, (2011). Transforming clinical education. *Journal of Nursing Education, 50,* 123-124.

Jackson. D, (2012). Testing a Model of Undergraduate Competence in Employability Skills and its Implications for Stakeholders. *Journal of Education and Work, 27, [2], 1-23.*

Jackson. D, (2013). Business Graduate Employability – Where are We Going Wrong?" *Higher Education Research and Development 32, [5],* 776–90.

Jang. H, Reeve. J, and Halusic. M, (2016). A New Autonomy-supportive Way of Teaching that Increases Conceptual Learning: Teaching in Students' Preferred Ways. *Journal of Experimental Education, 84, [4],* 686–701.

James. ME, (2013). *Educational Assessment, Evaluation and Research. The Selected Works of Mary E. James.* London, Routledge, Taylor Francis Group.

Jarman. B, Gault. S, Alves. B, Hider. A, Dolan. S, Cook. A, Hurwitz. B, and Iezzoni. LI, (1999). Explaining differences in English hospital death rates using routinely collected data. *British Medical Journal, 318, [7197]*, 1515–1520.

Jarvis. DSL, (2014). Regulating Higher Education: Quality assurance and Neo-liberal Managerialism in Higher Education—A Critical Introduction. *Policy and Society, 33, [*3*]*, 155–166.

Jayaweera. HK, Potts. HWW. Keshawani. K, Valerio. C, Baker. M, Mehdizadeh. L, and Sturrock. A. (2018). The GP Tests of Competence Assessment: Which Part best Predict Fitness To Practice. *BMC Medical Education, 18 [2]*, 1-9.

J-Metric: Open Source Software for Psychometric Analysis. Compact Disc (CD). In Meyer. JP, (2014). *Applied Measurement with J-Metrick.* New York: Routledge, Taylor and Francis Group.

Joanna Briggs, (2017). *joannabriggs.org/research/critical-appraisal-tools.html*. Retrieved February 22nd, 2018 from The Joanna Briggs Institute: www.joannabriggs.org.

Johns. C, (1995). Framing learning Through Reflection within Carper's Ways of Knowing in Nursing. *Journal of Advanced Nursing, 22, [2]*, 226-234.

Johns. C, (2004). *Becoming a Reflective Practitioner (2nd Edn).* Oxford: Blackwell Publishing.

Kamarudin. G, Penn. J, Chaar. B, and Moles. R, (2016). Educational Interventions to Improve Prescribing Competency: A Systematic Review. *BMJ Open Access, 3, [8],* e003291.

Kane. MT, (1990). Generalising Criterion-Related Validity Evidence for Certification Requirements Across Situations and Speciality Areas. *ACT Research Report Series, 90, [3]*, 1-20.

Kane, MT, (1992). An Argument-Based Approach to Validity. *Quantitative Methods in Psychology, 112, [3]*, 527-535.

Kane. MT, (1994). Validating Interpretive Arguments for Licensure and Certification Examinations. *Evaluation of the Health Professions, 17, [2]*, 133-159.

Kane. MT, (1994a). Validating the Performance Standards Associated With Passing Scores. *Review of Educational Research, 64, [3]*, 425-461.

Kane. MT, (2001). Current Concerns in Validity Theory. *Journal of Educational Measurement, 38, [4]*, 319-342.

Kane. MT, (2006). Validation. In R. Brennan, *Educational Measurement, (4th Edn),* (pp. 17–64). Westport. Connecticut: Greenwood Publishing.

Kane. MT, (2006a). Content Related Validity Evidence in Test Development. In a. H. Downing. SM, *Handbook of Test Development* (p. 131-155). Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Kane. RL, Shamliyan. TA, Mueller. C, Duval. S, and Wilt. TJ, (2007) The Association of Registered Nurse Staffing Levels and Patient Outcomes: Systematic Review and Meta-analysis. *Medical Care*, 45, [12], 1195–204.

Kane. MT, (2013). Validating the Interpretations and Uses of Test Scores. *Journal of Educational Measurement, 50, [1]*, 1-73.

Karppinen, K, and Moe. H, (2019). Texts as Data I: Document Analysis. In *The Palgrave Handbook of Methods for Media Policy Research* (p. 49-262). Palgrave Macmillan, Cham.

Kehm. BM, (2010). Quality in European Higher Education: The Influence of the Bologna Process. C*hange: The Magazine of Higher Learning, 42, [3],* 40-46.

Kelchtermans. G, (2008). Who I am in How I Teach is the Message: Self Understanding, Vulnerability and Reflection. *Teachers and Teaching: Theory and Practice, 15 [2]*, 257-272.

Kember. D, (1999). Determining the Level of Reflective Thinking from Students' Written Journals using a Coding Scheme Based on the Work of Mezirow. *International Journal of Lifelong Education, 18, [1]*, 18-38.

Kember. D, McKay. J, Sinclair. K, and Wong. FKY, (2008). A Four-category Scheme for Coding and Assessing. *Assessment and Evaluation in Higher Education, 33, [4]*, 369-379.

Kendall-Gallagher. D, Aiken. LH, Sloane. M, and Cinotti. JP, (2011). Nurse Speciality Certification, Inpatient Mortality and Failure to Rescue. *Journal of Nursing Scholarship, 43 [2]*, 188-194.

Kent. PS, McCarthy. M, Burrows. J, McCann. L, Pavilnac. J, Goeddeke-Merickel. CM, Weisen. K, Kruger. S, Byham-Gray. L, Pace. RC, Hannah. V and Benner. D, (2014). Academy of Nutrition and Dietetics and National Kidney Foundation: Revised 2014 Standards of Practice and

Standards of Professional Performance for Registered Dietitian Nutritionists (Competent, Proficient, and Expert) in Nephrology Nutrition. *Journal of the Academy of Nutrition and Dietetics, 114, [9],* 1448-1457.

Keogh. B, (2013). Review into the quality of care and treatment provided by 14 hospital trusts in England: Overview Report. http://www.nhs.uk/NHSEngland/bruce-keogh-review/Documents/outcomes/keogh-review-fi nal-report.pdf (accessed Jan 13, 2014).

Keutel. M, Michalik. B, and Richter. J, (2014). "Towards Mindful Case Study Research in IS: A Critical Analysis of the Past Ten Years", *European Journal of Information Systems, 23, [3]*, 256–272.

Khattab. AD, and Rawlings. B, (2008). Use of a Modified OSCE to Assess Nurse Practitioner Students. *British Journal of Nursing, 17*, 754–759.

Kim. J, Nelipovitz. D, Cardinal. P, and Chu. M, (2009). Comparison of Global Rating Scale and Checklist Scores in the Validation of an Evaluation Tool to Assess Performance in the Resuscitation of Critically Ill Patients During Simulated Emergencies (Abbreviated as "CRM Simulator Study IB"). *Simulation in Healthcare, 4, [1]*, 6-16.

King. PM, and Kitchener. KS, (1994). *Developing Reflective Judgment: Understanding and Promoting Intellectual Growth and Critical Thinking in Adolescents and Adults.* Higher and Adult Education Series and Jossey-Bass Social and Behavioral Science Series. San Francisco, Jossey-Bass Publication.

Kinsella. EA, (2009). Professional Knowledge and the Epistemology of Reflective Practice. *Nursing Philosophy,* 11:3-14.

Kitchener. KS, and King. PM, (1990). *The Reflective Judgement Model: Transforming Assumptions About knowing '*, in Jack Mezirow (ed.) Fostering Critical Reflection in Adulthood: A Guide to Transformative and Emancipatory Learning, 159– 76, San Francisco, CA: Jossey-Bass.

Kizior. R, (2016). *Saunders Nursing Drug Handbook* (1st Edn). USA, Elsevier Publication.

Klein. SP, Stecher.BM, Shavelson. RJ, McCaffrey. D, Ormseth. T, Bell. RM, Comfort. C and Othman. AR, (1998). Analytic Versus Holistic Scoring of Science Performance Tasks. *Applied Measurement in Education, 11, [2]*, 121-137.

Knight. PT, and Yorke. M, (2003). *Assessment, Learning and Employability.* Maidenhead: Society for Research into Education and Open University Press.

Koens. F, Custers. EJFM and tenCate. O, (2009) Clinical and Basic Science Teachers Opinions about the Required Depth of Biomedical Knowledge for Medical Students. *Medical Teacher, 28, [3],* 234-238.

Kogan. J, Holmboe. E, and Hauer. K, (2009). Tools for Direct Observation and Assessment of Clinical Skills of Medical Trainees: A Systematic Review. *JAMA The Journal of the American Medical Association 302, [12],* 1316-1326.

Kolb. DA, (1984). *Experiential Learning: Experience as a Source of Learning and Development.* Michigan: Prentice-Hall.

Koretz. D, (2008). *Measuring Up: What Educational Testing Really Tells Us*. Cambridge Massachusetts. Harvard University Press.

Koretz, D., (2015). Adapting Educational Measurement to the Demands of Test-based Accountability. *Measurement: Interdisciplinary Research and Perspectives, 13,[1],* 1-25.

Kramer. M, (2016) A Capacity to Get Things Right: Gilbert Ryle on Knowledge. *European Journal of Philosophy*, *25, [1],* 25-46.

Kreber. C, (2005). Reflection on Teaching and the Scholarship of Teaching: Focus on science Instructors. *Higher Education*, *50*, 323-359.

Kristiansen. L, Häggström. M, Hallin. K, Andersson. L, Bäckström. B, (2015). Swedish Translation, Qualitative Relevance Evaluation and Quantitative Reliability Test of Lasater Clinical Judgment Rubric. *Nordic Journal of Nursing Research, 35, [2], 113–22.*

Kroenzen. M, van Dijk. L, Groenewegen. PP, and Francke. AL, (2011). Nurse Prescribing of Medicines in Western in European and Anglo-Saxon Countries: A Systematic Review of the Literature. *BioMed Central, 11*, 1-17.

Kubinger. KD, and Gottschall. CH, (2007). Item Difficulty of Multiple Choice Tests Dependant on Different Item Response Formats: An experiment in Fundamental Research on Psychological Assessment. *Psychology Science, 49, [4]*, 361-374.

Kulasegaram. K, Manzone. JC, Ku. C, Skye. A, Wadey. V, and Woods. NN, (2015). Cause and Effect: Testing a Mechanism and Method for the Cognitive Integration of Basic Science. *Academic Medicine,* 90 (11 Suppl.):63–69.

Kumar. M, Singh. AJ, and Handa. D, (2017). Literature Survey on Student's Performance prediction in Education using Data Mining Techniques. *International Journal of Education and Management Engineering 6, [6]*, 40-49.

Kutney-Lee. A, Sloane. DM, and Aiken. LH, (2013). An Increase in the Number of Nurses with Baccalaureate Degrees is linked to the Lower Mortality Rates of Post-Surgery Mortality. *Health Affairs, 32, [3],* 579-589.

Lane. S. and Stone. CA, (2003). Consequences of a State Accountability Program: Examining Relationships between School Performance Gains and Teacher, Student, and School Variables. *Applied Measurement in Education, 16 [1],* 1–26.

Lane. S, (2013). The Need for a Principled Approach for Examining Indirect Effects of Test Use, Measurement: *Interdisciplinary Research and Perspectives, 11 [1-2],* 44-46.

Langley. ME, and Brown. ST, (2010). Perceptions of the use of Reflective Journals in Online Graduate Nursing Education. *Nursing Education Perspectives, 31, [1],* 12-27.

Latter. S, Blenkinsopp. A. Smith. A, Chapman. S, Tinelli. M, Gerard. K. Little. P, Celino. N, Granby. T, Nicholls. P, and Dorer. G, (2011). Evaluation of Nurse and Pharmacist Independent Prescribing. Department of Health Research Programme Project 0160108. University of Southampton and Keel University.

Lasater. K, (2007). Clinical Judgment Development: Using Simulation to Create an Assessment Rubric. *Journal of Nursing Education, 46 [11], 496-503.*

Larsen. DP, London. DA, and Emke. AR, (2016). Using Reflection to Influence Practice: Student Perceptions on Daily Reflection in Clinical Practice. *Perspectives on Medical Education, 5, [5],* 285-291.

Lave. J, and Wenger. E, (1991). *Situated Learning: Legitimate Peripheral Participation.* Cambridge: Cambridge University Press.

Leach. J and Moon. B, (2008). *The Power of Pedagogy.* London. Sage Publication.

LeDeist. FD, and Winterton. J, (2005). What is Competence? *Human Resource Development International, 8 [1]*, 27-46.

Lee. NJ, (2011). An Evaluation of CPD Learning and Impact upon Positive Practice Change. *Nurse Education Today, 31, [4],* 390-395).

Leijen. A , Valtna.K , Leijen DAJ and Pedaste. M, (2012). How to determine the quality of students' Reflections?, *Studies in Higher Education, 37 [2],* 203-217.

Levett-Jones. T, (2014). *Critical Conversations for Patient Safety.* Pearson Australia, Melbourne.

Levett-Jones. T, Hoffman. K, Dempsey. J, Sim- Jeong. Y, Noble. D, Norton. CA, Roche. J, and Hickey. N, (2010). The Five Rights of Clinical Reasoning, An Educational Model to Enhance Nursing Students Ability to Identify and Manage Clinically at Risk Patients. *Nurse Education Today, 30*, 515-520.

Levine. DM, Krehbiel. TC, and Berenson. ML, (2002). *Business Statistics: A First Course (3rd Edn).* Upper Saddle Row, New Jersey: Prentice-Hall.

Lewis. R, (2012) Vital Signs are Vital for a Reason*, Nursing Standard,* 26, [1], 72-72.

Liao. LM, Sun. XY, Yu. H, and Li. JW, (2016). The Association of Nurse Educational Preparation and Patient Outcomes: Systematic Review and Meta-analysis. *Nurse Education Today,* 42, 9–16.

Liaw. SY, Scherpbier. A, Klainin-Yobas. P, and Rethans. JJ, (2011). A Review of Educational Strategies to Improve Nurses' Roles in Recognizing and Responding to Deteriorating Patients. *International Nursing Review 58, [3],* 296–303.

Likic. R, and Maxwell. SR, (2009). Prevention of Medication Errors: Teaching and Training. *British Journal of Clinical Pharmacology, 67*, 656-661.

Lineberry. M, Park. YS, Cook. DA, and Yudowsky. R, (2015). Making the Case for Mastery in Learning Assessment: Key Issues in Validation and Justification. *Academic Medicine*, 90, 1445-1450.

Linn. R, Baker. EL and Dunbar. SB, (1991). Complex, Performance-Based Assessment: Expectations and Validation Criteria. *Educational Researcher, 20*, 16-21.

Lissitz. RW, and Samuelsen. K, (2007). A suggested change in terminology and emphasis regarding validity and education. *Educational Researcher, 36*, 437–48.

Litzinger. TA, Lattuca. LR, Hadgraft. RG, and Newstetter. WC, (2011). Engineering Education and the Development of Expertise. *Journal of Engineering Education*, *100, [1],* 123–150.

Little. B and Locke. W, (2011). Excellence in Teaching and Learning. In Questioning Excellence in Higher Education: Policies, Experiences and Challenges in National and Comparative Perspective. Rostan and Vairia [Eds] 119-137. Rotterdam, Sense Publishers.

Lizzio. A, and Wilson. K, (2013). First-year Students' Appraisal of Assessment Tasks: Implications for Efficacy, Engagement and Performance. *Assessment and Evaluation in Higher Education*, 38, *[4]*, 389-406.

Lokhoff. J Wegewijs. B, Durkin. K Wagenaar. R, González. J, Isaacs. AK, Donà dalle Rose. DF, and Gobbi. M, (2010). A Tuning Guide to Formulating Degree Programme Profiles Including Programme Competencies and Programme Learning Outcomes. Publicaciones de la Universidad de Deusto.

Lui. LYY, Winnie. KW, and Fong. DYT, (2008). Knowledge and Attitudes Regarding Pain Management Among Nurses in Hong Kong Medical Units. *Journal of Clinical Nursing, 17 [15], 2014-2021.*

Lundgren. K, and Poell. RF, (2016), "On Critical Reflection: A Review of Mezirow's Theory and it's Operationalisation". *Human Resource Development Review*, *15 [1],* 3-28.

Lyotard. JF. (1984). *The Postmodern Condition: A Report on Knowledge: Theory and History of Literature Volume 10.* Manchester: Manchester University Press.

Mackillop. L, Crossley. J, Vivekananda-Schmidt. P, Wade. W, and Armitage. M, (2011). A Single Generic Multi-source Feedback Tool for Revalidation of all UK Career-grade Doctors: Does One Size Fit All? *Medical Teacher, 33*:75–83.

Maguire. M, Perryman. J, Ball. SJ, and Braun. A, (2012). The Ordinary School- What is it? *British Journal of Sociology of Education, 32, [1],* 1-16.

Maier, NRF, (1933). An Aspect of Human Reasoning. *British Journal of Psychology*, 24, 144-155.

Malau-Adult. BS, Mulcahy. S, Warnecke. E, Otahal. P, Teague. PA, and Turner. R, (2012). Inter-rater Reliability: Comparison of Checklist and Global Scoring for OSCEs. *Creative Education Journal: Special Issue 3*, 937–942.

Mango. C, (2009). Demonstrating the Difference between Classical Test Theory and Item Response Theory Using Derived Test Data. *International Journal of Educational and Psychological Assessment, 1 [1]*, 1-11.

Mantzoukas. S, and Watkinson. S, (2007). Review of Advanced Nursing Practice: The International Literature and Developing the Generic Features. J*ournal of Clinical Nursing. 16, [1]*, 28-37.

Manzano-Vázquez. B, (2014). Lexical Transfer in the Written Production of a CLIL Group and a non-CLIL group. *International Journal of English Studies, 14, [2],* 57–76.

Mare. DS. Moreira. F, and Rossi. R, (2016). Nonstationary Z-Score Measures. *European Journal of Operational Research*, 1-11.

Marieb. E, and Hoehn. K, (2012). *Human Anatomy and Physiology, 9th Edn.,* London: Pearson.

Maton. K, (2013). 'Making Semantic Waves: A Key to Cumulative Knowledge-building'. *Linguistics and Education, 24,* [1], 8–22.

Markus. KA and Borsboom. D, (2013). *Frontiers of Test Validity Theory: Measurement, Causation and Meaning.* London. Routledge.

Matsunaga. M, (2011). How to Factor-Analyse your Data Right: Do's, Don'ts and How-tos. *International Journal of Psychological Research, 3, [1],* 97-110.

Mayhew. F, Swartz. N, and Taylor. JA, (2014). Implementing a Multi-method Competency Model: Experiences of the MPA Program at James Madison University. *Journal of Public Affairs Education, 20, [3]*, 321-334.

McAuliffe. G, and Lovell. C, (2006). The Influence of Counselor Epistemology on the Helping Interview: A Qualitative Study. *Journal of Counseling and Development, 84*, 308–317.

McCance. KL, and Huether. SE, (2006). *Pathophysiology: The Biological Basis for Disease in Adults and Children [5th Edn].* St Louis: Elsevier Mosby.

McClarty. KL, and Gaertner. MN, (2014). *Measuring Mastery. Best Practice for Assessment in Competency-Based Education.* Center for College and Career Success. Pearson

McGaghie. WC, (2014). A Critical Review of Mastery of Simulation-based Mastery Learning with Translational Outcomes. *Medical Education, 48,* 375-385.

McGaghie. WC, (2015). Mastery Learning: Is it time for Medical Education to Join the 21st Century.

Academic Medicine, 90, [11], 1438-1441.

McGaghie. WC, and Kristopaitis. T, (2015). *Deliberate Practice and Mastery Learning: Origins of Expert Medical Performance,* In Cleland. J, and Durning. SJ, (eds) Researching Medical Education, p. 219–30. New York: John Wiley and Sons.

McGaughey. J, O'Halloran. P, Porter. S, Trinder. J, and Blackwood. B, (2017). Early Warning Systems and Rapid Response to the Deteriorating Patient in Hospital: A Realist Evaluation. *Journal of Advanced Nursing, 73*, 3119-3132.

McHugh. ML, (2012). Interrater Reliability: The Kappa Statistic. *Biochem Med (Zagreb), 22, [3]*, 276-282. Free PMC Article available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052/ accessed October 2019.

McKay. FH, and Dunn. M, (2015). Student Reflections in a First-Year Public Health and Health Promotion Unit. *Reflective Practice, 16, [2]*, 242-253.

McKenna. HP, and Slevin. OD, (2008). Knowing and Knowledge: Importance for Nursing. In M. H. OD, *Nursing Knowledge, Theories and Practice* (pp. 63-97). Chichester: John Wiley and Sons Ltd.

McKinley. RK, Strand. J, Ward. L, Gray. T, Alun-Jones. T, and Miller. H, (2008). Checklists for Assessment and Certification of Clinical Procedural Skills Omit Essential Competencies: A Systematic Review. *Medical Education, 42, 338–349.*

McNeill. H, Brown JM, and Shaw. NJ, (2010). First Year Specialist Trainees' Engagement with Reflective Practice in the e-portfolio. *Advances in Health Science Education, 15: 547-558.*

Medicine and Healthcare Products Regulatory Agency (MHRA, 2005). *Consultation on Proposals to Introduce Independent Prescribing by Pharmacists.* London: The Cabinet Office.

Medicines for Human Use Act (MHUA, 2001). Available from http://www.legislation.gov.uk/uksi/2004/1031/contents/made Accessed November 2015.

Medicinal Products: Prescription for Nurses Act (1992), Chapter 28. London. Her Majesty's Stationery Office *HMSO.*

Megginson. D, and Whitaker. V, (2017). C*ontinuing Professional Development (2nd Edn.)* London: The Chartered Institute of Personnel and Development. Kogan Page Publishers.

Meleis. AI, (2012). *Theoretical Nursing: Development and Progress (5th Edn.)* New York, Lippincott Williams and Wilkins.

Merrill. MD, (1983). *Component Display Theory. In Reigluth. CM, [Ed.], Instructional Design Theories and models: An Overview of their Current States.* Hillsdale, New Jersey. Erlbaum Publication.

Merrill. MD, (1983). *Component Display Theory. In Reigluth. CM, [Ed.], Instructional Design Theories and models: An Overview of their Current States.* Hillsdale, New Jersey. Erlbaum Publication.

Merrill. MD, (1987). The New Component Design Theory: Instructional Design for Courseware Authorising. *Instructional Science, 16*, 19-34.

Mirriam. SB, and Tisdell. EJ, (2016). *Qualitative Research (4th Edn). A Guide to Design and Implementation. San Francisco,* Josey-Bass.

Messick. S, (1980). Test Validity and the Ethics of Assessment. *American Psychologist, 35 [11]*, 1012-1027.

Messick. S, (1989). Validity. In R. L. [Ed.], *Educational Measurement* (pp. 13-103). Washington DC: American Council on Education and National Council on Measurement in Education.

Messick. S, (1995). The validity of Psychological Assessment. Validation of Inferences from Persons Response and Performance as Scientific Inquiry into Score Meaning. *American Psychologist, 50 [9]*, 741-749.

Meyer. JP, (2014). *Applied Measurement with J-Metrick.* New York: Routledge- Taylor and Francis Group.

Mezirow. J, (1981). A Critical Theory of Adult Learning and Education. *Adult Education, 32 [1]*, 3-24.

Mezirow. J, and Associcates. (1990). *Fostering Critical Reflection in Adulthood. A Guide to Transformative Learning.* San Fransisco: Josey- Bass.

Mezirow. J, (1990a). How Critical Reflection Triggers Transformative Learning. In E. In Mezirow et al., *Fostering Critical Reflection in Adulthood. A guide to Transformative Emancipatory Learning.* (pp. 1-20). San Fransisco: Josey-Bass.

Mezirow. J, (2000). L*earning as Transformation: Critical Perspectives on a Theory in Progress*. [1st Edn]. The Jossey-Bass Higher and Adult Education Series. Jossey-Bass, San Francisco.

Miles. A, (2011). *The Reflective Coach. In Stafford. I, (2011) Coaching Children in Sport, Chapter 8* P109-120. London. Routledge.

Miller. GE, (1990). The Assessment of Clinical Skills, Competence, Performance. *Academic Medicine, 65 [9]*, S63-S67.

Miller. MD, Linn. RL, and Gronlund. NE, (2013). *Measurement in Education and Teaching*. New Jersey. Pearson Education, Inc.

Miller. NP, (2018). *Teaching Law: A Framework for Instructional Mastery* (2nd.Edn.), Missouri. Crown Management LLC Publication.

Moon. J, (2007). Getting the Measure of Reflection: Considering Matters of Definition and Depth *Journal of Radiotherapy in Practice, 6*, 191-200.

Mooney. BY, and Nowacki. M, (2011). *Understanding Teaching and Learning, St Andrews Studies in Philosophy and Public Affairs.* Exeter: Imprint Academic.

Moore. KL, Dalley. AF, and Agur. AM, (2009). *Clinically Oriented Anatomy (6th Edn)*. London. Lippincott, Williams and Wilkins.

Moore. DT, and Hoffman. RR, (2011). Sensemaking: A Transformative Paradigm. *American Intelligence Journal,* 29, 26–36.

Morgan. C, Dunn. L, Parry. S, and O'Reilly. M, (2004). *The Student Assessment Handbook: New Directions in Traditional and Online Assessment.* London: Routledge-Falmer.

Morton. PG, Fontaine. DK, Hudak. CM, and Gallo. BM, (2005). *Critical Care Nursing: A Holistic Approach [5th Edn.].* Philadelphia: Lippincott, Williams and Wilkins.

Mucklow. J, Bollington. L, and Maxwell. S, (2011). Assessing Prescribing Competence. *British Journal of Clinical Pharmacology, 74, [4]*, 632–639.

Mulcahy. CM, Mulcahy. DE, and Mulcahy. DG, (2015). *Pedagogy, Practice and Purpose in Education*, London. Routledge, Taylor Francis Group.

Mulholland. PJ, (2014) Pharmacists as Non-medical Prescribers: What Role can they Play? The Experience in a Neonatal Intensive Care Unit. *European Journal of Hospital Pharmacy*, *21, [6]*, 335-338.

Munoz. LQ, Byrne. C, Pugsley. J, and Austin. Z, (2005). Reliability, Validity, and Generalizability of an Objective Structured Clinical Examination (OSCE) for Assessment of Entry-to-practice in Pharmacy. *Pharmacy Education, 5, [1],* 33–43.

Murdoch-Eaton, D., and Whittle. S, (2012). Generic Skills in Medical Education: Developing the Tools for Successful Lifelong Learning. *Medical education, 46, [1],* 120–28.

Murphy. KR, and Davidshofer. CO, (2001). *Psychological Testing: Principles and Applications (5th Edn).* New Jersey: Prentice Hall International, Inc.

Naber. J, and Wyatt. TH, (2014). The Effect of Reflective Writing Interventions on the Critical Thinking Skills and Dispositions of Baccalaureate Nursing Students. *Nurse Education Today, 34*, 67-72.

Naidoo. R and Williams. J, (2015). The Neoliberal Regime in English Higher Education: Charters, Consumers and the Erosion of the Public Good. *Critical Studies in Education, 56, [2]*, 208-223.

National Health Service (NHS, 2019). Long Term Plan: Interim People Plan (IPP).London, NHS.

National Health Service (NHS, 2012). Education and Training- next stage. Available from https://www.gov.uk/government/publications/nhs-future-forum-recommenedations Accessed June 2020.

National Patient Safety Agency (NPSA, 2007). *Safer Care for the Acutely Ill Patient; Learning from Serious Incidents.* London: NPSA.

National Patient Safety Agency, (NPSA, 2007a). *Safety in Doses: Medication Safety Incident in the NHS.* London: The Fourth Report from the Patient Safety Observatory.

Nash. C, and Sproule. J, (2011). Insights into Experiences: Reflections of an Expert and Novice coach. *International Journal of Sports Science and Coaching, 6, [1],* 149–161.

Naughton. C, Drennan. J, Hyde. A, Allen. D, O'Boyle. K, Felle. P, and Butler. M, (2013). An Evaluation of the Appropriateness and Safety of Nurse and Midwife Prescribing in Ireland. *Journal of Advanced Nursing,* 69, [7], 1478–1488.

Newble. D, (2004). Techniques for Measuring Clinical Competence: Objective Structured Clinical Examination. *Medical Education; 38, [2]*, 199-203.

Newby. D, (2011). Competence and Performance in Learning and Teaching: Theories and Practices. *Selected Papers on Theoretical and Applied Linguistics*, *19*, 15-32.

Newton. JM., Jolly. BC, Billett. S, Cross. WM, and Kelly. C, (2007). Engaging in Learning in Nursing Practice. Royal College of Nursing, UK: 1st International Nurse Education Conference, Brighton, UK, 5–7 July.

Ng. SL, Kinsella. EA, Friesen. F, and Hodges. B, (2015). Reclaiming a Theoretical Orientation to Reflection in Medical Education. *Medical Education and Review, 49,* 461-475.

Nguyen. QD, Fernandez. N, Karsenti. T, and Charlin. B, (2014). What is Reflection? A Conceptual Analysis of Major Definitions and a Proposal of a Five-component Model. *Medical Education, 48 [12],* 1176–1189.

Nilsson. P, (2008). Teaching for Understanding: The Complex Nature of Pedagogical Content Knowledge in Pre-service Education. *International Journal of Science Education*, *30, [10]*, 1281-1299.

Ning. HK, and Downing. K, (2011). The Interrelationship between Student Learning Experience and Study Behaviour. *Higher Education Research and Development, 30, [6],* 765-778,

Nitko, AJ, and Brookhart. SM, (2011). *Educational Assessment of Students (6th Edn).* New Jersey: Pearson.

Newman. FM, (1990). Higher-order Thinking in Teaching Social Studies: A Rationale for the Assessment of Classroom Thoughtfulness. *Journal of Curriculum Studies, 22, [1]*, 41-56.

Nolte. E, Fry. CV, Winpenny. E, and Brereton. L, (2011). *Use of Outcome Metrics to Measure Quality in Education and Training in Healthcare Professionals: A Scoping Review of International Experience.* London: Department of Health and RAND Europe.

Norcini. JJ, (1994). Research on Standards for Professional Licensure and Certification Examinations. *Evaluation and the Health Professionals, 17 [2],* 160-177.

Norcini. JJ, (2003). Setting Standards on Educational Tests. *Medical Education, 37*, 464–469.

Norrie. C, Hammond. J, D'Avray. L, Collington. V, and Fook. J, (2012). Doing It Differently? A Review of Literature on Teaching Reflective Practice across Health and Social Care Professions. Reflective Practice, 13, 565-578.

Nursing and Midwifery Council (NMC, 2006). *Standards of Proficiency for Nurse and Midwife Prescribers.* London: NMC.

Nursing and Midwifery Council (NMC, 2008). *The Code: Standards of Conduct, Performance and Ethics for Nurses and Midwives.* London: NMC.

Nursing and Midwifery Council (NMC, 2010). *Standards for Pre-registration Nursing Education.* London: NMC.

Nursing and Midwifery Council (NMC, 2012). *Advanced Nurse Practitioners: An RCN Guide to the Advanced Nurse Practitioner Role Competencies Programme.* London: RCN.

Nursing and Midwifery Council, (2014). *Standards for Competence for Registered Nurses.* London: NMC.

Nursing and Midwifery Council (NMC, 2015). *The Code.* London: NMC.

Nursing and Midwifery Council (NMC, 2016). *Revalidation: Employers Guide to Revalidation* London: NMC.

Nursing and Midwifery Council (NMC, 2018) *Realising Professionalism: Standards for Education and Training Part 3: Standard for Prescribing Programs.* London, NMC.

Nursing and Midwifery Council (NMC, 2018a) *Future Nurse: Standards of Proficiency for Registered Nurses.* www.nmc.org.uk Accessed 08/01/19.

Nursing and Midwifery Council (NMC, 2018b) *Enabling Professionalism: Nursing and Midwifery Practice.* London, NMC.

Nuttall. D, and Rutt-Howard. J, (2011). *Textbook of Non-Medical Prescribing.* Hoboken, NJ. Wiley-Blackwell.

Ockey. G, Koyama. D, Setoguchi. E, and Sun. A, (2015). The Extent to which TOEFL iBT Speaking Scores are Associated with Performance on Oral Language Tasks and Oral Ability Components for Japanese University Students. *Language Testing*, *32* [1], 39–62.

O'Connor. P, Keogh. I, and Azuara-Blanco. A, (2016). *Teamwork and Communication*. In Flin. R, Youngson. GG, and Yule. S, (Eds). Enhancing Surgical Performance. A Primer in Non-technical Skills, p. 105–21. Florida, CRC Press, Taylor and Francis Group.

Oermann. MH, and Gaberson. KB, (2014). *Evaluation and Testing in Nurse Education (4th Edn.)* New York: Springer Publishing Co.

Oermann. MH, Saewert. KJ, Charasika. M, and Yarbrough. SS, (2009). Assessment and Grading Practices in Schools of Nursing: National Survey Findings Part I. *Nursing Education Perspectives, 30, [5],* 274–278.

Offredy. M, Kendall. S, and Goodman. C, (2008). The Use of Cognitive Continuum Theory and Patient Scenarios to Explore Nurse Prescribers' Pharmacological Knowledge and Decision-making. *International Journal of Nursing Studies, 45, [6]*, 855-68.

O'Keefe. P, and Linnenbrink-Garcia. L, (2014). The Role of Interest in Optimizing Performance and self-Regulation. *Journal of Experimental Social Psychology*, 53, 70-78.

Organisation for Economic Co-operation and Development (OECD, 2013). *Program for International Student Assessment* (PISA, 2012)*. Assessment and Analytical Framework: Mathematics, Reading, Science, Problem-solving and Financial Literacy*. Paris, France: OECD Publishing.

Organisation for Economic Co-operation and Development (OECD, 2014). Learning for Jobs. Paris, France: OECD Publishing.

Organisation for Economic Co-operation and Development (OECD, 2016*). Netherlands 2016: Foundations for the Future: Reviews of National Policies for Education*. Paris, France: OECD Publishing.

Organisation for Economic Co-operation and Development (OECD, 2017). P*rogram for International Student Assessment (PISA, 2015). Proficiency Scale Construction: Technical Report* Paris, France: OECD Publishing.

Organisation for Economic Co-operation and Development (OECD, 2019). *Program for International Student Assessment (PISA, 2018). Chapter 15: Assessment and Analytical Framework*. Paris, France: OECD Publishing.

Organisation for Economic Co-operation and Development (OECD, 2019a) O*ECD Future of Education and Skills 2030: OECD Learning Compass.*

Ogle, D.M. (1986). K-W-L: A Teaching Model That Develops Active Reading of Expository

Text. *Reading Teacher,* 39, 564-570.

Ogle. DM, and Muth. KD, (1989). The Know, Want to Know, Learn Strategy. Children's Comprehension of Text: *Research into Practice*. 205-223.

Orell, J. (2006). Feedback on Learning achievement: Rhetoric and Reality. *Teaching in Higher Education, 11* (4), 441-456.

Osbourne. JW, (2008). *Best Practices in Quantitative Research*. Thousand Oaks. Sage Publications, Inc.

Oxford Handbook of Adult Nursing (2009). Oxford. Oxford University Press.

Paans. W, Sermeus. W, Nieweg. RM, Krijnen. WP, and van der Schans. CP, (2010). Determinants of the Accuracy of Nursing Diagnoses: Influence of Ready Knowledge, Knowledge Sources, Disposition Toward Critical Thinking, and Reasoning Skills. *Journal of Professional Nursing, 26*, 232–241.

Padden. ML, (2013). A Pilot Study to Determine the Validity and Reliability of the Level of Reflection-on-Action Assessment. *Journal of Nursing Education, 52, [7],* 410-415.

Paolo. C, Golia. E, and Yeh. ETH, (2012). Role of C-Reactive Protein in Acute Myocardial Infarction and Stroke: Possible Therapeutic Approaches. *Current Pharmaceutical Biotechnology, 13, 4-16.*

Patricio. M, and Carneiro. AV, (2012). Systematic Reviews of Evidence in Medical Education and Clinical Medicine: is the Nature of Evidence Similar? *Medical Teacher, 34, [6],* 474-482.

Parahoo. K, (2006). *Nursing Research: Principles, Process and Issues, [2nd Edn.]*. Basingstoke: Palgrave.

Payne. E, (2010). Principles of Safe Medicines Administration. In P. L. Philip. J, *Nurse Prescribing, Chapter 14* (pp. 282-304). Oxford: Wiley-Blackwell.

Peach. L, (1999). *Fitness for Practice - The Report for the United Kingdom Central Council (UKCC) for Nursing, Midwifery and Health Visiting Commission.* London: UKCC.

Pell. G, Fuller. R, Homer. M, and Roberts. T, (2010). How to Measure the Quality of the OSCE: A Review of Metrics. *Medical Teacher, 32, [10]*, 802-811.

Pellegrino. JW, Chudowsky. N, and Glaser. R, (2001). *Knowing What Students Know: The Science and Design of Educational Assessment*, Washington, DC: National Academies Press.

Petticrew. M, and Roberts. H, (2006). *Systematic Reviews in The Social Sciences – A Practical Guide*. Oxford: Blackwell Publishing.

Plack. MM, Driscoll. M, Marquez. M, Cuppernull. L, Maring. J, and Greenburg. L, (2007). Assessing

Reflective Writing on a Pediatric Clerkship by Using a Modified Bloom's Taxonomy. *Ambulatory Pediatrics,* 7:285–291.

Ployhart. RE, (2012). The Content Validity of Cognitively Oriented Tests. Commentary on Schmidt (2012). *International Journal of Selection and Assessment, 20,* 19–23.

Polanyi. M, (1966). *The Tacit Dimension.* New York: Doubleday and Company. Inc.

Polit. DF, and Tatano-Beck. C, (2009). *Nursing Research: Generating and Assessing the Evidence for Nursing Practice.* London: Lippincott Williams and Wilkins.

Popham. WJ. (2012). *The Role of Rubrics in Testing and Teaching.* New Jersey: Pearson.

Posner. MI, and McLeod. P, (1982). Information Processing Models of Elementary Operations. *Annual Review of Psychology, 33*, 477-514.

Price. M, Carroll. J, O'Donovan. B, and Rust. C, (2011)."If I Was Going There, I Wouldn't Start from Here: A Critical Commentary on Current Assessment Practice."*Assessment and Evaluation in Higher Education, 36 [4],* 479–492.

Price. M, Rust. C, O'Donovan. B, Handley. K, and Bryant. R, (2012). *Assessment Literacy: The Foundation for Improving Student Learning.* Oxford: Oxford Centre for Staff and Learning Development.

Price. B, and Harrington. A, (2016). *Critical Thinking and Writing for Nursing Students (3rd Edn).* Sage Publications, Thousand Oaks, CA.

Price. B, (2017). Improving Nurses' level of Reflection. *Nursing Standard, 32, [1],* 52-63.

Prescription for Nurses Act. (1992). *Medicinal Products: Prescription for Nurses Act.* London: Majesty's Stationery Office HMSO.

Prosser. M, and Trigwell. K, (1999). *Understanding Learning and Teaching: The Experience of Higher Education.* Buckingham: Society into Research in Education. Open University Press.

Pugh. D, Halman. S Desjardin. I, Humphrey-Murto. S, and Wood. TJ, (2016). Done or Almost Done? Improving OSCE Checklists to Better Capture Performance in Progress Tests. *Teaching and Learning in Medicine, 28, [4]*, 406-414.

Pulcini. J, Jelic. M, Gul. R, and Loke. AY, (2010). An international Survey on Advanced Nurse Practice Education, Practice and Regulation. *Journal of Nursing Scholarship, 42, [1],* 31-39.

Putnam. H, (1975). Open Questions about Test Score Meaning. Cited in *Markus and Borsboom (2013) Frontiers of Test Validity: Measurement, Causation and Meaning*. p.247-272. London. Routledge, Taylor-Francis Publication.

Quality Assurance Agency (QAA, 2010). *Master's Degree Qualification*. Gloucester. The Quality Assurance Agency for Higher Education.

Quality Assurance Agency (QAA, 2012). *UK Quality Code for Higher Education. Part B: Assuring and Enhancing Academic Quality*. Chapter B3: Learning and Teaching. Technical Report.

Quality Assurance Agency (QAA, 2014). *UK Quality Code for Higher Education. Part A: Setting and Maintaining Academic Standards. The Frameworks for Higher Education Qualifications of UK Degree-Awarding Bodies.* Gloucester. The Quality Assurance Agency for Higher Education.

Quinn. FM, and Hughes. SJ, (2007). Quinn's Principles and practices of Nurse Education [5th Edn]. London. Cengage Learning.

Rafferty. AM (2002). *The Politics of Nursing Knowledge.* London Routledge.

Rafferty. AM, Clarke. SP, Coles J, Ball. J, James. P, McKee. M, and Aiken. LH, (2007). Outcomes of Variation in Hospital Nurse Staffing in English Hospitals: Cross-sectional Analysis of Survey Data and Discharge Records. *International Journal of Nursing Studies, 44, [2],* 175–182.

Randall. J, and Engelhard. G, (2010). Examining the Grading Practices of Teachers. *Teaching and Teacher Education, 26*, 1372–1380.

Rauhvargers. A Deane. C, and Pauwels. W, (2009). Bologna Process Stocktaking Report http://ehea.info/media.ehea.info/file/20090326-27-Prague/ 61/ 0/ BFUG CZ_16_5.1a_stocktaking_final_594610.pdf. Accessed 09/04/2020.

Raven. S, (2001). *The Conceptualisation of Competence; In Competence in the Learning Society, Chapter 17: 253-274.* New York: Peter Lang Publication.

Raykov. T, and Marcoulides. GA, (2016). On the Relationship Between Classical Test Theory and Item Response Theory: From One to the Other and Back. *Educational and Psychological Measurement, 76, [2]*, 325–338.

Reed. PG, and Shearer. NBC, (2012) Perspective on Nursing Theory (6th Edn.) Philadelphia. Lippincott, Williams and Wilkins.

Redfern. S, Norman. I, Calman. L, Watson. R, and Murrells. T, (2010). Assessing Competence to Practise in Nursing: A Review of the Literature. *Research Papers in Education, 17, [1]*, 51-77.

Reise. SP, (2012). The Rediscovery of Bifactor Measurement Models. *Multivariate Behavioral Research*, *47, [5]*, 667-696.

Reise. SP, and Rodriguez. A, (2016). Item Response Theory and the Measurement of Psychiatric Constructs: Some Empirical and Conceptual Issues and Challenges. *Psychological Medicine, 46 [10]*, 2025-2039.

Richards. JC, (2017). Teaching English through English: Proficiency, Pedagogy and Performance. *RELC Journal* 48, [1], 7-30.

Richardson. G, and Maltby. H, (1995). Reflection-on-Practice: Enhancing Student Learning. *Journal of Advanced Nursing, 22, [2]*, 235-242.

Rindskopf. D, (2010). Reliability: Measurement. *International Encyclopedia of the Social and Behavioral Sciences, 20*, 248-252.

Risjord. M, (2010). Nursing Knowledge: Science, Practice and Philosophy. London A John Wiley *and* Sons, Ltd., Publication.

Rittle-Johnson. B, and Alibali. MW, (1991). Conceptual and Procedural Knowledge in Mathematics. Does one lead the Other? *Journal of Educational Psychology, 91, [1]*, 175-189.

Rizopoloulos. D, (2006). An R package for Latent Trait Variable Modelling an Item Response Theory Analysis. *Journal of Statistical Software, 17 [5]*, 2-25.

Robbin. M, and Asselin. M, (2015). Reflection as an Educational Strategy in Nursing Professional Development: An Integrative Review. *Journal for Nurses in Professional Development 31*, [2], 61-72.

Rogers. RR, (2001). Reflection in Higher Education: A Concept Analysis. *Innovative Higher Education, 26, [1]*, 37–57.

Roget's (2019) *International Thesaurus (8th Edn,) Revised and Updated.* New York, Harper Collins Publication.

Rolfe. G, (1996). *Closing the Theory-Practice Gap: A New Paradigm for Nursing.* Oxford: Butterworth-Heinemann Ltd.

Rolfe. G, and Fulbrook. P, (2001). *Advanced Nursing Practice*. Oxford. Butterworth Heinemann.

Rolfe. G, (2014). Understanding Advanced Nursing Practice. *Nursing Times, 110, [27]*, 20-23.

Rolfe. G, (2014a). Rethinking Reflective Practice: What Would Dewey have Done? *Nurse Education Today, 34,* 1179–1183.

Román-Cereto. M, García-Mayor. S, Kaknani-Uttumchandani. S, García-Gámez. M, León-Campos. A, Fernández-Ordóñez. E, Ruiz-Garcia. ML, Marti-Garcia.C, Lopez-Leiva. I, Lasater.K, and Morlaes-Asencio. JM, (2018). Cultural Adaptation and Validation of the Lasater Clinical Judgment Rubric in Nursing Students in Spain. N*urse Edu*cu*cation Today, 64, 71–8*.

Rosenblatt. Z, (2017). Personal Accountability in Education: Measure Development and Validation. *Journal of Educational Administrati*on, 55, [1], 18-32.

Ross. JD, and Kettles. AM, (2012). Mental Health Nurse Independent Prescribing: What are Nurse Prescribers' Views of the Barriers to Implementation? *Journal of psychiatric and mental health nursing*, *19, [10],* 916-932.

Rossi. PH, Freeman. HE, and Lipsey. MW, (2018). *Evaluation, a Systematic Approach (8th Edn).* London. Sage Publication.

Roumbanis. L, (2017). Academic Judgments Under Uncertainty: A Study of Collective Anchoring Effects in Swedish Research Council Panel Groups. *Social Studies of Science, 47, [1], 95–116.*

Royal. KD, (2017). Four Tenets of Modern Validity Theory for Medical Education Assessment and Evaluation. *Advances in Medical Education and Practice*, *8*, 567- 570.

Rusch. T, Mair. P, and Hatzinger. R, (2013). Psychometrics with R: A review of CRAN Packages for Item Response Theory. Discussion Paper Series. *Cente for Empirical Research. Http://wu.ac.at/methods,* Accessed November 2014: p1-29.

Rushforth. HE, (2007). Objective Structured Clinical Examination (OSCE): Review of the Literature and Implications for Nursing Education. *Nurse Education Today, 27*, 481–490.

Russell. CL, (2005). Evaluating Quantitative Reports. *Nephrology Nursing Journal, 32 [1], 61-64.*

Ryan. S, Stevenson. K and Hassell. AB. (2007). Assessment of Clinical Nurse Specialists in Rheumatology using an OSCE. *Musculoskeletal Care, 5, [3],* 119–129.

Ryan. M, (2013). Theorizing a Model for Teaching and Assessing Reflective Learning in Higher Education. Higher Education Research and Development, 32, [2], 244-257.

Ryle. G, (1949). *The Concept of Mind: 60th Anniversary 2009 Edition.* Abingdon, Oxon: Hutchinson.

Sackett. PR, Borneman. MJ, and Connelly. BS, (2008). High-Stakes Testing in Higher Education and Employment: Appraising the Evidence for Validity and Fairness. *American Psychologist, 63 [4],* 215-227.

Sackett. PR, (2012). Cognitive Tests, Constructs, and Content Validity. A commentary on Schmidt (2012). *International Journal of- Selection and Assessment, 20*, 24–27.

Sadler. DR, (2007). Perils in the Meticulous Specification of Goals and Assessment Criteria. *Assessment in Education: Principles, Policy and Practice, 14*, 159–179.

Sadler. DR, (2009). Grade Integrity and the Representation of Academic Achievement. *Studies in Higher Education, 34, [7]*, 807-826.

Sadler. DR, (2009a). Indeterminacy in the use of Preset Criteria for Assessment and Grading. *Assessment and Evaluation in Higher Education, 34, [2]*, 159–179.

Salkind. NJ, (2017). Statistics for People Who (Think They) Hate Statistics (6th Edn). London. Sage Publication.

Salzman. DH, McGaghie. WC, Caprio. TW, Hufmeyer.KK, ISsa. N, Cohen.ER and Wayne. DB, (2011).

A Mastery Learning Capstone Course to Teach and Assess Components of Three Entrustable Professional Activities to Graduating Medical Students. *Teaching and Learning in Medicine, 31, [2]*, 186-194.

Sandelowski. M, (2000). Focus on research methods: Whatever happened to qualitative description? *Research in Nursing and Health*, *23, [4],* 334-340.

Sapsford. R, and Jupp. V, (2006). *Data Collection and Data Analysis [2nd Edn.]* London: Sage Publication.

Sargeant. J, Wong. BM, and Campbell. CM, (2018). CPD of the Future: A Partnership between Quality Improvement and Competency-based Education. *Medical Education*, *52, [1],* 125-135.

Scalese. JR, Obeso. VT, and Issenberg. SB, (2008). Simulation Technology for Skills Training and Competency Assessment in Medical Education. *Journal of General Internal Medicine. Supp 1*, 49-46.

Scott. D, (2008). *Critical Essays on Major Curriculum Theorists.* Oxon, Routledge Publication.

Sechrist. KR, and Berlin. LE, (2006). Psychometric Analysis of the Perceived Value of Certification Tool. J*ournal of Professional Nursing*, 22, 248-252.

Seaman. M, (2011). *Blooms Taxonomy: Its Evolution, Revision, and use in the Field of Education. In Curriculum and Teaching Dialogue, Chapter 3, p 29-4*5. North Carolina, USA. Information Age Publishing.

Scottish Government Health Service (SGHS, 2008). *Supporting the Development of Advanced Nurse Practice- A toolkit Approach*. Edinburgh: SGHS.

Schachter. M, (2009). The Epidemiology of Medication Errors: How Many, How Serious? *British Journal of Clinical Pharmacology, 67*, 621–623.

Schaub-De Jong. M.A, Schönrock-Adema. J. Dekker. H, Verkerk. M, and Cohen-Schotanus. J, (2011). Development of a Student Rating Scale to Evaluate Teachers' Competencies for Facilitating Reflective Learning. *Medical Education, 45 [2]*, 155-165.

Schellert. DL, (1982). The Significance of Knowledge: A Synthesis of Research Related to Schema Theory. In a. W. In Otto. W, *Reading Expository Prose* (p. 65-79). New York: Academic.

Schmidt. FL, (2012). Cognitive Tests Used in Selection Can Have Cognitive Tests Used in Selection Can Have A broader research review and implications for practice. *International Journal of Selection and Assessment, 20 [1]*, 1-13.

Schon. D, (1983). *The Reflective Practitioner.* London: Temple Smith.

Schon. D, (1990). *The Reflective Practitioner: How Professionals Think in Action.* London: Temple Smith.

Schraw. G, and Robinson. DR, [Eds] (2011). *Assessment of Higher Order Thinking Skills.* Charlotte. North Carolina. Information Age Publishing.

Schubert. M, Clarke. SP, Aiken. LH, and de Geest. S, (2012) Associations Between Rationing of Nursing Care and Inpatient Mortality in Swiss Hospitals. *International Journal for Quality in Health Care, 24, [3],* 230–238.

Schultz. KS, Whitney. DJ, and Zickar. MJ, (2005). *Measurement Theory in Action.* New York: Sage Publications.

Schuwrith LWT, and van der Vleuten. CPM, (2014). *How to Design a useful Assessment. In Understanding Medical Education: Evidence, Theory and Practice, 195- 207.* Hoboken, Wiley-Blackwell.

Schuwrith. LWT, and van der Vleuten. CPM, (2011). Pragmatic Assessment: From Assessment of Learning to Assessment for Learning. *Medical Teacher, 33, [6],* 478-485.

Scobie. SD, Lawson. M, and Cavell. G, (2003). Meeting the Challenge of Prescribing and Administering Medicines Safely. Structured Teaching and Assessment for Final Year Medical Students. *Med Educ, 37*, 434 –437.

Scully. NJ, (2011). The Theory-Practice Gap and Skill Acquisition: An Issue for Nursing Education. *Collegian 18, [2] 93–98.*

Secolsky. C, and Denison. DB, [Eds] (2012). *Handbook of Measurement and Evaluation in Higher Education.* Oxon, Routledge Publication.

Seidel. T, Blomberg. G, and Renkl. A, (2013). Instructional Strategies for Using Video in Teacher Education. *Teaching and Teacher Education, 34,* 56–65.

Selim. A Ramadan. FH, El-Gueneidy. MM, and Gaafer.MM, (2012). Using Objective Structured Clinical Examination (OSCE) in Undergraduate Psychiatric Nursing Education: Is it Reliable and Valid. *Nurse Education Today, 32*, 283–288.

Shavelson. RJ, (2010). On the measurement of competency. *Empirical Research in Vocational Education and Training, 2,* 41–63.

Shepard. LA, (1993). Evaluating Test Validity. *Review of Research in Education, 19*, 405-450.

Shepard. L A, (1997). *Measuring Achievement: What Does it Mean to Test for Robust Understanding?* Princeton, NJ: Educational Testing Service.

Shepard. LA, (2000), The Tole of Assessment in a Learning Culture. *Educational Researcher, 29, [7]*, 4-14

Shepard. LA, (2016). Evaluating Test Validity: Reprise and Progress. *Assessment in Education: Principles, Policy and Practice, 23, [2]*, 268-280.

Shepard. L A, Penuel. W R, and Davidson. KL, (2017). Design Principles for New Systems of Assessment. *Phi Delta Kappa, 98, [6],* 47– 52.

Sibbald. B, Laurant. GM, and Reeves. D, (2006). Advanced Nurse Roles in UK Primary Care. *The Medical Journal of Australia, 85, [1]*, 10-12.

Silverman. D, (2014). *Interpreting Qualitative Data. Methods for Analyzing Talk, Text and Interaction.* London, SAGE Publications.

Silverston. P, (2012). Using a Model of Illness to Aid Symptom-based Learning in Primary Care. *Education for Primary Care, 23,* 443– 445.

Silverston. P, (2013). Consultation Skills Teaching in Primary Care. *Education for Primary Care, 24,* 206-218.

Sim. J, and Wright. CC, (2005). The Kappa Statistic in Reliability Studies: Use, Interpretation and Sample Size. *Physical Therapy, 85, [3]*, 257-268.

Sipress. JM, and Voelker. DJ, (2011). The End of the History Survey Course: The Rise and Fall of the Coverage Model. *Journal of American History, 97, [4]*, 1050-1066.

Sireci. SG, (2007). On Validity Theory and Test Validation. Educational Researcher, 36, 477-481.

Smith. E, (2008). *Using Secondary Data in Education and Social Research.* Berkshire: Open University Press.

Smith. GB, (2002). ALERT: A Multiprofessional Training Course in the Care of the Acutely Ill Adult Patient. *Resuscitation, 52*, 281-286.

Society for Industrial and Organizational Psychology (SIOP), (2003). *Principles for the Validation and Use of Personnel and Use of Personnel Selection Procedures.* Bowling Green, OH: SIOP.

Soleimani. H, and Khieri. S, (2016). An Evaluation of TEFL Postgraduates' Testing Classroom Activities and Assignments Based on Bloom's Revised Taxonomy. *Theory and Practice in Language Studies, 6, [4]*, 861-869.

Stenner. K, Carey. N, and Courtenay. M (2010). How Nurse Prescribing Influences the Role of Nursing. *Nurse Prescribing, 8, [1],* 29-34.

Stone. CA and Ting. X, (2012). Using IRT Trait Estimates versus Summated Scores in Predicting Outcomes. *Educational and Psychological Measurement*, *72, [3],* 453-468.

Stone. CA, and Tang. Y, (2013). Comparing Propensity Score Methods in Balancing Covariates and Recovering Impact in Small Sample Educational Program Evaluations. *Practical Assessment, Research, and Evaluation*, *18 [13],* 1-12.

Stott. NC, and Davis. RH, (1979). The Exceptional Potential in each Primary Care Consultation. *Journal of the Royal College of General Practitioners, 29*, 201-205.

Stough. T, Ceulemans. K, Lambrechts. W, and Cappuyns. V, (2018). Assessing Sustainability in Higher Education Curricula: A Critical Reflection on Validity Issues. *Journal of Cleaner Production*, *172*, 4456-4466.

Stubbings. L, and Scott. JM, (2004). NHS Workforce Issues: Implications for Future Practice. *Journal of Health Organisation Management, 18,*149-194.

Stubbings. L, Chaboyer. W, and McMurray. A, (2012). Nurses' use of Situation Awareness in Decision-making: An Integrative Review. *Journal of Advanced Nursing*, *68, [7]*, 1443-1453.

Striener. DL, (2010). Measure for Measure: New Developments in Measurement and Item Response Theory. *The Canadian Journal of Psychiatry, 55, [3],* 180-186.

Strickland. HP, (2013). Comparing Lasater's Clinical Judgment Rubric Scores Across Faculty, Self-Assessment and Outcome Scores. Dissertations and Theses - Gradworks.

Strydom. P, (2011). C*ontemporary Critical Theory and Methodology.* London, Routledge. Taylor and Francis Group.

Sullivan. GM, (2011). A Primer on the Validity of Assessment Instruments. *Journal of Graduate Medical Education*, *3, [2],* 119-120.

Sulistyawati. K, Wickens. CD, and Chui. YP, (2011). Prediction in Situation Awareness: Confidence Bias and Underlying Cognitive Abilities. International Journal of Aviation Psychology, 21, 153–174.

Sun. Y, and Cheng. L, (2013). Teachers' Grading Practices: Meaning and Values Assigned. *Assessment in Education, 21*, 326–343.

Surscock. A, (2012). *Quality Assurance and the European Transformational Agenda. In: Curaj A., Scott P., Vlasceanu L., Wilson L. [Eds]. European Higher Education at the Crossroad*s: 247-265. Dordrecht, Springer Publication.

Swanwick. T, [Ed], (2010). *Understanding Medical Education, Theory and Practice.* Hoboken: Wiley-Blackwell.

Tait. D, (2010). Nursing Recognition and Response to Signs of Clinical Deterioration. *Nursing Management, 17, [*6*],* 31-35.

Tanner. CA, (2006). Thinking like a Nurse: A Research-based Model of Clinical Judgment in Nursing. *The Journal of Nursing Education, 45*, 204–211.

Tavakol. M, and Dennick. R, (2011). Post-examination Interpretation of Objective Test Data: Monitoring and Improving the Quality of High-stakes Examinations: AMEE Guide No. 66. Medical Teachacher, 34, e161-75.

Tavakol. M, and Dennick. R, (2017). The Foundations of Measurement and Assessment in Medical Education. *Medical teacher*, *39, [10],* 1010-1015.

Tavakol. M, and Pinner. G, (2018). Enhancing Objective Structured Clinical Examinations through Visualisation of Checklist Scores and Global Rating Scale. *International Journal of Medical Education, 9*, 132-136.

Taylor. EW, and Cranton. P, (2012). *The Handbook of Transformative Learning: Theory, Research, and Practice*. San Francisco, CA: Jossey-Bass.

Teddlie. C, and Tashakkori. A, (2009). *Foundations of Mixed Method Research: Integrating Quantitative and Qualitative Approaches in the Social and Behavioural Sciences.* London: Sage Publications.

tenCate. O, and Scheele. F, (2007). Competency-Based Postgraduate Training: Can We Bridge the Gap between Theory and Clinical Practice? *Academic Medicine, 82 [6]*, 542–547.

Terwilliger. J (1997) Semantics, Psychometrics and Assessment Reform. A Close Look at Authentic Assessment. *Educational Researcher, 26, [8],* 24-27.

The Medicines Act (1968). Her Majesty's Government of the United Kingdom. www.Legislation.gov.uk. Accessed November 2015.

The British National Formulary (BNF, 2017). *BNF 73, March-September*. London. BMJ Group and the Royal Pharmaceutical Company.

The Crown Report, (1999). *Review of Prescribing Supply and Administration, Final Report.* London. Department of Health.

The Kings Fund (2015). *Workforce Planning in the National Health Service [NHS].* London: The Kings Fund.

The Kings Fund (2017). *Making the Case for Quality Improvement: Lessons for the NHS Boards and Leaders*. London: The Kings Fund.

The Royal College of Nursing (RCN, 2007). *A Joint Position Statement on Continuing Professional Development for Health and Social Care Practitioners.* London, RCN.

The Royal College of Nursing (RCN, 2010). *Integrated Core Career and Competency Framework for Registered Nurses.* London, RCN.

The Royal College of Nursing (RCN, 2012). *Advanced Nurse Practitioners: An RCN Guide to the Advanced Nurse Practitioner Role Competencies and Programme Accreditation.* London, RCN.

The Royal College of Nursing (RCN, 2016). *Frontline Nurse Leadership: An international*

Perspective. Case Studies from Australia, New Zealand, and the United States. Policy Briefing 02/16. RCN Policy and International Department. London, RCN.

The Royal College of Nursing (RCN, 2018). *Advanced Level Nursing Practice: Introduction.RCN Standards for Advanced Level Nursing Practice, Advanced Nurse Practitioners, RCN Accreditation and RCN Credentialing.* London, RCN.

The Royal College of Nursing (RCN, 2018a). Realising Professionalism: Standards for Education and Training. Part 3: Standards for Prescribing Programmes. London, RCN.

The Oxford Dictionary of English (2010) Oxford. Oxford University Press.

The Royal Pharmaceutical Society (RPS, 2016). *A Competency Framework for all Prescribers.* The Royal Pharmaceutical Society. Great Britain.

The World Health Organisation (WHO, 2008). *Where are the Patients in Decision Making About their Own Care?* Estonia. Health Systems and Policy Analysis.

The World Health Organisation (WHO, 2017). *Patient Safety: Making Healthcare Safer.* London: The WHO.

Thompson. P, and Smith. C, [Eds] (2010). Working life: Renewing Labour Process Analysis. London. Palgrave.

Thorpe. K, (2004). Reflective Learning Journals: From Concept to Practice. *Reflective Practice*, *5 [3]* 327-343.

Trigwell. K, Ashwin. P, and Millan. ES, (2013). Evoked Prior Learning Experience and Approach to Learning as Predictors of Academic Achievement. *British Journal of Educational Psychology 83*, 363–378.

Tognolini. J and Stanley. G, (2007). Standards-based Assessment: A Tool and means to the Development of Human Capital and Capacity Building in Education. *Australian Journal of Education, 51, [2]*, 129–145.

Torbin. G, and Bagley. C, (2004). Methodological Rigour within a Qualitative Framework. *Journal of Advanced Nursing, 48, [4],* 388-396.

Torrance. H (1995). *Evaluating Authentic Assessment.* Buckingham: Open University Press.

Toulmin. SE, (2008). *The Uses of Argument.* Cambridge. Cambridge University Press.

Traynor, M. (2013). *Nursing in Context: Policy, Politics, Profession*. Palgrave Macmillan International Higher Education.

Trochim. WMK, and Donnelley. JP, (2008). *The Research Methods Knowledge Base.* Mason, OH: Cengage Learning.

Tufford. L, and Newman. PA, (2012). Bracketing in Qualitative Research. *Qualitative social work*, *11, [1],* 80-96.

Unwin. A, and Fuller L, (2010). 'Knowledge Workers' as the New Apprentices: The Influence of Organisational Autonomy, Goals and Values on the Nurturing of Expertise. *Vocations and Learning 3,[3],* 203-222.

Uys. LR and Gwele. NS, (2001). *Curriculum Development in Nursing: Process and Innovations.* London and New York: Routledge: Taylor and Francis Group.

Vaismoradi. M, (2012). Nursing Education Curriculum for Improving Patient Safety. *Journal of Nursing Education and Practice,* 2, [1], 101-104.

van der Vleuten. CP, Schuwirth. LW, (2005). Assessing Professional Competence: From Methods to Programmes. *Medical Education, 39, [3],* 309–17.

van Woerkom. M, (2010). Critical Reflection as a Rationalistic Ideal. *Adult Education Quarterly, 60, [4]*, 339-356.

Veenman. MVJ, Wilhelm. P, and Beishuizen. JJ, (2004). The Relation between Intellectual and Metacognitive Skills at the Onset of Metacognitive Skill Development. *Instructional Science, 33*, 193–211.

Viera. AJ, and Garrett. JM, (2005). Understanding Interobserver Agreement. *Family Medicine*, 360-373.

Vinoviskis. M, (2015). *From a Nation at Risk to No Child Left Behind. National Education Goals and the Creation of Federal Education Policy.* London. Teachers College, Columbia University Press.

Vitale. P, and Exley. B, [Eds] (2016). *Pedagogic Rights and Democratic Education: Bernsteinian Explorations of Curriculum, Pedagogy and Assessment*. London Routledge, Taylor Francis Group.

von Davier. M, and Carstensen. CH, (2007). *Statistics for Social and Behavioral Sciences: Multivariate and Mixture Distribution Rasch Models.* London: Springer.

Wagenaar. R, (2014). Competences and Learning Outcomes: A Panacea for Understanding the (new) Role of Higher Education? *Tuning Journal of Higher Education, 1, [2],* 279-302

Wajizigha. BC and Sireci. SG, (2011). Importance of Equating High-Stakes Educational Measurements. *International Journal of Testing, 11, [1]*, 38-52.

Wallman. A, Lindblad. AK, Hall. S, Lundmark. A, and Ring.LA (2008). A Categorization Scheme for

Assessing Pharmacy Students' Levels of Reflection During Internships. *American Journal of Pharmaceutical Education*, *71, [1],* Article 05:1-10.

Walsh. M, Hil- Bailey. P. Mossey. S, and Koren. I, (2010). The Novice Objective Structured Clinical Evaluation Tool: Psychometric Testing. *Journal of Advanced Nursing, 66, [12]*, 2807–2818.

Walters. K, Osborn. D, and Raven. P, (2005). The Development, Validity and Reliability of a Multimodality Objective Structured Clinical Examination in Psychiatry. *Medical Education, 39 [4]*, 292-298.

Walton. D, and Gordon. TF, (2009). Jumping to a Conclusion: Fallacies and Standards of Proof. *Informal Logic 29,* 215-243.

Walton. M, Woodward. H, Van Staalduinen. S, Lemer. C, Greaves. F, Noble. D, Ellis. B, Donaldson. L, and Barraclough. B, (2010). The WHO Patent Safety Curriculum Guide. *Quality and Safety in Healthcare, 19,* 542-546.

Waring. M, and Evans. C, (2015). Understanding Pedagogy: Developing a Critical Approach to Teaching and Learning. London, Routledge, Taylor and Francis Group.

Watson. R, Stimpson. A, Topping. A, and Porock. D, (2002). Clinical Competence Assessment in Nursing: A Systematic Review of the Literature. *Journal of Advanced Nursing, 39 [5]*, 421-431.

Weeks. G, George. J, Maclure. K, and Stewart. D, (2016). Non-medical Prescribing versus Medical Prescribing for Acute and Chronic Disease Management in Primary and Secondary Care. *Cochrane Database of Systematic Reviews*, (11).

Wehlburg, CM. (2011). A Scholarly Approach to Assessing Learning. *International Journal for the Scholarship of Teaching and Learning, 5,[2],* 1–4.

Weigle. SC, (2010). Validation of Automated Scores of TOEFL iBT Tasks Against non-test Indicators of Writing Ability. *Language Testing*, *27, [3]*, 335–353.

Weingarten. HP, (2017). What is academic quality? Available from http://blog-en.heqco.ca/2017/03/harvey-p-weingarten-what-is-academic-quality/. Retrieved 17/05/20.

West. L, (2001). *Doctors on Edge.* London. FA Books.

Wheelahan. L, (2010). *Why Knowledge Matters in Curriculum: A Social Realist Argument.* Abingdon, Oxon: Routledge Publication.

Wheelahan. L, and Moodie. G, (2011). *Rethinking Skills in Vocational Education and Training: From Competencies to Capabilities.* Education and Communities Office of Education. New South Wales.

Wheelahan. L, (2016). Patching Bits won't fix Vocational Education in Australia: A New Model is

Needed. *International Journal of Training Research, 14, [3],* 180–196.

Willis Commission (2012). *Quality with Compassion: The Future of Nursing Education. Report of the Willis Commission on Nursing Education.* London, The Royal College of Nursing.

Willis. P, (2012). *An Existential Approach to Transformative Learning*. In E. W. Taylor and P. Cranton[Eds.], Handbook of Transformative Learning: Theory, Research, and Practice, pp. 212-227. San Francisco: Jossey-Bass.

Winch. C, (2013). Curriculum design and epistemic ascent. *Journal of Philosophy of Education, 47, [1],* 128–46.

Winch, C. (2010). *Dimensions of Expertise: A Conceptual Exploration of Vocational Knowledge.* London: Continuum Press.

Windsor. C, Douglas. C, and Harvey. T, (2012). Nursing and Competencies — a Natural Fit: The Politics of Skill ⁄Competency Formation in Nursing. *Nursing Inquiry, 19 [3],* 213–222.

Wittek. L., and Kvernbekk. T, (2011). On the problems of asking for a definition of quality in education. *Scandinavian Journal of Educational Research*, *55, [6],* 671–684.

Wolf. A, (2001). *Competence-Based Assessment; In Competence in the Learning Society. Chapter 25: 453-466.* New York: Peter Lang Publishers.

Wollack J. A., Maynes. D, (2011). *Data Forensics: What Works, What Doesn't. In Kingston. NM, and Clarke. A, (2014). Test Fraud: Statistical Detection and Methodology* Abingdon Oxon, Routledge, Taylor and Frances.

World Economic Forum (2016). *The Future of Jobs: Employment, Skills and Workforce Strategy for the Fourth Industrial Revolution.* Geneva: World Economic Forum.

Wright. BD, and Mok. MMC, (2004). An Overview of the Family of Rasch Measurement Models. In a. S. Smith. EV, *Introduction to Rasch Measurement: Theory, Models and Applications* (pp. 1-24). Maple Grove. MN: JAM Press.

Wu. A, and Adams. R, (2007). *Applying the Rasch Model to Psycho-social Measurement: A Practical Approach.* Melbourne.: Educational Measurement Solutions.

Yanhua. C , And Watson. R, (2011). A Review of Clinical Competence Assessment in Nursing. *Nurse Education Today, 31 [8]*, 832-836.

Young. MR, (2018). Refection Fosters Deep Learning: The 'Reflection Page and Relevant to You' Intervention. *Journal of Instructional Pedagogies, 20,* 1-17.

Young. M, (2010). Alternative Educational Futures for a Knowledge Society. *European Educational Research Journal,* 9, [1], 1–12.

Young. M, and Muller. J, [Eds] (2014) *Knowledge, Expertise and the Professions.* London, Routledge. Taylor Frances Group.

Young. HN, Len-Rios. ME, Brown, R., Moreno. MM, and Cox. E, (2017). How does Patient-Provider Communication Influence Adherence to Asthma Medications? *Patient Education and Counseling*, *100, [4]*, 696-702.

Yin. RK, (2003). *Case Study Research: Design and Methods [3rd Edn.].* Thousand Oaks, CA: Sage Publications.

Yin. RK, (2014). *Case Study Research: Design and Methods [6th Edn.].* Thousand Oaks, CA: Sage Publications.

Zepke. N, (2013). Threshold Concepts and Student Engagement: Revisiting Pedagogical Content knowledge", *Active Learning in Higher Education 14, [2],* 97-107.

Zhu, X. and Stone, CA, (2011). Assessing Fit of Unidimensional Graded Response Models using Bayesian Methods. *Journal of Educational Measurement, 48, [1]*, 81-97.

Zhu. XW, Yu. LM, Zheng. J, Fang. J, Hou. S, Lu. M, Ma, W, Wang. H, Wu. Z, and Zhang L (2012). Nurse Staffing Levels make a Difference in Patient Outcomes: A Multisite- study in Chinese Hospitals. *Journal of Nursing Scholarship*, *44, [3],* 266–73.

Zlatkin-Troitschanskaia, O., Shavelson, RJ and Kuhn. C, (2015). The International State of Research on Measurement of Competency in Higher Education. *Studies in Higher Education, 40,* 393–411.

Zumbo, BD, (2009). Validity as Contextualized and Pragmatic Explanation and its Implications. In R. W. [Ed], *The Concept of Validity: Revisions, New Directions and Applications* (pp. 65–82). Charlotte, NC: Information Age Publishing.

Zumbo. BD, and Forer. B, (2011). *Testing and Measurement from a Multilevel View: Psychometrics and Validation.* In *High-stakes Testing in Education: Science and Practice in K–12 Settings, p.* 177-190. American Psychological Association.