



University of Portsmouth
Faculty of Technology
Department of Electronic and Computer Engineering,

**INTELLIGENT PERFORMANCE ASSESSMENT IN A
VIRTUAL ELECTRONIC LABORATORY**

By

IFEYINWA EUCHARIA ACHUMBA

(formerly IFEYINWA EUCHARIA CHIKA)

BSc, MSc

Thesis submitted to the academic registry, in partial fulfilment of the requirements for the award of the degree of Doctor of Philosophy of the University of Portsmouth

April 2011



ABSTRACT

Laboratory work, in the Undergraduate Engineering course, is aimed at enhancing students' understanding of taught concepts and integrating theory and practice. This demands that laboratory work is synchronised with lectures in order to maximise its derivable learning outcomes, measurable through assessment. The typical high costs of traditional engineering laboratory, which often militates against its increased use and the synchronisation of laboratory and lectures, have, in addition to other factors, catalysed the increased adoption of virtual laboratories as a complement to the traditional engineering laboratory. In extreme cases, virtual laboratories could serve as alternative means of providing, albeit simulated, meaningful practical experiences. A Virtual Electronic Laboratory (VEL), which can be used to undertake a range of undergraduate electronic engineering curriculum-based laboratory activities, in a realistic manner, has been implemented as part of the work presented in this thesis. The VEL incorporates a Bayesian Network (BN)-based model for the performance assessment of students' laboratory work in the VEL. Detailed descriptions of the VEL and the assessment model are given. The evaluation of the entire system is in two phases: evaluation of the VEL as a tool for facilitating students' deeper understanding of fundamental engineering concepts taught in lectures; and evaluation of the assessment model within the context of the VEL environment. The VEL is evaluated at two different engineering faculties, in two separate universities. Results from the evaluation of the VEL show the effectiveness of the VEL to enhance students' learning, in the light of appropriate learning scenarios, and provide evidence and support for the use of virtual laboratories in the engineering educational context. Performance data, extracted from students' behaviour logs (captured and recorded during the evaluation of the VEL) are used to evaluate the assessment model. Results of the evaluation demonstrate the effectiveness of the model as an assessment tool, and the practicability of the performance assessment of students' laboratory work from their observed behaviour in a virtual learning environment.



DECLARATION

Whilst registered as a candidate for the above degree, I have not been registered for any other research award. The results and conclusions embodied in this thesis are my own (the named candidate) and have not been submitted for any other academic award.



ACKNOWLEDGEMENT

My most profound thanksgiving and reverence to the Almighty and Everlasting God. I owe my deepest gratitude to my sponsors, The Schlumberger Foundation and the Faculty For The Future (FFTF) team, for the opportunity to transform my academic experience. I am heartily thankful to my Director of Studies, Dr Djamel Azzi, whose sustained encouragement, guidance, and support facilitated this work. My gratitude goes to James Stocker for all his assistance and support. I would also like to express my gratitude to: the head of department, Dr Misha Filip for his unwavering support and assistance; Mr Victor Dunn for his assistance and encouragement; and to the other lecturers and staff in the department whose assistance in various forms was invaluable. I offer my regards and thanks to all, who assisted me in any respect in the course of the work.



DEDICATION

To

The Almighty and Everlasting God



TABLE OF CONTENTS

ABSTRACT	ii
DECLARATION.....	iii
ACKNOWLEDGEMENT.....	iv
DEDICATION	v
TABLE OF CONTENTS	vi
LIST OF TABLES	ix
LIST OF FIGURES.....	xi
DISSEMINATION.....	xv
ABBREVIATIONS.....	xvii
CHAPTER 1.....	1
INTRODUCTION.....	1
1.0 BACKGROUND.....	2
1.1 LABORATORY IN THE UNDERGRADUATE ENGINEERING COURSE.....	3
1.1.1 Traditional Laboratory.....	5
1.1.2 Remote Laboratory.....	6
1.1.3 Virtual Laboratory	7
1.2 EFFECTIVENESS OF THE VIRTUAL LABORATORY AS A LEARNING ENVIRONMENT	7
1.3 VIRTUAL LABORATORY IN UNDERGRADUATE ENGINEERING LABORATORY EDUCATION.....	9
1.4 LABORATORY WORK ASSESSMENT	10
1.5 MOTIVATION AND OBJECTIVES OF THE RESEARCH WORK.....	11
1.6 RESEARCH METHODOLOGY	12
1.6.1 Implementation Phase.....	12
1.6.2 Evaluation Phase.....	12
1.7 OVERVIEW OF THE THESIS	13
1.8 SUMMARY.....	14
CHAPTER 2.....	15
LITERATURE REVIEW	15
2.0 INTRODUCTION	16
2.1 RELATED WORK ON VIRTUAL LABORATORY	16
2.2 TRADITIONAL LABORATORY WORK ASSESSMENT PRACTICES	19
2.3 PROBLEMS OF TRADITIONAL LABORATORY WORK ASSESSMENT PRACTICES	20
2.4 PERFORMANCE-BASED ASSESSMENT	21
2.5 RELATED WORK ON INTELLIGENT PERFORMANCE ASSESSMENT IN THE VIRTUAL LEARNING ENVIRONMENT	22
2.6 BAYESIAN NETWORK ARTIFICIAL INTELLIGENCE TECHNIQUE	24
2.6.1 Probability Theory, Bayes Theorem, and Joint Probability Distribution.....	26
2.6.2 Conditional Independence	28
2.6.3 D-Separation.....	29
2.6.4 Marginalization.....	30
2.6.5 Inference	31
2.6.6 Advantages of Bayesian Network Modeling	33
2.6.7 Disadvantages of Bayesian Network Modeling.....	35
2.7 BEHAVIOUR LOG ANALYSIS FOR PERFORMANCE DATA EXTRACTION	35
2.8 COGNITIVE TASK ANALYSIS	37
2.9 SUMMARY.....	38
CHAPTER 3.....	40
THE VIRTUAL ELECTRONIC LABORATORY: DESCRIPTION AND EVALUATION.....	40

3.0	INTRODUCTION	41
3.1	DESIGN AND IMPLEMENTATION CONSIDERATIONS	41
3.2	DESCRIPTION	42
3.2.1	The Front-end (The client-side).....	44
3.2.2	The Back-end (The Server-side).....	46
3.2.2	Netlist Generation.....	47
3.3	EVALUATION	48
3.3.1	Methodology and Context	48
3.3.2	Sample Student Cohorts	49
3.3.3	Evaluation Process.....	50
3.4	DATA ANALYSIS	53
3.4.1	Statistical Analysis	54
3.4.2	Concentration Analysis.....	59
3.5	INDIVIDUAL STUDENT ADVANTAGE IN THE VEL ENVIRONMENT	62
3.6	FEEDBACK FROM STUDENTS	63
3.6	SUMMARY.....	65
CHAPTER 4.....		66
BAYESIAN NETWORK-BASED INTELLIGENT ASSESSMENT MODEL: STRUCTURE REALISATION AND CALIBRATION		66
4.0	INTRODUCTION	67
4.1	THE MODEL FOUNDATIONAL FRAMEWORK	67
4.1.1	Psychology of Learning: the Basis for Derivation of the Foundational Framework	67
4.1.2	Derivation of the Foundational Framework.....	68
4.2	STRUCTURE REALIZATION	71
4.3	CALIBRATION	77
4.3.1	Knowledge Elicitation	77
4.3.2	The Calibration Process.....	78
4.4	OVERVIEW OF THE MODEL BELIEF ESTIMATION PROCESS.....	81
4.5	SUMMARY.....	83
CHAPTER 5.....		84
INVESTIGATION OF THE OPTIMAL CONSTRUCTION APPROACH FOR THE BAYESIAN NETWORK-BASED ASSESSMENT MODEL		84
5.0	INTRODUCTION	85
5.1	BAYESIAN NETWORK MODEL CONSTRUCTION APPROACHES	87
5.1.1	Expert-Centred Approach.....	87
5.1.2	Data-Centred Approach.....	88
5.2	INVESTIGATION PROCEDURE.....	91
5.2.1	Construction of the Data-centred Models used for the Investigation.....	92
5.2.2	Evaluation Method	94
5.3	OPTIMALITY CRITERIA AND COMPARATIVE TOOLS	95
5.3.1	Structure Comparison	95
5.3.2	Performance Comparison Metrics	96
5.4	RESULTS	97
5.6	COMMENTS	103
5.7	SUMMARY.....	105
CHAPTER 6.....		106
ASSESSMENT MODEL: DATA EXTRACTION, EVIDENCE VARIABLE QUANTIZATION AND INSTANTIATION		106
6.0	INTRODUCTION	107
6.1	PERFORMANCE DATA EXTRACTION	107
6.2	EVIDENCE QUANTIZATION CASE STUDIES.....	111
6.2.1	Quantization of Node 1: “ <i>UseOfCorrectComponents</i> ”	111
6.2.2	Quantization of Node 2: “ <i>CorrectPlacementOfComponents</i> ”	112
6.2.3	Quantization of Node 33: “ <i>ExperienceUsingVEL</i> ”	114
6.3	EVIDENCE INSTANTIATION	114
6.4	STUDENT LABORATORY WORK SCHEMA	117
6.5	SUMMARY	119
CHAPTER 7.....		120

ASSESSMENT MODEL VERIFICATION, EVALUATION AND APPLICATION	120
7.0 INTRODUCTION	121
7.1 VERIFICATION.....	121
7.1.1 Functional Verification.....	121
7.1.2 Sensitivity Analysis.....	122
7.2 EVALUATION	125
7.2.1 Reliability	125
7.2.2 Validation	127
7.3 MODEL APPLICATION	142
7.4 SUMMARY.....	144
CHAPTER 8.....	145
DISCUSSION, LIMITATIONS, CONTRIBUTIONS, FURTHER WORK, AND CONCLUSION.....	145
8.0 INTRODUCTION	146
8.1 DISCUSSION	146
8.2 CHALLENGES	148
8.3 CONTRIBUTIONS	149
8.4 FURTHER WORK	153
8.5 CONCLUSION	154
REFERENCES	157
APPENDIX A: PRE-TEST AND POST-TEST SCORES	171
APPENDIX B: CONCENTRATION ANALYSIS FORMULA, DETAILS, AND STUDENTS' RESPONSE DISTRIBUTIONS.....	173
APPENDIX B1: THE CONCENTRATION ANALYSIS FORMULA AND DETAILS	173
APPENDIX B2: STUDENTS' RESPONSE DISTRIBUTION FOR THE PRE-TEST POST-TEST ITEMS.....	174
APPENDIX B3: THE CONCENTRATION ANALYSIS RESULTS AND DISCUSSION	176
APPENDIX C: STUDENTS' SCORES ON THE LEARNING STYLE PREFERENCES SCALES AND USABILITY SURVEY ITEMS	179
APPENDIX D: DOMAIN EXPERTS' LINK RATINGS	183
APPENDIX E: EMPIRICAL INVESTIGATION RESULTS	185
APPENDIX F: QUANTIZATION OF EVIDENCE VARIABLES	187
APPENDIX F1: Quantization of Node 2: "CorrectPlacementOfComponents"	187
APPENDIX F2: Quantization of the node 3: "UseCorrAlternativeComponent"	192
APPENDIX F3: Quantization of the node 5: "ObserveSafetyProcedure".....	193
APPENDIX F4: Quantization of the node 6: "OperateEquipmentCorrectly".....	194
APPENDIX F5: Quantization of the nodes 8: "ApplyCorrectFormula"	194
APPENDIX F6: Quantization of the node 9: "UseCorrectValues"	195
APPENDIX F7: Quantization of the node 12: "SpecificActivityConstraints"	195
APPENDIX F8: Quantization of the node 17: "AbilityToMakeAppropriateMeasurements"	195
APPENDIX F9: Quantization of the node 18: "AbilityToPlotGraph"	196
APPENDIX F10: Quantization of the node 19: "FollowInstructions"	196
APPENDIX F11: Quantization of the node 20: "ReadDiagramsAndGraphs"	196
APPENDIX F12: Quantization of the node 22: "AbilityToInterpretExpResults"	197
APPENDIX F13: Quantization of the node 38: "ConceptualUnderstanding"	197
APPENDIX F14: Quantization of the node 36: "FactualKnowledge"	198
APPENDIX F15: Quantization of the node 30: "PreLaboratoryTestScore"	198
APPENDIX F16: Quantization of the node 31: "PreRequisiteUnitTestScore"	198
APPENDIX F17: Quantization of the node : "PerceivedInfoType"; "InfoCommChannel"; "InfoProcessing"; and "ProgressionToUnderstanding"	199
APPENDIX G: ASSESSOR CONSISTENCY	200
APPENDIX H: INTER ASSESSOR AGREEMENT.....	209
APPENDIX J: HUMAN ASSESSORS VS MODEL	212



LIST OF TABLES

Table 2.1: Performance Factors for a Physical Observation technique.....	20
Table 3.1: Netlist file description of the circuit of Figure 3.4.....	48
Table 3.2: A knowledge goal and its outcome and pre-requisite concepts	49
Table 3.3: Statistical significance test results.....	58
Table 3.4: An arbitrary sample set of students' response distributions for test items	59
Table 3.5: Concentration analysis of students' responses to the pre- and post-test items.	61
Table 4.1: A HLAS and its associated BACs.....	73
Table 4.2: A BAC and its associated LLBACs.....	73
Table 4.3: Parameter assignment	79
Table 4.4: Fair Prior values for an Observable variable, A	80
Table 4.5: The CPT for Node 4 of Figure 4.3: “ <i>AbilityToWorkWithComponents</i> ”	81
Table 5.1: Number of possible BN structures for various numbers of variables.	91
Table 5.2: Training and Test Datasets.....	98
Table 5.3: Parametric Models: Performance indices.....	99
Table 5.4: Non-parametric models: Performance indices	101
Table 6.1: A list of some of the actions that can be taken in the VEL.....	110
Table 6.2: Summary highlight of the fuzzification rules.....	116
Table 7.1: Inter-Assessor correlation factors, r	133
Table 7.2: Correlation Factors and P-values of Assessors Scores for <i>PUE</i> , <i>WWC</i> , <i>SRC</i> , <i>MAC</i> , and Performance, Compared to the Model Belief Estimates.	141
Table A1: UoP participants’ pre-/post-test Scores and Percentage Gain	171
Table A2: UTON participants’ pre-/post-test Scores and Percentage Gain	172
Table B1: Three-level coding scheme for Score and Concentration Factor.....	174
Table B2.1: UoP Response Distribution for each Choice of the Pre-Test Items	174
Table B2.2: UoP Response Distribution for each Choice of the Post-Test Items	175
Table B2.3: UTON Response Distribution for each Choice of the Pre-Test Items.....	175
Table B2.4: UTON Response Distribution for each Choice of the Post-Test Items	176
Table B3.1: Score, Concentration factor, and SRPs for the UoP Cohort.....	176
Table B3.2: Score, Concentration factor, and SRPs for UTON Cohort.....	177
Table C1: UoP students’ scores on the LS preferences scales	179
Table C2: UTON students’ scores on the LS preferences scales	180
Table C3: UoP Scores for some of the Usability Survey Items	181
Table C4: UTON Scores for some of the Usability Survey Items	182
Table D1: Domain Expert Link Weight Assignments	183
Table D2: Pearson Correlation factor of the expert link weightings.....	184

Table E1: Average performance metric values for the parameteric groups of models and their maximum standard deviation	185
Table E2: Average performance metric values for the non-parameteric groups of models and their maximum standard deviation	186
Table G1: Assessor Consistency Correlation factors and p-values.....	208



LIST OF FIGURES

Figure 2.1: A sample BN model: parent-child relationship	25
Figure 2.2: Sample BN models illustrating the concept of conditional independence	28
Figure 2.3: Different types of dependency topologies (d-connections)	29
Figure 2.4: A sample BN for explaining marginalization	31
Figure 2.5: Student Topic and Concept Mastery BN model	32
Figure 3.1: Structural overview of the VEL	43
Figure 3.2: The VEL GUI	45
Figure 3.3: Sample listing of events log	46
Figure 3.4: Sample circuit schematic	47
Figure 3.5: The sample circuit schematic after wire removal and node number assignment	48
Figure 3.6: One of the laboratory activities undertaken by students in the VEL environment	52
Figure 3.7: A student's built circuit for the laboratory activity of Figure 3.6, in the VEL environment	52
Figure 3.8: A student's graphical output for the laboratory activity of Figure 3.6	53
Figure 3.9: UoP cohort's pre-/post-tests scores	55
Figure 3.10: UoP cohort's percentage knowledge gain scattergram	55
Figure 3.11: UTON cohort's pre-/post-tests scores	56
Figure 3.12: UTON cohort's percentage knowledge gain scattergram	57
Figure 4.1: Basic framework of learning	69
Figure 4.2: The foundational framework of the LAP model	70
Figure 4.3: Graphical view of the LAP model	76
Figure 4.4: Categorization of the LAP model nodes	76
Figure 4.5: Number line knowledge elicitation instrument	78
Figure 4.6: A Simple five node BN	78
Figure 4.7: A sub-model of the LAP Model	82
Figure 5.1: A form of a dataset	88
Figure 5.2: Average performance indices of the <i>parameteric</i> models and the Reference model	100
Figure 5.3: Average performance indices of the <i>non-parameteric</i> models and the Reference model	102
Figure 5.4: The CPT of a node in a <i>non-parameteric sedacen</i> model, PND2np	104
Figure 6.1: Data extraction and evidence quantization and instantiation processes.	108
Figure 6.2: Interaction events transformation: Selection and Aggregation	108
Figure 6.3: Interaction events transformation: Abstraction	109
Figure 6.4: Aggregation of atomic events to highlight an action.	110
Figure 6.5: Sequence detection and count and summary statistics	111

Figure 6.6: Circuit schematic of one of the laboratory activities used for the evaluation of the VEL: <i>Twin-T Notch filter</i>	113
Figure 6.7: The graph, G , describing the network of Figure 6.6.....	113
Figure 6.8: Fuzzification Engine.....	114
Figure 6.9: The trapezoidal function representation for node 1 of Figure 4.3, “ <i>UseCorrectComponents</i> ”	116
Figure 6.10: Stud_A laboratory work schema.....	117
Figure 6.11: Stud_B laboratory work schema.....	117
Figure 6.12: Stud_C laboratory work schema.....	118
Figure 7.1: Extract from LAP model before node 2 was instantiated.....	122
Figure 7.2: Extract from LAP model after node 2 was instantiated.....	122
Figure 7.3: Initial sensitivity of the target node to the evidence nodes.....	124
Figure 7.4: Sensitivity of the target node, after parameter adjustment.....	124
Figure 7.5: Test for reliability.....	127
Figure 7.6: Assessor A1 1 st and 2 nd assessment scores for the performance factor, <i>WWC</i>	129
Figure 7.7: Assessor A2 1 st and 2 nd assessment scores for the performance factor, <i>WWC</i>	130
Figure 7.8: Assessor A3 1 st and 2 nd assessment scores for the performance factor, <i>WWC</i>	130
Figure 7.9: Assessor A4 1 st and 2 nd assessment scores for the performance factor, <i>WWC</i>	131
Figure 7.10: Assessors scores for the performance factor, <i>PUE</i>	132
Figure 7.11: Assessors’ scores for the performance factor, <i>WWC</i>	132
Figure 7.12: Assessor A1 scores vs. model belief estimates for <i>PUE</i>	134
Figure 7.13: Assessor A2 scores vs. model belief estimates for <i>PUE</i>	134
Figure 7.14: Assessor A3 scores vs. model belief estimates for <i>PUE</i>	135
Figure 7.15: Assessor A4 scores vs. model belief estimates for <i>PUE</i>	135
Figure 7.16: Assessors average scores vs. model belief estimates for <i>PUE</i>	136
Figure 7.17: Assessor A1 scores vs. model belief estimates for <i>WWC</i>	136
Figure 7.18: Assessor A2 scores vs. model belief estimates for <i>WWC</i>	137
Figure 7.19: Assessor A3 scores vs. model belief estimates for <i>WWC</i>	137
Figure 7.20: Assessor A4 scores vs. model belief estimates for <i>WWC</i>	138
Figure 7.21: Assessors average scores vs. model belief estimates for <i>WWC</i>	138
Figure 7.22: Assessor A1 performance scores vs. model belief estimates for <i>Performance</i>	139
Figure 7.23: Assessor A2 performance scores vs. model belief estimates for <i>Performance</i>	140
Figure 7.24: Assessor A3 performance scores vs. model belief estimates for <i>Performance</i>	140
Figure 7.25: Assessor A4 performance scores vs. model belief estimates for <i>Performance</i>	140
Figure 7.26: Assessors average performance scores vs. model belief estimates for <i>Performance</i>	141
Figure 7.27: A view of the instructor interface GUI.....	143
Figure F1.1: The schematic of one of the circuits constituting part of the laboratory activities designed for the evaluation of the VEL: the <i>Twin-T Notch filter</i>	187
Figure F1.2: G : The graph description of the network of Figure F1.1.....	187
Figure F1.3: A chosen tree, T , of the graph, G	189

Figure F1.4: Oriented Loops of the graph, G	190
Figure F1.5: FSC associated with the tree, T , of the graph G	191
Figure G1: Assessor A1 1 st and 2 nd assessment scores for the performance factor, APG.....	200
Figure G2: Assessor A1 1 st and 2 nd assessment scores for the performance factor, IER	200
Figure G3: Assessor A1 1 st and 2 nd assessment scores for the performance factor, MAC.....	201
Figure G4: Assessor A1 1 st and 2 nd assessment scores for the performance factor, MAM.....	201
Figure G5: Assessor A1 1 st and 2 nd assessment scores for the performance factor, PUE	201
Figure G6: Assessor A1 1 st and 2 nd assessment scores for the performance factor, SRC	202
Figure G7: Assessor A2 1 st and 2 nd assessment scores for the performance factor, APG.....	202
Figure G8: Assessor A2 1 st and 2 nd assessment scores for the performance factor, IER	202
Figure G9: Assessor A2 1 st and 2 nd assessment scores for the performance factor, MAC.....	203
Figure G10: Assessor A2 1 st and 2 nd assessment scores for the performance factor, MAM.....	203
Figure G11: Assessor A2 1 st and 2 nd assessment scores for the performance factor, PUE	203
Figure G12: Assessor A2 1 st and 2 nd assessment scores for the performance factor, SRC	204
Figure G13: Assessor A3 1 st and 2 nd assessment scores for the performance factor, APG.....	204
Figure G14: Assessor A3 1 st and 2 nd assessment scores for the performance factor, IER	204
Figure G15: Assessor A3 1 st and 2 nd assessment scores for the performance factor, MAC.....	205
Figure G16: Assessor A3 1 st and 2 nd assessment scores for the performance factor, MAM.....	205
Figure G17: Assessor A3 1 st and 2 nd assessment scores for the performance factor, PUE	205
Figure G18: Assessor A3 1 st and 2 nd assessment scores for the performance factor, SRC	206
Figure G19: Assessor A4 1 st and 2 nd assessment scores for the performance factor, APG.....	206
Figure G20: Assessor A4 1 st and 2 nd assessment scores for the performance factor, IER	206
Figure G21: Assessor A4 1 st and 2 nd assessment scores for the performance factor, MAC.....	207
Figure G22: Assessor A4 1 st and 2 nd assessment scores for the performance factor, MAM.....	207
Figure G23: Assessor A4 1 st and 2 nd assessment scores for the performance factor, PU	207
Figure G24: Assessor A4 1 st and 2 nd assessment scores for the performance factor, SRC	208
Figure H1: Inter Assessor Agreement: Assessors scores for the performance indicator, SRC	209
Figure H2: Inter Assessor Agreement: Assessors scores for the performance indicator, MAM	210
Figure H3: Inter Assessor Agreement: Assessors scores for the performance indicator, IER	210
Figure H4: Inter Assessor Agreement: Assessors scores for the performance indicator, MAC	211
Figure H5: Inter Assessor Agreement: Assessors scores for the performance indicator, APG	211
Figure J1: Human-Model Agreement--Assessor A1 scores vs. model belief estimates for SRC	212
Figure J2: Human-Model Agreement--Assessor A2 scores vs. model belief estimates for SRC	212
Figure J3: Human-Model Agreement--Assessor A3 scores vs. model belief estimates for SRC	213
Figure J4: Human-Model Agreement--Assessor A4 scores vs. model belief estimates for SRC	213
Figure J5: Human-Model Agreement--Assessors average scores vs. model belief estimates for SRC	214
Figure J6: Human-Model Agreement--Assessor A1 scores vs. model belief estimates for MAC	214
Figure J7: Human-Model Agreement--Assessor A2 scores vs. model belief estimates for MAC	215
Figure J8: Human-Model Agreement--Assessor A3 scores vs. model belief estimates for MAC	215

Figure J9: Human-Model Agreement--Assessor A4 scores vs. model belief estimates for *MAC*. 216

Figure J10: Human-Model Agreement--Assessors average scores vs. model belief estimates for *MAC*
..... 216



DISSEMINATION

Articles

1. **I. E. Chika**, "Building Better Engineers: Extending and Consolidating the Gains", *IEEE GOLDRush*, September, 2008, pp. 12 - 13

Conference Proceedings

- i. **I. E. Achumba**, D. Azzi, and R. Khusainov, "Bayesian Network Model for Students' Laboratory Work Performance Assessment: An empirical investigation of the optimal construction approach", Accepted for oral presentation at the *International Conference on Computational Intelligence (ICCI 2011)*, Paris, France, July 27 – 29, 2011.
- ii. **I. E. Achumba**, "The Virtual Electronic Laboratory (VEL): Description and Demonstration", *European Conference on Library of Labs (LILA 2011)*, Selwyn College, University of Cambridge, United Kingdom (UK), 11 - 12 April, 2011.
- iii. **I. E. Chika**, D. Azzi, A. Hewitt, & J. Stocker, "A Holistic Approach To Assessing Students' Laboratory Performance Using Bayesian Networks", *Proceedings of IEEE Workshop on Computational Intelligence in Virtual Environments (CIVE '09)*, Mar. 29 to April 2, 2009, Nashville, USA, pp. 26 -32.
- iv. **I. E. Chika**, D. Azzi, Barry P. Haynes, and Roy Williams, "Students Laboratory Work Performance Assessment", *Proceedings of the International Conference on Interactive Computer Aided Learning (ICL 2008)*, Villach, Austria, September 24 – 26, 2008.
- v. **I. E. Chika**, D. Azzi, J. Stocker, and B. P. Haynes, "Genuine Lab Experiences for Students in Resource Constrained Environments: The RealLab with Integrated Intelligent Assessment", *Proceedings of the International Conference on Interactive Mobile and Computer Aided Learning (IMCL)*, Amman, Jordan April 16 - 18, 2008
- vi. D. Azzi, **I. E. Chika**, and B. P. Haynes, "Engineering Learning and Assessment: Current and Emerging Trends", *Proceedings for the National Engineering Conference*, Abuja, Nigeria, 3 – 7 December, 2007.
- vii. **I. E. Chika**, D. Azzi, and B. P. Haynes, "A Paradigm Shift in Engineering Education: Implications for Assessment, Instruction and Curriculum", *Proceedings of the International Conference for Adaptive Science and Technology (ICAST' 07)*, Accra, Ghana, 10 – 13 December, 2007.

Journal Papers

1. D. Azzi, **I. E. Achumba**, V. L. Dunn, and G. A. Chukwudebe, “Intelligent Performance Assessment of Students’ Laboratory Work in a Virtual Electronic Laboratory Environment”, Paper submitted to *IEEE Transactions on Learning Technologies*, for review. Submitted on 09/Dec./2010.
2. **I. E. Achumba**, and D. Azzi, “A Virtual Electronic Laboratory for Genuine Practical Experiences: Description and Evaluation”, Paper submitted to *IEEE Transactions on Learning Technologies*, for review. Revised and resubmitted on 09/Dec./2010.
3. **I. E. Achumba**, D. Azzi, and J. Stoker, “Low-Cost Virtual Laboratory Workbench for Electronic Engineering”, *International Journal of Virtual and Personal Learning Environments (IJVPLE)*, vol. 1, no. 4, pp. 6 – 22, Oct.-Dec., 2010
4. **I. E. Chika**, Djamel Azzi, James Stocker and Barry P. Haynes, “Genuine Lab Experiences for Students in Resource Constrained Environments: The RealLab with Integrated Intelligent Assessment”, *IEEE Multidisciplinary Engineering Education Magazine*, vol. 3, no. 4, December, 2008.
5. **I. E. Chika**, D. Azzi, J. Stoker, and . P. Haynes, “Genuine Lab Experiences For Students in Resource Constrained Environments: The RealLab with Integrated Intelligent Assessment”, *International Journal of Online Engineering (IJOE)*, vol. 4, no 3, August 2008.



ABBREVIATIONS

▪ ABET	Accreditation Board for Engineering and Technology
▪ AI	Artificial Intelligence
▪ ANN	Artificial Neural Network
▪ ATS	Adaptive Tutoring System
▪ BAC	Basic Ability Component
▪ BN	Bayesian Network
▪ BS	Brier Score
▪ C	Concentration Factor
▪ CAM	Concentration Analysis Method
▪ CI	Conditional Independence
▪ CM	Cutset Matrix
▪ COE	Communication Engineering
▪ CPT	Conditional Probability Table
▪ CSS	Cascading Style Sheets
▪ CTA	Cognitive Task Analysis
▪ DAG	Directed Acyclic Graph
▪ DAQ	Data Acquisition
▪ DLD	Digital Logic Design
▪ ECE	Electronic and Computer Engineering
▪ EEE	Electrical and Electronic Engineering
▪ ESA	Evidence-based Sensitivity Analysis
▪ FCM	Fundamental Cutset Matrix
▪ FG	Function Generator
▪ FL	Fuzzy Logic
▪ FLM	Fundamental Loop Matrix
▪ GTT	Greedy Thick Thinning
▪ GUI	Graphical User Interface
▪ HTML	HyperText Markup Language
▪ IAM	Intelligent Assessment Model
▪ ILE	Intelligent Learning Environment
▪ ILS	Intelligent Learning Systems
▪ IM	Incident Matrix
▪ IT	Information Technology
▪ ITS	Intelligent Tutoring System
▪ JPD	Joint Probability Distribution
▪ KCL	Kirchoff Current Law
▪ KVL	Kirchoff Voltage Law
▪ LAP	Laboratory Performance
▪ LLBAC	Lower Level Basic Ability Component
▪ LM	Loop Matrix
▪ Logloss	Logarithmic Loss
▪ LS	Learning Style
▪ MCI	Multiple Choice Items
▪ MySQL	My Structured Query Language

- OSCE Objective Structural Clinical Examination
- OSTE Objective Structural Technical Examination
- PHP (recursive acronym for PHP: Hypertext Preprocessor)
- PPT Prior Probability Table
- PS Power Supply
- PSA Parameter-based Sensitivity Analysis
- PSE Power System Engineering
- PTPT Pre-Test and Post-Test
- QAA Quality Assurance Agency
- RIM Reduced Incident Matrix
- S Score
- SAM Statistical Analysis Method
- SaS Score and Search
- SEM Structural Expectation Maximization
- SQL Structured Query Language
- SRP Student Response Pattern
- UE Undergraduate Engineering
- UK-HEA UK Higher Education Agency
- UoP University of Portsmouth
- UTON University of Technology, Owerri
- UVL Universal Virtual Lab
- VEL Virtual Electronic Laboratory
- VLE Virtual Learning Environment



CHAPTER 1

INTRODUCTION

The beginning is the most important part of the work.

Plato

327BC - 427BC

1.0 BACKGROUND

The aim of engineering faculties all over the world is to “build better engineers”. Consequently, many engineering faculties have embraced the student-centred learning educational paradigm, which seeks to improve the quality of students’ learning through a redefinition of the learning environment, the role of the instructor, the role of the learner, and the relationship among them. Faculty, in this context, refers to the entity or a division within the organizational structure of a University, comprising one subject area or a number of related subject areas. The student-centred paradigm demands active learning by students and the increased use of the laboratory in the instruction process, to integrate theory and practice. This is imperative because, to build better engineers, students in their early undergraduate years, need to overcome the challenge of grasping the fundamental engineering concepts taught in lectures, in order to have a more solid foundation for their later undergraduate years.

The typical high costs of the traditional engineering laboratory often mean inadequate laboratory facilities in engineering faculties. This may especially be the case in developing countries where engineering faculties are more likely to be challenged by resource constraints and/or large class sizes, such that some students may graduate without meaningful laboratory experiences. Hence, engineering faculties are often challenged with the provision of meaningful laboratory experiences for their students. This challenge has catalytically contributed to the increasing adoption of virtual laboratories in the Undergraduate Engineering (UE) course. The aim is not to replace the traditional laboratory with virtual laboratories, but to complement it. However, the virtual laboratory could serve as an alternative in extreme cases [1]. Virtual laboratories are listed by The UK Higher Education Agency (UK HEA) as one of the alternative approaches to laboratory education because they can offer equivalent laboratory experiences [2].

The work reported in this thesis is focused on the implementation and evaluation of a low-cost Virtual Electronic Laboratory (VEL), and the application of probabilistic reasoning Artificial Intelligence (AI) technique to the construction of an assessment model, for the performance assessment of students’ laboratory work in the VEL environment. The assessment tool is vital because, engineering faculties, in addition to the ongoing challenge of providing meaningful laboratory experiences for students, are also challenged with the performance assessment of students’ laboratory work, especially where there are large class sizes.

First, in section 1.1, the concept of laboratory in the UE course is explored, differentiating between real, remote and virtual laboratories. Next, the effectiveness of virtual laboratories as learning environments is discussed in section 1.2, before looking at virtual laboratories in the context of UE laboratory education, in section 1.3. Section 1.4 highlights the challenges of students' laboratory work assessment in the UE course. The motivation and objectives of the research work are presented in section 1.5, while section 1.6 highlights the research methodology. An overview of the thesis is given in section 1.7, and the chapter summarised in section 1.8.

1.1 LABORATORY IN THE UNDERGRADUATE ENGINEERING COURSE

Laboratory work in the UE course have been classified into three types, controlled assignments, experimental investigation, and projects, and their differences are the way they are organized in terms of the [3]:

- nature of the activities to be undertaken,
- style of the instruction given to the students,
- relative amounts of time spent on the activities,
- stages in the overall course at which they are introduced.

Controlled assignments (also referred to as instructional laboratories) are described as relatively short laboratory exercises, devised and used by the instructor to enhance students' understanding of concepts [3]. Controlled assignments will be interchangeably referred to, in this context, as laboratory work or activities, and is the focus of the research work presented in this thesis.

In the early years of the UE course, students are challenged with grasping the fundamental concepts. For example, it has been highlighted that Electrical and Electronic Engineering (EEE) students find it difficult to understand the fundamental concepts in taught in lectures, in the early years of their study [4][5][6]. Failure to understand the fundamental concepts may affect their learning of more advanced concepts [6][7]. The course units in the early years serve two purposes: lay a foundation for the students to build on in their later undergraduate years of study; and equip students in some other engineering courses and related disciplines with the knowledge to support learning in their fields of study [8]. Fundamental concepts are the curriculum components that constitute the foundation for most other component parts of the curriculum.

The need to enhance students' deeper understanding of fundamental concepts is justifiably the main reason why laboratory activities are concentrated in the early years of the UE course. Laboratory activities, used to involve students in practical experiences with taught concepts, help deepen their understanding of the concepts by integrating theory with practice [2]. The objectives of laboratory work have been highlighted by [3][9][10][11][12]. These objectives include:

1. illustration and/or reinforcement of theory learned from lectures and tutorials.
2. illustration of topics more readily understood if introduced via laboratory rather than by means of lectures.
3. development of abilities/skills in the apt use of components, techniques, and equipment.
4. development of the ability to observe and check data.
5. stimulating interest in a subject

These objectives demand the synchronisation of laboratory activities with lectures in order to maximise their derivable learning outcomes, measurable through assessment. The timing of laboratory activities is important, to maximise its effectiveness [13].

Large class sizes often entail the division of students into groups, for weekly laboratory time slots. This, in addition to the high costs of the traditional engineering laboratory, often militates against the:

- synchronisation of laboratory activities and lectures.
- increased use of the laboratory to enhance students' understanding of taught concepts.

That is, providing meaningful laboratory experiences for students, synchronisation of laboratory activities and lectures, and increasing the use of laboratory exercises in the instructional process, present challenges to engineering faculties. The challenges crystallised the need for alternative laboratory arrangements.

For instance, at the University of Technology, Owerri (UTON), an additional complimentary approach to Electrical and Electronic Engineering (EEE) laboratory education may be beneficial. The engineering faculty at UTON offers a 5-year undergraduate EEE programme. All the EEE students take the same course units in their first three years of study. During this three-year period, the students are exposed to the foundations of science and general studies, fundamental EEE course principles/concepts and circuit theory. Laboratory activities are concentrated mainly in the second and third years of study. At the end of the third year, students choose their preferred EEE course options: Communication Engineering (COE), Power Systems Engineering (PSE) and Electronic and Computer Engineering (ECE). As the

students enter into their fourth and fifth years of study, they concentrate mainly on course units for their respective options. At this level, they begin to apply the learning gains from their earlier years, to the learning of more advanced concepts and specific problem contexts. This is when learning gaps, if any, may be highlighted.

Learning gaps may exist in students' learning, in an engineering faculty, if, for example:

1. laboratory activities are based more on the traditional laboratory setting.
2. there are resources constraints, in the face of high costs of laboratory equipment.
3. there are large class sizes such that students out-number the available laboratory resources.

Basically, there are three types of laboratories: real (traditional), remote and virtual. Nowadays, all three types of laboratories are computer mediated, the major difference being the psychology of presence and the degree of computer mediation [15]. The real laboratory is the traditional fixed-space laboratory. Both remote and virtual laboratories are Virtual Learning Environments (VLE), but throughout literature, their definitions are inconsistent and confusing. The terms are often used interchangeably, despite the significant differences in their technologies.

1.1.1 Traditional Laboratory

The traditional laboratory is characterized by:

- real data, equipment, and components.
- space requirements and time restrictions.
- supervisor and user physical presence requirements.
- high costs.
- waste of consumables.

In the traditional laboratory, “student cohorts are often subdivided into groups to undertake a rolling programme of practical sessions that is often out of sync with the lecture programme delivering the associated theoretical concepts” [2]. Students are often required to re-produce routine experiments that yield well-known results and to focus on experimental methodology rather than on developing skills [2]. In addition, they are often exposed to the use of expensive laboratory equipment that provide many complex functions oriented to the professional market, which are not needed for educational purposes [16]. Furthermore, laboratory sessions are time consuming to organize, manage and assess [2].

Despite these often emphasised limitations of the traditional laboratory, its proponents, according to [17], argue that students need to be exposed to physical contact with laboratory equipment, components, and the uncertainties of laboratory work, because hands-on experience is the best form of learning and there are more information and cues in working with real equipment and people. [17] stated that the argument is not backed by appropriate research study, and that the preference for a particular type of communication method might be the resultant effect of social factors and opportunities for choice, rather than technologically determined. Also, according to [18], individuals adapt to whatever media of communication are available and to which they are continually exposed. For example, [19] stated that, presently, technologies have pervaded and improved education to the extent that instructors and students now expect their educational experiences to include the use of interactive web-based online instruction/learning tools.

1.1.2 Remote Laboratory

Remote laboratories are characterized by mediated reality and are mainly for data acquisition and experimental monitoring and control [15]. They are real laboratories with remote user presence and internet/web mediated access and operations [20]. They require physical space to setup and can employ both real and virtual instruments. Laboratory devices are interfaced to local computers, using interface hardware (such as Data Acquisition (DAQ) boards) and real-time communication software. Middleware software facilitates remote clients' connection to the local computers, through the Internet and/or local network, for laboratory device manipulation and/or real world data acquisition.

Remote laboratories make it possible for many users to share remote and/or specific scarce laboratory resources. Control systems engineering, robotics, and mechanical engineering are areas that seem to have benefited from the remote laboratory technology. [21] described a remote laboratory for the tele-operation of an omni-directional vehicle, and [22] reported a remote laboratory for the tele-operation of robots. A rotating plate is remotely controlled to keep a ball in a specific position by modifying the parameters of the controller, while visualizing the setup using a camera. The user can introduce a disturbance to verify the stability of the system. They are commonly employed in distance learning education [20].

1.1.3 Virtual Laboratory

Virtual laboratories are software versions of the traditional laboratory where experimental setups are implemented in software such that a Personal Computer (PC) can take the place of a laboratory workbench [23]. They provide enabling environments for demonstrating/exploring concepts taught in lectures and/or concepts not readily explored in physical ways. The characteristics and benefits of virtual laboratories are articulated by [14], which include enhancing students' understanding of concepts.

Despite their acknowledged virtues, virtual laboratories are said to be limited in their function as laboratory environments because, they are not able to convey the seriousness of technical procedures [22]. The dangers and risks involved are not taken seriously and a certain degree of thoughtlessness in technical procedures occurs which may be a disadvantage later in the professional years [22]. This thoughtlessness with regards to procedures also exists in the traditional laboratory environment where students' disregard rules due to either forgetfulness or complacency [25]. To address this problem in the virtual laboratory environment, [25] developed a series of virtual reality based laboratory accidents to argue virtual laboratories and provide valuable learning experiences with respect to the importance of safe laboratory practices.

1.2 EFFECTIVENESS OF THE VIRTUAL LABORATORY AS A LEARNING ENVIRONMENT

Literature indicates that Virtual Learning Environments (VLEs) are seen to be representative of real world learning environments, and that users are able to learn effectively in the VLE. They have been proven to have positive impact on students' learning and performance [26] [27] [28]. In some cases, the evidence contradicts popular hypotheses that face-to-face instructor-student interaction is a more valuable experience and produces better results. [17] found students' learning outcomes with virtual laboratories to be equal or superior to that of the traditional laboratory, in terms of promoting students' understanding of course concepts specifically related to the laboratory topic, suggesting that in courses where the laboratory is intended primarily to aid the conceptual understanding of the course content, virtual laboratories can be valuable tools.

Virtual laboratories have been shown to be equivalent to the physical laboratory for explaining and reinforcing concepts [29]. [30] investigated the extent to which a software

simulation of electronic circuits' laboratory can support beginner electrical engineering students, and the extent to which laboratory simulations can replace some physical electronics laboratories. They found equivalent, and in some cases improved, learner performance with decreasing time requirements for laboratory activities and the number of questions students asked of laboratory assistants.

[31], in his research on whether alternatives to laboratory practicals do meet the desired needs, found that the group that had undertaken simulation based laboratory practicals performed better than the control group. The author concluded that simulations are more effective than the traditional laboratory in helping students achieve assessed objectives, and can provide a learning aid as effective as the physical laboratory. [2] stated that “many qualities and higher level learning goals traditionally gained through [traditional] laboratory-based investigation can be achieved without undertaking any hands-on practical work”.

According to [32] and [17], it is not clear what factors influence the efficacy of virtual laboratories as pedagogically useful tools for promoting understanding of course concepts.

[17] suggested three possible factors:

- much of the attention and time of students in traditional hands-on laboratories is devoted to understanding the procedures to be followed and to laboratory equipment setup and teardown, which is an inherent part of every laboratory work in the traditional laboratory environment. Consequently they focus less on developing conceptual understanding of how the data and relevant theories/concepts relate. In the virtual laboratory environment, students' attention can be directed towards specific learning objectives.
- students, in the virtual laboratory, more often collect laboratory data individually and have more opportunities to repeat laboratory activities, vary parameters, and observe their effects, and structure their learning. Such affordances are especially valuable in enhancing learning.
- in the virtual laboratory, “students usually run the laboratories individually. This is different from the traditional laboratory situation where the students run laboratories as a group, where often, only one student actually interacts with the equipment, while the others watch (witness). If witnessing the actual physical experiment is the important thing, we would expect the traditional laboratories to result in better learning outcomes. However, if individual interaction and the potential for multiple

runs of the procedure are more important, then both simulations and remote laboratories may have an advantage” [17].

In addition, [33] found in their study that students positive relationships with computers and attitudes to computer use were key factors in the VLE success. According to [34], the interactive nature of VLEs enhances student performance and retention, as well as faculty productivity. The VLE encourages freedom and students often thrive in it [35].

1.3 VIRTUAL LABORATORY IN UNDERGRADUATE ENGINEERING LABORATORY EDUCATION

The challenge facing engineering faculties, with respect to the adoption of virtual laboratories in the undergraduate engineering course, is how to place the real laboratory online [36]. Introducing the laboratory learning environment online is a difficult task but the rapid development of technologies makes it simpler [36]. The rate of adoption of virtual laboratories in the educational context is fuelled and driven by the possibilities and benefits offered by today’s affordable PCs, Object-Oriented and Graphical programming languages, and other technologies. There is a large body of literature on virtual laboratory. [15] and [37]] assert that reports on virtual laboratories are dispersed in more than 100 different journals and conference proceedings because of their wide disciplinary nature, with most concentrated on engineering.

However, the use of virtual laboratories in engineering laboratory education, to undertake curriculum-based laboratory activities, is still a challenge, due to the numerous diverse requirements and experiential needs of engineering curricula. Thus, for various fields of engineering, research is on-going on virtual laboratories that can emulate the traditional laboratory processes. Hence, more virtual laboratory implementation and research efforts, for the UE course, are required. Every implementation depends on the identified needs and application, the basic requirement being that each outcome should be functionally satisfying.

There is need for a virtual laboratory environment that will enable students to undertake curriculum-based laboratory activities, in the undergraduate electronic engineering course, in a realistic manner. The existing virtual laboratories in engineering are mainly interactive simulations or animated software solutions for demonstrating and/or visualizing specific concept(s) and are not intended for use to undertake varying instructor designed curriculum-based laboratory activities; or are implemented using commercial software tools that may be

expensive to license. Also, to the best of the author's knowledge, non of the existing engineering oriented virtual laboratories integrates an assessment tool for the assessment of students' laboratory work performance in the virtual laboratory environment.

1.4 LABORATORY WORK ASSESSMENT

The learning outcomes derivable from laboratory activities are only measurable through assessment. Learning is often assessed to provide a measure of students' performance as an index of the effectiveness of both the instruction and learning processes. Effective instruction and learning lead to improved performance [38]. This highlights a close association between instruction, learning, and assessment [39]. Without assessment, there is no quantitative or qualitative measure of learning [40]. Assessment generates data for information about the learning that has or has not taken place, relative to specific objectives and goals [40] [41]. The roles of assessment in instruction and learning have been articulated by [42] and include:

- direct students' efforts towards key aspects of a course unit
- engage students in learning activities appropriate to the unit
- reveal students' strengths and weaknesses
- facilitate feedback to students, for improvement.
- compare students' works based on some criterion level (e.g. a pass-mark) or with a student's previous work
- other stakeholders appreciate what standard a given student has attained
- instructors and administrators to evaluate the effectiveness of the course unit and improve it where necessary

The assessment of student's laboratory work performance is another challenge faced by engineering faculties generally, and particularly by faculties additionally challenged by resource constraints and/or large class sizes. The assessment requirements for students' laboratory work in the virtual laboratory environment are essentially the same as in the traditional laboratory setting, in terms of fairness, reliability, and validity [43]. In the virtual learning environment, assessment is no less critical [44]. This highlights the need for laboratory work performance assessment models for the virtual laboratory environment.

1.5 MOTIVATION AND OBJECTIVES OF THE RESEARCH WORK

The major motivating factors for the work presented in this thesis are the ongoing challenges faced by engineering faculties, with respect to the:

- provision of meaningful laboratory experiences for students, in view of the typical high costs of the traditional engineering laboratory and/or large class sizes
- the performance assessment of students' laboratory work

The critical role of laboratory work in the Undergraduate Engineering (UE) course makes these challenges ones of utmost importance. These challenges, together, inform the aim (goal) of the work reported in this thesis. The aim is to provide the instructor with a tool which presents a realistic laboratory environment (albeit virtual) to facilitate the increased use of laboratory activities to enhance students' understanding of concepts taught in lectures, in order to lay a more solid foundation for the students. This aim includes the provision of an assessment tool, for the instructor, to facilitate the timely, consistent, and fair performance assessment of students' laboratory work within the context of the virtual laboratory environment.

Consequently, objectives of the work reported in this thesis are:

- implementation of a low-cost Virtual Electronic Laboratory (VEL).
- evaluation of the VEL as a tool for enhancing students' understanding of concepts taught in lectures.
- construction of a Bayesian network-based model for the performance assessment of students' laboratory work, in the VEL environment, based on the students' "observed" behaviour.
- empirical investigation of the optimal approach for the construction of a Bayesian network-based model for the performance assessment of students' laboratory work.
- verification and evaluation of the assessment model within the context of the VEL environment, in order to demonstrate:
 - ★ its effectiveness as a tool for the performance assessment of students' laboratory work
 - ★ the practicability of the performance assessment of students' laboratory work based on performance data extracted from their "observed" behaviour.

These objectives collectively lead to the realisation of the aim of the research work.

1.6 RESEARCH METHODOLOGY

The methodology adopted for the work presented herewith could be classified as action research, which is often employed when proposing a new approach for grafting on to or replacement of an existing approach [44]. Action research is neither a method nor a technique but applied research carried out because of the identified need for change or improvement [44]. The aim of the action research methodology is to arrive at recommendations for improved practice [46]. Action research is a progressive problem solving led by an individual(s) working with others in a team, in order to propose a new course of action or strategy aimed at improving an existing work practice(s). The research work is undertaken in two phases: implementation and evaluation.

1.6.1 Implementation Phase

The implementation of the VEL will entail its functional specification and GUI design. The construction of the assessment tool will necessitate the committed participation of domain experts. Cognitive Psychologists will also be involved in the construction process, during cognitive task analysis (discussed in Chapter 2, section 2.8).

1.6.2 Evaluation Phase

The evaluation of the entire system is in two stages: evaluation of the VEL with student cohorts; and the evaluation of the assessment tool using students' behaviour logs to be generated from the evaluation of the VEL. The VEL will be evaluated at two different engineering faculties in two separate universities. Over 60 students from both engineering faculties will participate in the evaluation study. Domain experts will assist with the design of the laboratory activities to be used for the evaluation of the VEL. The activities will be designed to address concepts the students have already been taught in lectures. Pre-Test and Post-Test (PTPT) observation technique will be used to generate data for statistical and concentration analysis, while a survey instrument will be used to elicit students' feedback about the VEL.

The evaluation of the assessment tool (referred to in this context as the Laboratory Performance (LAP) model) will involve the participation of a team of assessors, domain experts different from the ones that will help to create the model and design the laboratory activities. The assessors will also be different from the instructors that taught the students the concepts to be addressed by the laboratory activities. The assessors will assess students'

laboratory work using performance data to be extracted from students' behaviour logs that will be generated during the evaluation of the VEL. An events tracker/recorder agent in the VEL environment will unobstructively observe, capture, and log students' behaviour as they undertake laboratory activities in the VEL environment. Hence, the evaluation of the LAP model will be based on the *virtual observation* method, whereby a software agent undertakes the observation of students and the recording of their laboratory work behaviour, in place of human assessors. The assessors, rather than physically observe the students themselves (*physical observation technique* which is time, space, and logistically demanding and subjective), will use the observations (virtual observation) that will be made and recorded by the software agent in the VEL environment to assess and score students' laboratory work performance, based on a set of criteria. The assessors will each assess the same laboratory activities for each of the students that will participated in the evaluation of the VEL.

1.7 OVERVIEW OF THE THESIS

The background, motivation, and objectives of the work presented in this thesis have been detailed. The remainder of the thesis is organized as follows:

Chapter 2 presents the review of literature for related work on virtual laboratories in the UE course, and intelligent assessment in VLEs. The chapter also detailed the traditional laboratory work assessment practices and their inherent problems, before giving an overview of performance-based assessment. In addition, details on Bayesian networks, their underlying theories, and associated concepts are also given, before highlighting the role of psychology in the research work, and summarising the chapter.

Chapter 3 gives a description of the VEL, and its evaluation processes and results.

Chapter 4 highlights the structural realization and parameterization of the assessment model.

The specific tools and reasoning used in the construction of the model are described, together with the knowledge elicitation processes and model calibration technique.

Chapter 5 details empirical investigation of an optimal approach for construction of Bayesian network-based model for the performance assessment of students' laboratory work.

Chapter 6 illustrates the performance data extraction and evidence quantization and instantiation processes which encompass the analysis of students' laboratory work behaviour logs.

Chapter 7 presents verification and evaluation of the assessment model within the context of the VEL, and its possible application processes.

Chapter 8 summarises the thesis in terms of discussion, limitations, contributions, further work, and conclusions.

1.8 SUMMARY

The background, motivation, and objectives of the work presented in this thesis have been given, and the organization of the thesis highlighted. The next chapter presents a review of literature on the various concepts that are key to the work undertaken and presented in the thesis.



CHAPTER 2

LITERATURE REVIEW

*Twice and thrice over, as they say, good is it to repeat and review
what is good.*

Plato

347BC - 427BC

2.0 INTRODUCTION

It is appropriate to review the literature on relevant and related works in order to place the work presented in this thesis into context. In this chapter, the broad areas of virtual laboratories and intelligent assessment are reviewed, and underlying key concepts detailed, in addition to the basics of Bayesian reasoning and networks.

First, in section 2.1, related works on virtual laboratories are reviewed. Next, the traditional laboratory work performance assessment practices (often extended to the virtual laboratory environment) and their inherent problems are detailed in sections 2.2 and 2.3, respectively. Section 2.4 presents the concept of performance-based assessment. A review of literature on intelligent performance assessment models in the VLE is presented in section 2.5, while section 2.6 details the Bayesian Network (BN) AI technique and its underlying theories and concepts, highlighting its inference process and advantages. A review of behaviour log analysis techniques for performance data extraction, for input into the BN-based model, is presented in section 2.7, before discussing the Cognitive Task Analysis (CTA) framework in section 2.8. The chapter is summarised in section 2.9.

2.1 RELATED WORK ON VIRTUAL LABORATORY

A review of virtual laboratory implementations that have contextual relevance to this work is presented in this section. These include: Virtual Workbench for Electronic Instrumentation Teaching (VIEW) by [47]; electromagnetic fields virtual laboratory by [48]; power system dynamics and control virtual laboratory by [49]; Fourier Transform Tool by [50]; Digital Signal Processing Soft Laboratory (DSPL) by [50]; microelectronics technologies virtual laboratory by [52], and DigiLab by [53]. Others are the virtual laboratories presented by Zhou & Lo [54], and [55].

VIEW [47], for teaching Analogue-to-Digital Conversion (ADC), DSP, digital filters and Fourier transforms, was developed using a combination of tools: LabVIEW, Matlab, Multisim, Spice, and Electronic Work Bench (EWB). The virtual laboratory by [48], for teaching electromagnetic fields, was based on Matlab, while the one by [49] was realized by obtaining a customized computer application for laboratory topics, using Matlab and Simulink. The virtual laboratory by [50], for teaching Fourier transform, was implemented with LabVIEW. The DSPL by [51] was created with VC ++ and Matlab mixed programming. The microelectronics virtual laboratory by [52] serves as a subsidiary learning tool by

demonstrating how microelectronic technologies work. It is based on HyperText Markup Language (HTML), Cascading Style Sheets (CSS), PHP (recursive acronym for PHP: Hypertext Preprocessor), My Structured Query Language (MySQL) and Macromedia flash technologies. The DigiLab by [53] for demonstrating the basic concepts of Digital Logic Design (DLD) is based on Java and Structured Query Language (SQL). The interface is designed with HTML as the programming language with animation and graphic display. The virtual laboratory by [54], like that of [48], is a set of java simulation applets used for enhancing students' understanding of the concepts taught in a course unit on random signals and noise.

The Ngee Ann Polytechnic virtual laboratory [56] is only used to measure the resistance of and the voltage across a resistor placed on a virtual breadboard. Resistors are the only components seen to be provided in the virtual laboratory environment. The Universal Virtual Laboratory (UVL) by [55], specifically focused on people with motor disability (especially upper extremities), also made intensive use of commercial software tools including Matlab, Simulink, PSpice, LabVIEW, Multisim, Authorware, and ORCAD.

In addition to the reports in literature, numerous virtual laboratory websites exist, a few of which are highlighted from the list given by [57].

- <http://www.ipes.ethz.ch>

The interactive Power Electronics Seminar (iPES) site offers interactive tools for demonstrating the theory of three phase systems, DC-AC converters, and fundamentals of electric and magnetic fields, among others, based HTML and Java applets.

- <http://www.isep.pw.edu.pl/ICG/vlab/>

The virtual laboratory for power electronics site offers interactive animation tools for demonstrating Pulse Width Modulation (PWM) techniques, based on java applets. The theory is summarized on an HTML surface.

- <http://www.sia.co.jp/~icass/index.html>

The Interactive Circuits and Systems Seminar (ICSS) site offers Java applet-based interactive animations for explaining the current and voltage sharing among resistors connected in parallel and in series and the behaviour of resistors, inductors and

capacitors in single phase networks, and the basic series and parallel RL, RC and RLC circuits.

- <http://thayer.dartmouth.edu/other/3Dcircuits/animations/index.shtml> The Electronic Circuits Animation (ECA) site contains animations for the visualization of voltage, current, and flux with respect to electronic circuit components (such as op-amp, comparator, and Schmitt-trigger circuits) and power system circuit components (such as diode, bridge, and half-wave rectifiers).
- <http://tams-www.informatik.uni-hamburg.de/applets/hades/html/> The HAmBurg DEsign System (HADES) is another virtual laboratory site. HADES is a Java-based framework for simulation which presents an editor to create digital circuit schematics. It provides about 200 simulation components for digital circuit.

These existing virtual laboratories are:

- either java applet-based animations for the demonstration of specific concept(s) or based on commercial software tools (such as Matlab, Simulink, PSpice, LabVIEW, Multisim and Authorware, ORCAD) that may be expensive to licence and therefore may not be ideal for engineering faculties challenged by resource constraints.
- not designed to undertake instructor-designed curriculum-based laboratory activities.
- not available for the whole spectrum of engineering courses.

Furthermore, they do not integrate any form of assessment tool or mechanism for:

- the automatic generation of laboratory activity-related graphs, by students, in the virtual laboratory environment.
- the performance assessment of students' laboratory work in the virtual laboratory environment. As stated earlier, in the virtual learning environment, assessment is no less critical [44]. Generally, the tendency is to extend the laboratory work assessment practices in the traditional laboratory setting to the virtual laboratory environment [58].

The work reported in this thesis, undertaken as a contribution towards addressing these identified research gaps, encompass the design and implementation of a low virtual electronic laboratory, for the field of electronic engineering, and the construction of a Bayesian Network-based assessment tool for the performance assessment of students' laboratory work, within the context of the virtual electronic laboratory environment. The virtual electronic laboratory, designed for undertaking instructor-designed curriculum-based laboratory

activities, incorporates a graphics tool for the automatic generation of laboratory activity-related graphs, by students.

2.2 TRADITIONAL LABORATORY WORK ASSESSMENT PRACTICES

Engineering faculties, as institutional divisions, often require students' to keep records of laboratory activities using journals and/or logbooks, and in some cases to produce written reports. Assessment of students' laboratory work is generally based on marking these pieces of written evidence and/or students' responses to questions relating to the laboratory activities. What and how it is assessed varies among faculties. [59] reported that, in their faculty, students' laboratory journals are inspected but not assessed. Assessment for each laboratory activity is based on marking a set of questions constituting part of the laboratory instruction sheet. [60] reported a similar assessment practice where, though students keep journals of all laboratory activities, assessment is based on marking a set of questions related to the performance and results of the laboratory work. [61] use the logbook as the primary mechanism for the assessment of students' laboratory work but not all laboratory activities are assessed. A set of randomly selected activities (which may be different for each student), are marked (students are encouraged to assume that every laboratory activity can potentially contribute to their final mark).

A similar practice is reported by [62], but with a different marking method. First, a student's logbook is marked in consultation with the student who is then given a chance to revise the logbook based on any adverse feedback received. The revised logbook is marked. The final mark for the laboratory work is made up of the mean of the marks assigned to the original logbook, the revised log book, and two written reports. A student is only allowed to progress to the next set of laboratory activities if he/she achieves a mean mark $\geq 50\%$. Otherwise, the student takes a practical examination which the student must pass in order to progress to the next stage. [63] reported an assessment scheme that is based on logbook and the *physical observation* of students as they undertake laboratory activities. The assessment by physical observation is based on a set of performance factors (shown in Table 2.1). Each factor has a set of criteria to be used in awarding marks. For example, "Intelligent approach" factor has the following four marking criteria: *original*, *full of ideas*, *questioning each step*, and *not contributing at all (no ideas)* [3]. All the assessment is made and completed in the presence of the students and the marks from the logbook and the physical observation contribute equally to the final mark. [64] reported an assessment scheme in which the final mark is made up of

marks from written reports, portfolios, and individual laboratory examinations. [65] proposed a scheme involving team-based approach to writing laboratory reports by students, using a journal-style format, whereby the team is assessed as a group and share the mark on the report.

The traditional laboratory work assessment practices seem to embody some limitations, with possibilities of inconsistency, bias and unfairness. For example, the assessment scheme of [63] puts emphasis on attitude, qualities, and other factors that neither reflect the objectives of laboratory work nor allow them to occur [3].

Table 2.1: Performance Factors for a Physical Observation technique
(Source [3])

Student	Experiment					
		Poor	Doubtful	Satisfactory	Good	Excellent
Performance	Total	0	1	2	3	4
Planning and preparation						
Understanding of expt. And background						
Intelligent approach						
Application and effort						
Rate of progress and progress made						
Lab. Log.	Total	0	1	2	3	4
Sectioning and layout (ease of access)						
Data presentation and handling (tables and graphs)						
Experimental notes and observations						
Treatment of uncertainties						
Presentation of results and conclusions						
Grand Total	Date					

2.3 PROBLEMS OF TRADITIONAL LABORATORY WORK ASSESSMENT PRACTICES

Laboratory work assessment practices based only on written evidence has been criticized [26]. The criticism is not necessarily indicative of the questionability of the value of students' written reports (the importance of technical communication and the maintenance of proper records are not underestimated [3], rather, it emphasises the need for the assessment of students' laboratory work from a holistic approach. "The requirements of formal laboratory reports, their structure, and the assessment methodology, are often ambiguous" [66]. Moreover, assessment of laboratory reports is challenging/demanding, especially where there are large class sizes [67] [68]. The time requirement is typically in excess of 33 working hours per week for class sizes of around 200 students [68]. It is therefore, difficult to sustain this

approach because of increasing student numbers. Moreover, students often manipulate laboratory reports [26], resulting in inconclusive assessment of the students' work. Large class sizes pose the additional problems of fair and consistent assessment.

An important flaw of assessment of practical work based only on written evidence is that it focuses on assessing only the end product, neglecting the work process and the specific practical abilities/skills (competencies) which are part of the learning outcomes of engineering education [2]. The assessment method is not in line with the principles of the prevailing educational paradigm which demands that students demonstrate their abilities/skills for laboratory work, and their knowledge and understanding of concepts addressed by the laboratory activities. It is therefore imperative that students' laboratory work assessment practices should change, not only in accordance with the prevailing educational theories, but in response to statements of criteria for engineering education by Quality Assurance Agency (QAA) [11] and Accreditation Board for Engineering and Technology (ABET) [12]. This need for change towards a holistically based assessment scheme can possibly be met through performance-based assessment which is discussed below.

2.4 PERFORMANCE-BASED ASSESSMENT

"Performance [-based] assessment is the direct, systematic observation of an actual student performance and the rating of that performance according to previously established performance criteria" [69]. Performance assessment entails that students are assessed on both the process and the end result of their work [69]. "Performance [-based] assessment calls upon students to demonstrate specific skills and competencies and to apply the skills and knowledge they have mastered" [70]. It highlights what a student knows and can actually do [70]. It is said to be valid in terms of consequence, impartiality, content coverage, cognitive complexity, significance, judgment, cost, and efficiency [71]. Its outcome provides information on what students can actually do [72]. Performance-based assessment has been demonstrated as an effective way of evaluating students' laboratory skills [73].

Efforts at performance-based assessment, in the traditional laboratory setting, used the *physical observation* technique based on a list of performance criteria [63] [74]. [75] proposed the use of a survey instrument to assess students' practical-intelligences (knowledge and skills) gained from engineering laboratory activities. The instrument consists of a set of domain specific problems/tasks, such as wire stripping. For each problem/task, a set of 10 to

20 response items, each describing (in words, pictures, and/or illustrations) a possible solution approach to the problem/task is constructed. Students are required to rate the appropriateness of each response item to the problem/task, on a 7-point Likert scale, and the students' responses are scored. This assessment scheme is not performance-based and the authors acknowledged that constructing the assessment instrument is challenging.

The principles of Objective Structured Clinical Examination (OSCE) have also been proposed, by [76], for the assessment of engineering students' laboratory work performance. The OSCE, originated by [77], is a widely adopted physical observation-based performance assessment tool in healthcare education. [76] proposed Objective Structured Technical Examination (OSTE) as an engineering version of the OSCE. The assessment scheme is space, time, and assessor intensive, logistically demanding, and difficult to apply to large class sizes, unlike the *virtual observation* technique proposed in this thesis as an alternative to the *written evidence* and *physical observation* assessment techniques, for the virtual laboratory environment, using a Bayesian network-based assessment model.

Virtual behavioural observation is the collection and logging of behavioural data while subjects are engaged in activities in an interactive virtual environment [78]. It is a common tool of user modelling, an activity that attempts to describe the users' behavioural patterns, preferences, interests, and characteristics, for providing personalized or adaptable web applications [78]. It is also used in the field of Intelligent/Adaptive Tutoring System (ITS/ATS) for student modelling (describing students' behavioural patterns, preferences, interests, and characteristics), in order to provide personalized or adaptable teaching. Virtual observation has also been used in other areas of study and applications including examination of moon phases [79], clinical studies [80], and teachers' professional development experience [81]. [79] found virtual observations more effective than direct physical observations and acknowledged that it produced better data.

2.5 RELATED WORK ON INTELLIGENT PERFORMANCE ASSESSMENT IN THE VIRTUAL LEARNING ENVIRONMENT

Intelligent assessment consists of: assessment scenarios, AI, and the use of a cognitively-grounded measurement model [82]. The assessment scenario consists of context specific tasks, the performance of which will be assessed by automated routines that emulate the behaviour of an expert (the model-based measurement tool, based on AI technique), to

provide an assessment outcome that impact instructionally useful information [82]. Educational assessment has been subject to numerous studies. The emergence of Bayesian reasoning AI technique as a powerful probabilistic inference tool catalyzed research on intelligent assessment in the educational context. [83] presented a preliminary study of the progress of research on educational assessment using BNs, highlighting the different areas of application of BN-based models, in the educational assessment context. [84] reported the construction of a model for estimating an educational institution's performance, with respect to a specific subject. The aim was to develop a diagnostic model of an institution's effectiveness or quality. They identified and modelled the causal relationship between factors affecting the effectiveness of an educational system, such as instructors' knowledge and experience, class size, student and instructor attendance, among others. [85] presented a model for evaluating students learning styles based on the Index of Learning Styles by [86]. Data recorded from the observation of a student's learning behaviour in web-based courses are used to automatically detect the student's learning style. [87] presented a model for assessing students' grasp of a particular concept. The model assesses students' current state in an open/distance informatics course based on such factors as: use of computers, web-based material, textbook, and internet, family status, and hours of work. The model is aimed at predicting whether a student is likely to dropout or continue with a course.

Other related works that demonstrate the feasibility and benefits of performance assessment of students learning in the virtual learning environment (VLE), using BN AI technique, are hereby cited. The design of a BN-based model for the prediction of students' performance in a test and the application of a BN-based model in educational computerized adaptive testing to assess students' performance were reported by [88] and [89], respectively. [90] applied BN in adaptive testing of multiple student latent traits in a single item-based test, using granularity hierarchies. Bayesian inference is used to propagate knowledge over the hierarchies. Each trait is either mastered or not mastered. [91] presented a BN- and rule-based model for the assessment of students' learning, using students' knowledge map and analysis of their responses to tests items. The model was designed for the assessment of students' software usage abilities and skills in a computer course. [92] reported a BN-based model of student activity performance for assessing students' learning processes in a VLE, from the analysis of the students' web portfolios. Assessment is based on such variables as self reflection, homework, read frequency, login frequency, and peer interactions.

These intelligent assessment models, though BN-based, are not aimed at students' laboratory work performance assessment. That is, literature search has highlighted previous work on performance assessment in virtual learning environments but has not identified performance-based assessment in the context of virtual laboratories generally and virtual electronic laboratories specifically. Furthermore, most of these existing BN-based assessment tools, though designed for VLEs, are focused mostly on the ITS/ATS environment, the common and initial area of application of intelligent assessment models. ITS/ATS are software tools for teaching, aimed at achieving a one-to-one tutoring process, without the instructor. The output of the assessment model, in an ITS/ATS, is often used to determine what instructional material or assessment item to give the learner next or what concept to teach next [93]. Also, some of the existing models are focused on the assessment of such constructs as learning style, effectiveness, quality, grasp of concept, and learning, while others are focused on item-based tests assessment.

On-Line Assessment of Expertise (OLAE) presented by [93] and [94] is another BN-based performance assessment model designed for the assessment of students' knowledge of Newtonian mechanics in a VLE. OLAE maintains a list of the correct and incorrect rules for solving problems in the domain, against which it matches a student's problem solving solution steps, using a BN and a solution graph, in order to assess the set of rules that have been mastered by a student. OLAE does not assess students' abilities to undertake laboratory activities or analyze experimental results [93]. Also, it "does not monitor laboratory tasks, large-scale projects or their [students'] hands-on activities" [95]. The BN-based assessment model by [96] designed for a VLE consisting of an ITS and a robotics virtual laboratory, was used in the ITS component of the VLE to assess students' knowledge of course themes.

A BN-based assessment tool, which is part of the work presented in this thesis, was constructed for the performance assessment of students' laboratory work in the VEL environment. The next section describes Bayesian networks and their underlying theory and associated concepts, highlighting the advantages of BNs over other AI techniques.

2.6 BAYESIAN NETWORK ARTIFICIAL INTELLIGENCE TECHNIQUE

A BN consists of two component parts:

- *network structure* (the qualitative part of a BN): a set of random variables (nodes) and a set of directed edges interconnecting the nodes without creating directed loops, so

that the nodes, together with the edges, form a Directed Acyclic Graph (DAG). The network structure defines the probabilistic relations among the variables.

- *parameters* (the quantitative part of a BN): the conditional and prior probability distributions of the dependent nodes (child nodes – nodes with parents) and independent nodes (nodes without parents), respectively, in the network.

The building of a BN model involves the following three ordered tasks: identification of the variables (nodes) and their possible values (states), identification of the relationships between the variables, and obtainment of the parameters.

Each node in a BN represents a random variable or hypothesis and each directed edge represents a relationship (dependency) between the two linked nodes, thereby creating a parent→child relationship. Though the term, parent→child relationship, is generally viewed from a biological perspective, it is conventionally used in BN literature to represent relationship(s) between a dependent variable (the child node) and the variable(s) on which it depends (the parent node(s)). In the context of BN models, a child node (a dependent variable) can have many parents (the variables on which it depends). Node *A*, for instance, is a parent of a node *B*, the child node, if there is a directed edge from *A* to *B* (see Figure 2.1).

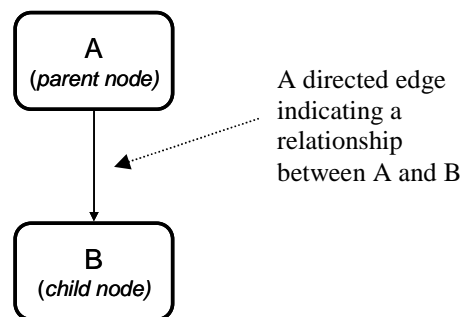


Figure 2.1: A sample BN model: parent-child relationship

Often, semantics are used to give the edges and their directions, particular meanings, reflecting the relationship modelled by the edges. The commonest semantic is causality, which implies an asymmetric cause-effect relationship. Causality is not however, a requirement for BNs [97]. The edges in a BN can represent other relationships, such as containment, ownership, part, requirement, or any relationship that has meaning with the context of the focus domain. For example, [98], defined the edges in their BN model as representing the relationship of “skill → sub-skill” where the parent node represents a skill and the child node represents a sub-skill (component) of the parent skill.

A node can have from zero to a finite number of parents, which interact to produce a combined effect on the child node. All child nodes (nodes with parents) have Conditional Probability Tables (CPTs) that model the combined impact of their parents, while each leaf node (node without parents) has Prior Probability Tables (PPTs) describing the prior knowledge about them. The number of possible values of a node and the number of its parents impact on the size of its CPT. There is, an exponential relationship between the size of the CPT of a node, the number of its parent nodes, and the number of possible states of its parents. For example, if a node, A , has k parents and each parent has two possible states, then the CPT for node A will contain, 2^k entries ($k = 2$, $k = 8$, and $k = 16$, will result in 4, 256, and 65536 entries respectively). If each of the parents has three possible values, 3^k entries are required ($k = 2$, $k = 8$, and $k = 16$, will result in 9, 6561, and 43046721 entries respectively). Generally, the number of entries in the CPT of a child node, A , that has k number of parents, where each parent has n possible states, is n^k .

Each discrete variable in a BN can take a finite set of values and can assume only one of its values at any one time, and the sum of the probabilities of all the states should be 1. In this way a BN model is subject to the standard axioms of probability theory [99]. The basic axioms of probability theory, Bayes theorem, and the concept of Joint Probability Distribution (JPD) underpin BNs. Other important concepts associated with BNs are conditional independence, d-separation, marginalization, and inference.

2.6.1 Probability Theory, Bayes Theorem, and Joint Probability Distribution

The basic axioms of probability theory states that, for events, A and B :

1. $P(A) = 1$ iff A is certainly true. If A is a multinomial variable with the set of possible states $a = \{a_1, a_2, \dots, a_i, \dots, a_n\}$, the probability that A is in state a_i is denoted $P(A = a_i)$. The sequence of probabilities $P(A = a_1), P(A = a_2), \dots, P(A = a_i), \dots, P(A = a_n)$ define a probability vector simply written as $P(A)$, and $\sum_{i=1}^n P(A = a_i) = 1$.
2. $P(A) = 0$ iff A is certainly false.
3. $0 \leq P(A) \leq 1$, the numeric bound of $P(A)$

4. If A and B are mutually exclusive (the two events are disjoint and cannot occur together) then, the probability of A or B , $P(A \cup B)$, is $P(A \cup B) = P(A) + P(B)$.
5. If A and B are dependent, then $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, and the probability of A , given B , is $P(A|B) = \frac{P(A, B)}{P(B)}$, while the probability of B , given A , is $P(B|A) = \frac{P(A, B)}{P(A)}$. The *Joint Probability Distribution* (JPD) of A and B , $P(A \cap B)$, also written as $P(A, B)$, is:

$$P(A, B) = P(A|B)P(B) = P(B|A)P(A) \quad (2.1)$$

Equation 2.1 is a basic rule of probability calculus referred to as the *product rule*.

$P(A|B)$ denotes the *conditional probability distribution* of A given B . For example, $P(B=b_i|A=a_i) = r$ represents the statement: if $A=a_i$ is true, and no other information to hand is relevant to B , then the probability of $B=b_i$ is r .

Rearranging the product rule yields *Bayes theorem* given as:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

and

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (2.2)$$

where B represents some hypothesis whose degree of truth is to be determined, A represents new data (evidence) from an observation. $P(B|A)$ is the posterior probability, updated as a result of obtaining the evidence A . $P(B)$ is the prior probability of B , and $P(A|B)$ is the likelihood of B given A [Jensen, 2001] or the likelihood that B will materialize if A is true [Pearl, 1988]. The factor $1/P(A)$ is the normalization or proportionality constant. The ability to update the posterior probability of B is the major strength of Bayes theorem infused into BNs, which can be applied repeatedly as new pieces of information (evidence/findings) are obtained.

The Joint Probability Distribution (JPD) over the set of n variables, $U = (A_1, A_2, A_3, \dots, A_n)$, in a BN, is defined recursively, by applying the product rule repeatedly, which gives the *chain rule* [100]:

$$\begin{aligned}
P(U) &= P(A_1, A_2, A_3, A_4, \dots, A_n) \\
&= P(A_1 | A_2, \dots, A_n) P(A_2 | A_3, \dots, A_n) \dots P(A_{n-1} | A_n) P(A_n) \quad (2.3) \\
&= \prod_{i=1}^n P(A_i | A_{i+1}, \dots, A_n) \\
&= \prod_{i=1}^n P(A_i | pa(A_i)), \text{ where } pa(A_i) \text{ represents the parent set of } A_i.
\end{aligned}$$

The JPD can be used for belief updating in BNs, which amounts to the algorithmic estimation of $P(U)$, in the light of new evidence. The chain rule expression depends on the way the set of variables, $U = (A_i)$, is ordered. Different possible orders give different expressions, but each expression yields the same value for $P(U)$, for the same set of variables and values [101].

2.6.2 Conditional Independence

The concept of Conditional Independence (CI) is used to minimize the size of the CPT of a child node by restricting the number of its parents through a reduction in the number of edges in the model. BN structures encode CI, and the independencies determine which beliefs are updated given evidence. Hence, CI is advantageous for specifying a BN compactly, and reduces the complexity of inference. Figure 2.2 illustrates the CI concept.

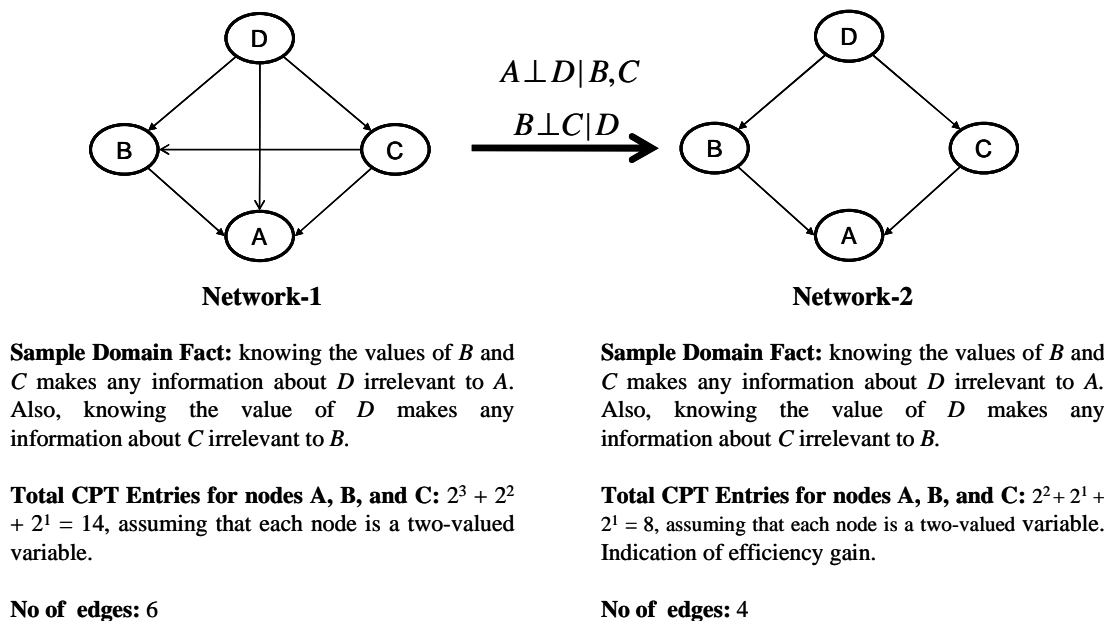


Figure 2.2: Sample BN models illustrating the concept of conditional independence

Figure 2.2 implies, with respect to the sample domain fact, that node A is independent of node D , given nodes B and C , and node B is independent of node C , given node D . Network-1 did not exploit these conditional independencies implied by the domain fact example, to reduce the complexity of the network, while Network-2 did. Reduced model complexity translates to improved computational tractability and gains in storage requirements. Reducing model complexity is an active issue in uncertainty reasoning with BN models.

2.6.3 D-Separation

The term, d-separation, is generally used in BN literature, and the author understands the d in the term to imply dependency. A BN implies more CIs than just those involving the parent nodes. In a BN, there are three different types of dependency topologies (d-connections): serial, converging, and diverging (depicted graphically in Figure 2.3).

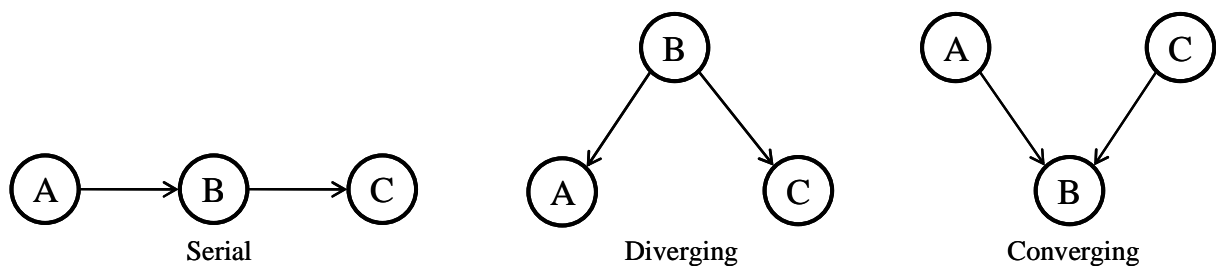


Figure 2.3: Different types of dependency topologies (d-connections)

Serial: Node C is conditionally dependent on node B , and B is conditionally dependent on A .

Entering hard evidence at node A or C will lead to an update in the probability distribution of B . However, if evidence is entered at node B , nodes A and C become conditionally independent given the evidence at node B . This means that evidence at node B “blocks the pipeline” [99].

Diverging: Nodes A and C are conditionally dependent on node B . Entering hard evidence at

node B will update beliefs at nodes A and C . If evidence is entered at node A , it will not cause belief update at node C , when there is evidence at node B . Here nodes A and C are conditionally independent given evidence at node B [99].

Converging: Node B is conditionally dependent on nodes A and C . Entering hard evidence at

node A will update beliefs at node B but will have no effect on node C . Also, entering hard evidence at node C will update beliefs at node B but will have no effect on node A . If hard

evidence is entered at node B, then entering evidence at node A will update node C, and entering evidence at node C will update node A. Hence, nodes A and C are conditionally dependent given evidence at node B [99]. That is, entering evidence at node B opens the otherwise blocked pipeline.

Generally, two variables are *d-separated* if evidence about one cannot influence the other. The paths in serial and diverging connections are blocked if node B is instantiated, so that B effectively d-separates A and C. If A and C converge to B, it transpires that they will be d-separated as long as B and none of B's descendents are observed; if B or any of its descendents is observed, A and C become dependent. This property of converging connections arises because A and C represent multiple explanations of node B. For example, using the analogy given by [101], suppose A is the proposition that student X *has not mastered the topic*, C, the proposition that student X *performs poorly on exams*, and B the proposition that student X *failed the exam*, then A and C are two possible explanations for B and would therefore converge to B in a Bayesian network. If B is subsequently observed, (say, it was found that the proposition represented by node B is true, that is student X did fail the exam), then B's causes, nodes A and C, become dependent. This is because if A is subsequently observed to be true as well, after B had been found to be true, then it has some bearing on C (that is the probability of C is revised downwards) and vice versa. Thus, d-separation characterises independence arising from lack of evidence as well as evidence. Any system for reasoning under uncertainty must capture these properties, as they are basic attributes of reasoning. "By using these ideas (d-separation topologies) we can postulate topologies connecting small numbers of nodes and hypothesise the effects of entering evidence at one node on another. The answer to the question 'would entering data here affect the conclusion reached here, given that we know this datum over here?' might help indicate the type of d-connection at play in the expert's reasoning" [99]. This helps to build a good BN structure. Changing the direction of an edge in a BN (ensuring directed cycle is not induced) may change the d-separation properties of the network, but the overall JPD remains invariant. Hence, BNs differing only in edge directionality are considered equivalent.

2.6.4 Marginalization

While the product rule is used to compute JPDs, marginalisation reduces a JPD to a distribution over a subset of its universe of variables, in order to improve computational tractability. It is used to sum out variables (variable elimination) that are not directly involved in resolving a query. For example, for the sample BN of Figure 2.4, with the JPD,

$P(U) = P(A, B, C, D) = P(A)P(B|A,C)P(C)P(D|C)$, where the state of each variable is either true or false, resolving the query $P(D|C=true)$ by complete enumeration and computation of $P(U)$ would entail summations and multiplications over all the variables.

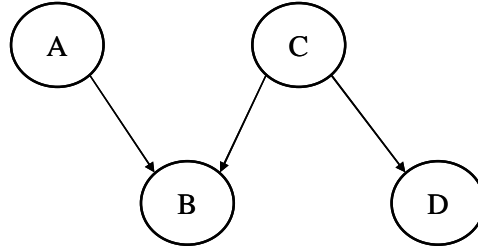


Figure 2.4: A sample BN for explaining marginalization

Alternatively, the query can be resolved by eliminating (summing out) the variables A and B which are not directly involved in the query so that only the *sub*-JPD, $P(D|C)$, is computed instead of the complete JPD, $P(A)P(B|A,C)P(C)P(D|C)$. The marginalization function or rule is generally written as:

$$P(X) = \sum_{U-X} P(U) \quad (2.4)$$

where $P(X)$ represents the joint distribution of the subset of variables directly involved in the resolution of the query, $P(U)$ the complete joint probability distribution of all the variables in the BN, and $U - X$ the subset of variables that are not directly involved in the resolution of the query. \sum_{U-X} means marginalization (summing out) of the subset of variables, $U-X$, out of $P(U)$ to get $P(X)$. Marginalization is used to make query resolution more tractable, which is why it is seen as an inference tool in BNs.

2.6.5 Inference

Once a BN model has been constructed, verified, and evaluated, it can be used for reasoning (inference). Inference in BNs (also referred to as belief/model update and model conditioning) is performed by the “flow of information” through the network, which is not limited to the direction of the edges [103]. Inference is essentially the computation of the posterior probabilities of target/query nodes, given findings/evidence. That is, given a BN model, a piece of evidence $E=e$, and a query variable Q , inference is the computation of $P(Q|E=e)$. Reasoning is about querying and interpreting the appropriate posterior probability of an

unobserved node(s) of interest, given the observed data. The different types of reasoning or queries include [101]:

- *diagnostic query* (finding causes from symptoms). This type of reasoning is mainly used in causal models and occurs in the opposite direction of the causal edges.
- *predictive or inferential query*. This reasoning occurs in the direction of the orientation of the edges in the BN.
- A combination of the above two types of reasoning, if the reasoning task does not neatly fit into any one of the above.

The difference between the above types of reasoning is highlighted using a description by [101] based on Figure 2.5, which depicts a BN model of a student's mastery of domain topics (T_1 and T_2) and concepts (C_1 , C_2 , C_3). The model has the set of variables, $X = (T_1, T_2, C_1, C_2, C_3, Q_1, Q_2, Q_3, Q_4)$, and the JPD

$$P(X) = P(Q_1 | C_1, C_3)P(Q_2 | C_1)P(Q_3 | C_2, C_3)P(Q_4 | C_3)P(C_1 | T_1)P(C_2 | T_1)P(C_3 | T_2)P(T_1)P(T_2).$$

Mastery of the topics implies mastery of the various concepts, which in turn influences performance on the test questions (Q_1, Q_2, Q_3, Q_4). A test question may require the student to have mastered more than one concept, e.g. Q_3 . With this BN, inferential queries, such as $P(Q_1 | T_1 = \text{Mastered})$, and diagnostic queries such as $P(T_2 | Q_3 = \text{failed})$; or a combination of both, such as $P(C_2 | Q_4 = \text{passed}, T_1 = \text{notMastered})$ can be performed. Evidence for evaluating a query or inference is received when an observed node is assigned a value or instantiated (hard evidence).

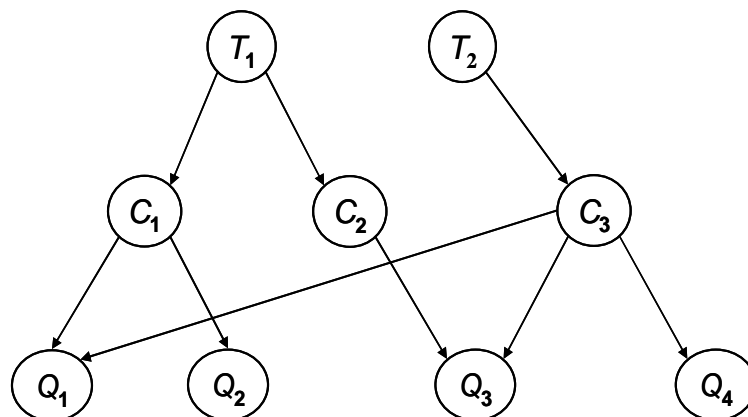


Figure 2.5: Student Topic and Concept Mastery BN model
(Source: [Mayo, 2001])

The two computation methods for evaluating a query are *global* and *local*. To evaluate a query, $P(C1|Q1 = \text{failed})$, for example, the global computation method would involve the computation of the joint distribution, $P(X) = P(Q1|C1, C3)P(Q2|C1)P(Q3|C2, C3)P(Q4|C3)P(C1|T1)P(C2|T1)P(C3|T2)P(T1)P(T2)$, in order to resolve the query. The local computation method would involve the marginalization (elimination) of the variables that are not directly involved in resolving the query so that the query resolution reduces to computing $P(C1|Q1 = \text{failed}) = \frac{P(Q1 = \text{failed} | C1)P(C1)}{P(Q1 = \text{failed})}$. The global computation method

is intractable because even the simplest query would require summations and multiplications over all the variables in the model. Though the two methods give the same inference result, local computation is the more efficient approach because the local inference algorithm takes advantage of the conditional independence relations defined by the model, and the marginalization rule, to increase efficiency. Details of the two computation methods are given by [104] [105].

2.6.6 Advantages of Bayesian Network Modeling

Bayesian networks offer a wide range of advantages over other AI techniques, such as Artificial Neural Networks (ANN) and Fuzzy Logic (FL). These advantages, as highlighted by [99][106][107], include that:

- Bayesian networks enable reasoning under uncertainty and combine the advantages of visual representation with sound mathematical basis.
- expert opinions can be articulated about dependencies between different variables and the impact of evidence on the probabilities of unobserved variables is propagated consistently, in any direction.
- not all the variables in a BN model need to be instantiated for each model update and inference instance. The models works by using the instantiated variables to update belief about the un-instantiated variables and enable inference. This is unlike FL where if a case is missed (a variable is not instantiated), the model would not work or respond properly.
- expert knowledge can be combined with data in a very practical way.
- a large BN can be constructed from various BN fragments which enables BN models to be varied from problem instance to problem instance.

- Bayesian networks are a convenient way to model and solve problems in which it is desired to arrive at conclusions that are not warranted logically but rather probabilistically. This is unlike FL models where each output is warranted logically, which makes it unsuitable for applications in uncertainty modelling, despite that FL applications generate outputs using approximate and ambiguities human languages, based on logic.
- Bayesian networks are robust in that small alterations in the model parameters do not affect the performance of the system dramatically, which implies that maintaining and updating existing models is relatively easy since the functioning of the system changes smoothly as the model parameters are modified.
- BNs facilitate clear semantic interpretation of model nodes, links, and parameters. That is, every edge (link), node (variable), and parameter in a Bayesian network is contextually relevant, and semantically meaningful and understandable to both domain experts and the model constructor. Hence, resulting models the resulting model can be analyzed and explained in understandable terms. This is unlike ANN that presents a “black box” approach with limited interpretability, where only the inputs and outputs are meaningful.
- Bayesian networks do not require extensive training and retraining. The training process for Bayesian networks is taken to be the creation of its Conditional Probability Tables (CPTs), and the retraining process the review and modification of the parameters of the CPTs. Even the learning of the parameters from data does not necessarily require large amounts of data. This is unlike ANN that requires large amounts of training data, and when new data becomes available, the ANN needs to be retrained from the scratch. The training time needed to produce an ANN model is often relatively high.
- Bayesian network models can handle different both continuous and discrete variables at the same time, whereas many alternative model technologies are designed for some single specific types of variables.
- when sufficient amount of data is available, both the structure and the parameters of a Bayesian network model can be learned from data (with or without expert knowledge), based on well established algorithmic techniques.

2.6.7 Disadvantages of Bayesian Network Modeling

While Bayesian network models are a useful way to model expert knowledge for reasoning under certainty, it has some disadvantages, as highlighted below:

- In creating a BN model with expert knowledge, it may be difficult for experts to agree on the nodes (variables) that are important to be included in the BN model, overall structure of the model, and probability distributions for the nodes. Experts are said to rarely agree. That is, it may be more difficult to elicit expert knowledge for expert-dependent BN models. Experts may be challenged to express their knowledge in the form of probability distributions; hence the knowledge elicitation is often an iterative process to ensure that experts are comfortable with the nodes, their states, and interrelationships in the BN model, and subsequently, their their statements about probability distributions for the variables.
- Though Bayesian networks are good for modeling discrete variables, they may not model continuous variables accurately.
- In learning a BN model from data, BN software packages are required. Most existing BN software packages have no or limited ability to deal with continuous data. Each software package may have its own strengths and shortcomings.
- Each Bayesian Network is specifically constructed for a specific or unique purpose/application, such that an existing Bayesian network model may not be applicable to other problem areas. Therefore different problem situations require different Bayesian networks.
- The reliability of a Bayesian network model depends on the reliability of the elicited expert knowledge, which is often said to be subjective judgement or estimation.
- The time requirements for the construction of a Bayesian network model is relatively high, depending on the domain experts, and other factors, including the domain itself (some domains may be easier to model than others), the model constructor (in terms of experience and knowledge), and tools used (software packages, knowledge elicitation tools, model construction techniques, etc), and requires a detailed analysis of the domain.

2.7 BEHAVIOUR LOG ANALYSIS FOR PERFORMANCE DATA EXTRACTION

Behaviour logs are generated as products of User Interface (UI) events which indicate user behaviour with respect to the components that make up an application's UI. UI events are one of the broad spectrums of Human Computer Interface (HCI) events which lend themselves to

the techniques for collecting observational data [108]. They are relatively high frequency events, usually sequential, ordered by time, and atomically detailed, and provide sufficient data for user characterization, classification, and performance assessment, in online environments [108]. The events are limited by the communication window open for interaction between the user and the virtual environment, and constitutes of mainly mouse-click events, with respect to application windows, and key-press events with respect to application input fields.

The grain size of (amount of information in) a behaviour log depends on the context and the actions that make sense in that particular context. A behaviour log of fine grain size affords more opportunities for research [109] [101], while a coarse grain size log may lead to loss of data [109]. However, what is important is that the granularity of the log should fit its intended analyses [109]. Interaction events can be logged at multiple grain sizes to facilitate multiple types and levels of analysis and data extraction [109]. Behaviour logs may be analyzed at multiple levels of abstraction [108]. For instance, one may be interested in analyzing low-level mouse movement and timing information, or one may be more interested in higher-level information regarding the steps taken by users in completing tasks such as placing components and building a circuit or adherence to safety issues in the laboratory.

There is no systematic approach to behaviour log analysis, for information and data extraction [111][112]. Researchers resort to different approaches, with perspective on the particular learning setting, in order to extract the relevant performance data from students' behaviour logs. "UI events provide detailed performance information that can be searched, counted, and analyzed, using automated techniques" [108]. "Performance data" is any information that is useful in measuring (or identifying potential issues affecting) the performance of a student.

Many researchers are not explicit regarding their log analysis and data extraction approaches. Only a few highlighted their adopted approach. [112] indicated the use of java, mySQL, and machine learning technique, to analyze and extract performance data, from a database of interaction logs, for a mathematics VLE. The author stated that the log analyzer was not an automated component of the virtual laboratory. [93][95] stated that, in On-Line Assessment of Expertise (OLAE -- an assessment tool for the problem-solving domain) presented by [95], equations that consist of students' problem solution steps are recorded as they solve problems. The recordings are analysed off-line, using a rule-based system that compares the student's problem solving behaviour step by step to the behaviour of an ideal model that solved the

same problem, to determine the pieces of domain knowledge used by the student. The identified pieces of domain knowledge used by a student are articulated for input into the assessment model. [92] stated that students' learning activities, in the form of behavioural portfolios, are logged and stored in a database. An information retrieval facility, IBM textminer, is used to extract data and information on concepts of portfolios from the logs. The data constitutes input to a BN model that identifies students' learning states.

[108] surveyed different approaches for extracting usability-related information and data, from UI events logs. The adopted behaviour log analysis technique, in this context, is based on ideas from those approaches and is detailed in Chapter 7.

2.8 COGNITIVE TASK ANALYSIS

Most systems can be modelled in terms of *input*, *output* and the *operations* which map the input to the output, introducing terminologies as needed to represent desired operation and data constructs [113]. The design of a system for use in a cognitive task domain often requires the analysis of both the task procedures and the knowledge required for the performance of the task, as part of the *operations* definition, which entails Cognitive Task Analysis (CTA). [114] defined CTA as "the general term used to describe a set of methods and techniques that specify the cognitive structures and processes associated with task performance." That is, CTA is the description of the cognitive abilities/skills needed to perform a task proficiently [115]. It aims to capture information about both the overt observable task performance behaviour and the covert functions behind it to form an integrated whole, in order to facilitate the breakdown of the procedure into discrete identifiable steps [115].

CTA grew out of the need to explicitly identify and take into account the cognitive elements that underlie observable task performance [117]. Hence, it is seen as being rooted in cognitive psychology, and has been increasingly informed by advances in cognitive science. Cognitive psychology is devoted to the study of human cognition (the act or process of knowing, reasoning, and understanding, particularly as it affects learning and behaviour, and examines internal mental processes with respect to learning and task performance). Often, CTA uses a variety of techniques to capture (elicit) from domain experts, a description of the abilities/skills and knowledge required to perform tasks within the context of the domain. The description could then be used to develop or construct the intended system [115] [118].

There are different methods of undertaking CTA because CTA has become an important component for the design of systems in many domains [118]. For example, the CTA framework was used to develop training programs in the Air Traffic Control domain [119]. CTA has also proven useful in generating models of cognitive processes and human-computer interaction in a number of domains, including [119]: en-route air traffic control [120], air antisubmarine warfare [121], and naval command and control [122].

The CTA framework is relevant in this context. It is used to support the construction of the BN-based intelligent assessment model (detailed in Chapter 4), for the performance assessment of students' laboratory activities in the VEL environment. This is because, laboratory work, like most learning tasks, is limited by the cognitive function. Individuals come to a given cognitive task with differing backgrounds of internal traits/states, including abilities/skills, resulting from learning which support and determine the limit of individual performances [121]. Internal traits/states combine to produce instances of observable behaviour [124]. "Teaching, learning and assessment are all human behaviours and part of the natural subject matter of psychology" [125]. The next chapter (Chapter 3) describes the VEL and its evaluation processes and results.

2.9 SUMMARY

There is a large body of work on virtual laboratories for the UE course. A review of literature on contextually relevant virtual laboratories has been presented. The need for more implementation and research efforts on virtual laboratories that can emulate the traditional engineering laboratory has been highlighted. In addition, literature highlighted the non-existence of assessment tools for the performance assessment of students' laboratory work in the virtual laboratory environment. These identified needs and gaps in literature constitute the basis of the work presented in this thesis.

Performance-based assessment of students' laboratory work, based on *virtual-observation*, using a Bayesian network-based assessment model, is proposed in this thesis as an alternative to the *written evidence* and *physical observation* assessment techniques, for the virtual laboratory environment. The behaviour log is analyzed and performance data extracted for input into the assessment model. The strengths of Bayesian networks are harnessed to make inferences about students' competencies and performances from their observed behaviour. A demonstration of this proposed performance-based assessment scheme, in a VEL

environment, is part of the work presented in this thesis. The next chapter gives a description of the VEL (implemented and evaluated as part of this work), and presents the results of its evaluation in two different engineering faculties.



CHAPTER 3

THE VIRTUAL ELECTRONIC LABORATORY: DESCRIPTION AND EVALUATION

*A first-rate laboratory is one in which mediocre scientists
can produce outstanding work.*

Patrick M.S. Blackett

1897 - 1974

3.0 INTRODUCTION

The Virtual Electronic Laboratory (VEL) is aimed at facilitating the increased use of laboratory activities in the instructional process, in order to enhance students' understanding of taught concepts. In resource constrained engineering faculties, the VEL could provide complimentary or alternative means to deliver meaningful and effective laboratory education. The expected benefits of the system include availability on a one-to-one basis, reduced laboratory preparation time, and meaningful laboratory learning experiences for students. The requirements specification, descriptions, and the evaluation processes and results, of the VEL, are given in this chapter.

First, in section 3.1, the issues put into consideration, in the design and implementation of the VEL, are outlined. Next, a description of the VEL is given in section 3.2. Section 3.3 describes the evaluation of the VEL in terms of the methodology and context, sample student cohorts, and the process, while section 3.4 presents the data analysis. Section 3.5 highlights possible individual student advantage in the VEL environment in terms of learning styles, and students' feedbacks are presented in section 3.6, before summarizing the chapter in section 3.7.

3.1 DESIGN AND IMPLEMENTATION CONSIDERATIONS

The design and implementation of a system often necessitates the highlight of requirement specifications, based on the objectives, which are driven by the identified need or problem situation. In this context, the system requirements are:

- **Low-cost:** The virtual laboratory needs to be low-cost in order to be suitable for deployment in resource constrained engineering faculties. Hence, the use of commercial software packages that may be expensive to licence is avoided. Rather, freeware software tools are used.
- **Portability:** Portability is a software codebase feature that enables the reuse of existing code instead of having to create new code when moving the software from an platform to another. This necessitates a generalized abstraction between the application logic and hardware interfaces. The VEL needs to be portable so that platform will not be a deployment issue. Java technologies are well suited for addressing this requirement.
- **Concurrent Client/server environment:** The VEL should present a concurrent client/server laboratory environment where the client, the GUI, will not reside on the user machine to avoid making demands on users' resources. The client/server

communication connection should be persistent to eliminate unnecessary server overhead. The VEL will be implemented using Java technologies which are well suited for meeting these requirements. Java eliminates most of the problems exhibited by codes developed in other languages, including deceptive interactivity [126].

- Avoidance of user interface clustering and popup windows: This is necessary in making user experience in the VEL environment more intuitive. To this end, the classical and multi-window user interface format is avoided [127]. Rather, an Integrated Graphical User Interface (IGUI) approach is adopted. Popup windows are considered to be annoying [128]. The technique of removing panes that are not immediately in use and replacing them with ones that are immediately in use is resorted to in order to avoid popup windows. A removed pane is replaced when required.
- Non-complex tools and equipment: The equipment in the traditional laboratory are often near-research-grade complex, multi-mode, and multi-function instruments which are often intimidating and difficult for students to understand and use [129]. The instructor ends up providing a laboratory instruction sheet that is more like a “button pushing recipe” that impinges on the limited laboratory time slot [129]. In view of this, given the target students’ level and limited experience, only instruments that are no more complex than is necessary to undertake laboratory activities related to fundamental engineering concepts will be provided in the VEL environment.

3.2 DESCRIPTION

The VEL is based on the client/server programming concept. The client consists of the VEL workbench, designed to run in a browser and provides the user with an interactive GUI with which to realistically construct and simulate a circuit. The components and devices required to build a circuit are laid out on the workbench GUI. The server facilitates user access to the workbench and interfaces between the workbench, and the application software that run on the server, including the circuit simulator and graphic tool. That is, the server acts as the interface between the client and the back-end applications, and facilitates the logging of user behavioural events. Communication between the client and the server is achieved using the Remote Method Invocation (RMI) java technology. Also, java DataBase Connectivity (JDBC) technology is used for communication with the database. The structural overview of the VEL is shown in Figure 3.1, which highlights the two sides of the VEL: the front-end and the back-end.

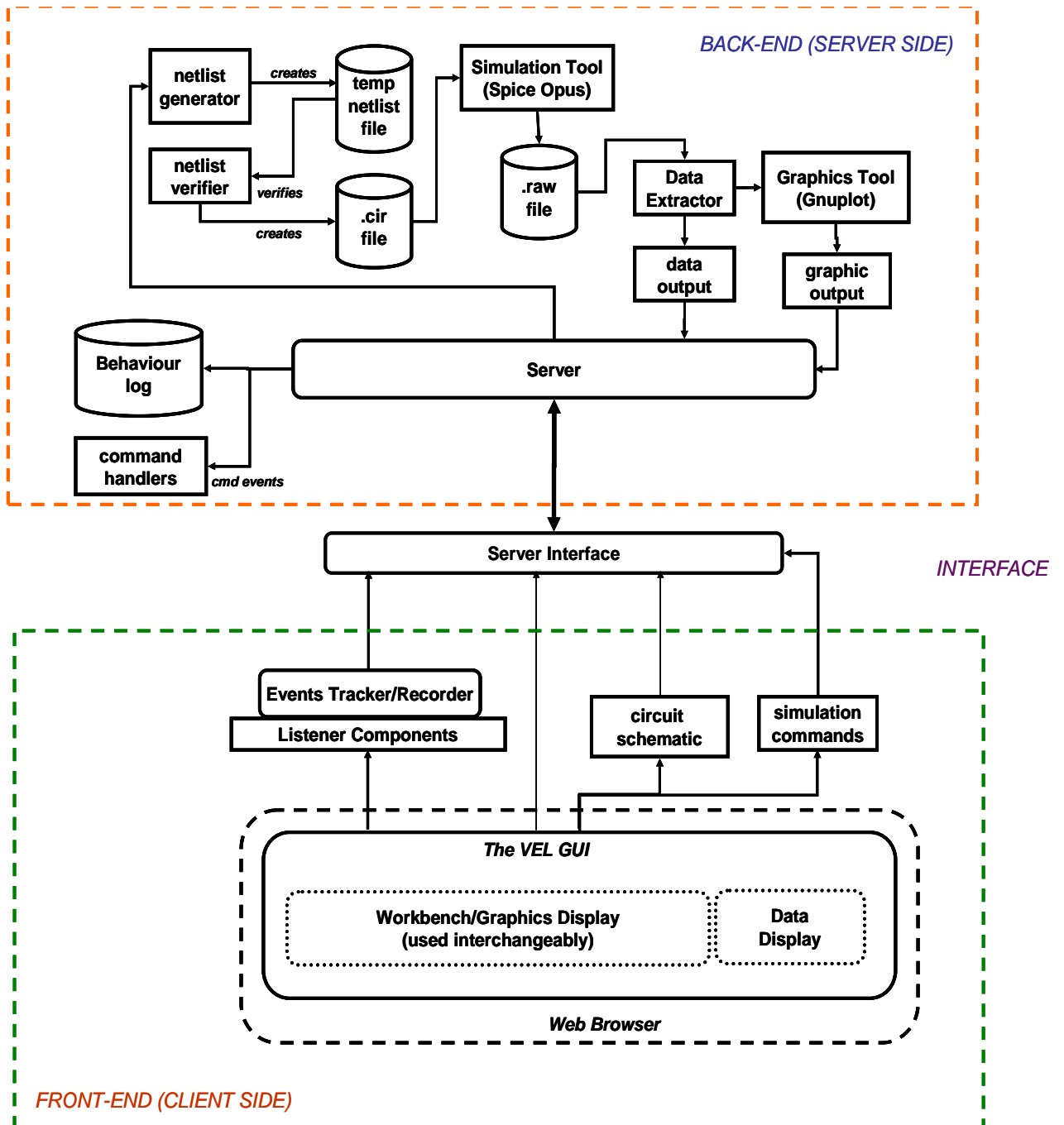


Figure 3.1: Structural overview of the VEL

3.2.1 The Front-end (The client-side)

The virtual workbench front-end, the client application, consists of a java applet that provides a GUI for the user. Figure 3.2 shows the workbench front-end which provides the user with the facilities and features to perform curriculum-based laboratory activities. The breadboard, used for building a circuit, is placed on the workbench, and provides both vertical and horizontal grids, with power and ground rails. The GUI has a number of panels including: Components, Tools, Workbench/Graphic display, Simulation Results, and Placed Components panels. The Components panel holds the components' containers. Each container contains a single type of component, such as resistors or capacitors, of different values. A user can search for, select, and pick up a component, drop a picked up component, place a picked-up component on the breadboard, remove a placed component from the breadboard, reposition the pins of components placed on the breadboard, and make connections using wires of different colours. The components placed on the breadboard are numbered in the order in which they are placed, according to their component types, such as R1, R2, R3, for resistors, and displayed in the Placed Components panel, together with their values and the nodes to which they are connected. This helps the user to identify each component on the breadboard at any point in time and is also used for issuing commands for circuit simulation.

The Tools panel contains two basic pieces of equipment: the Power Supply (PS) and the Function Generator (FG). The user can use either positive or negative voltage, from the power supply, as desired. The function generator can be set to output either a square, triangular or sinusoidal signal at different frequencies. Also, located in the tools panel is the menu bar which consists of the clickable buttons for clearing the workbench or the breadboard, switching between single and cascaded breadboard, among other functions. Laboratory instructions are displayed in the Instruction panel. After circuit construction, simulation is initiated by clicking on the SIM button in the Tools panel. This triggers a number of events at the back-end of the virtual workbench. The simulation output is displayed on the Simulation Results panel. The dummy voltage source behaves as an ammeter in accordance with Spice Opus use syntax [130].

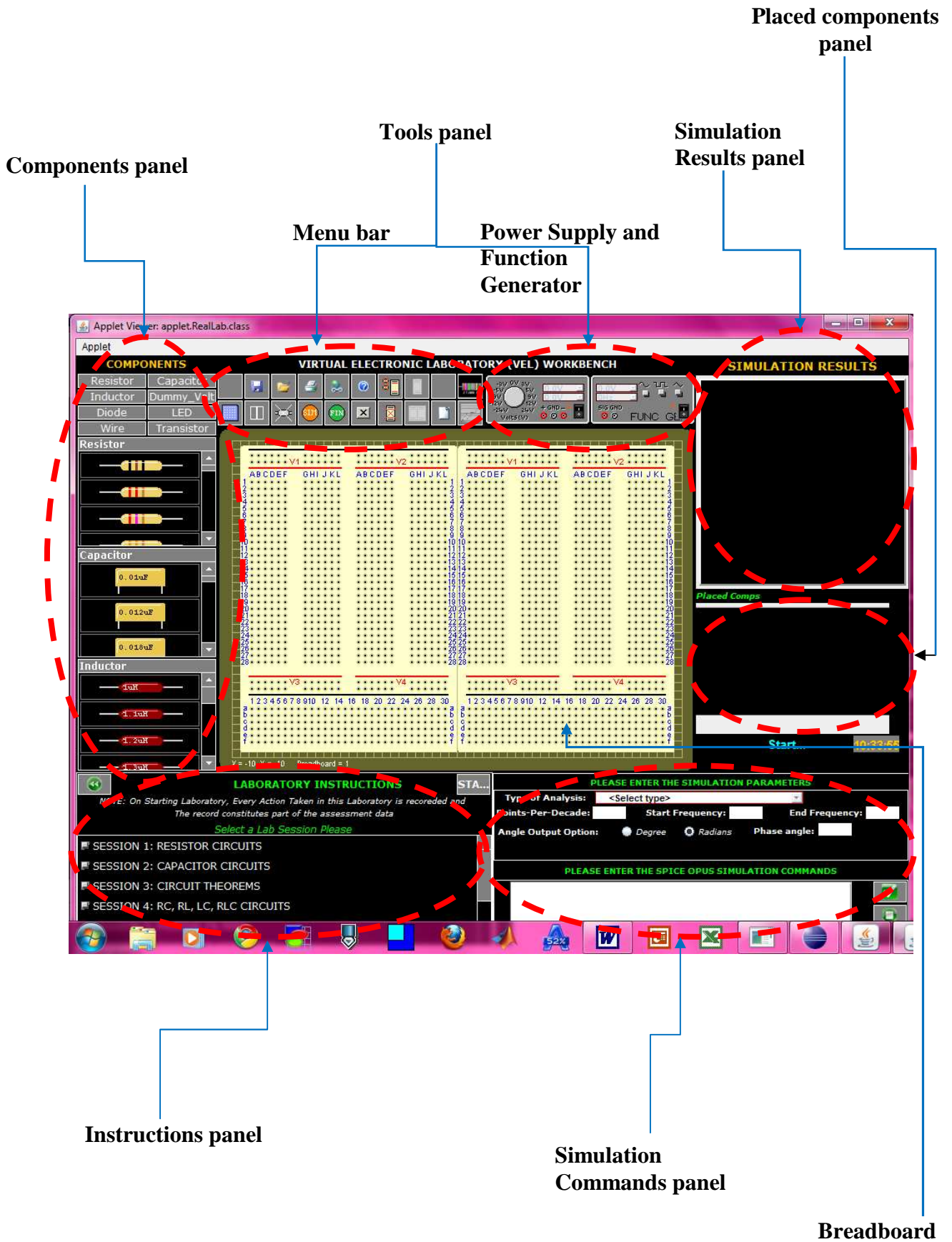


Figure 3.2: The VEL GUI

The events tracker/recorder in the VEL environment tracks and records students' behavioural events, with respect to the VEL GUI components and input fields. Figure 3.3 shows a sample events listing. On completion of a laboratory activity by a student, the behaviour log is sent to the server for storage, for analysis and extraction of performance data.

```

1242223587571 SINGLE_BB_BUTTON_CLICKED COMPONENT_VOID : VALUE_VOID MOUSE
1242223587742 SINGLE_BB_BUTTON_HOVER COMPONENT_VOID : 421ms MOUSE
1242223587898 DOUBLE_BB_BUTTON_HOVER COMPONENT_VOID : 141ms MOUSE
1242223593229 COMPONENT_ICON_CLICKED Capacitor : 6.8µF MOUSE
1242223595021 FIRST_CONNECTION_POINT_CLICKED Capacitor : 6.8µF MOUSE
1242223596096 SECOND_CONNECTION_POINT_CLICKED Capacitor : 6.8µF MOUSE
1242223598762 RESISTOR_CONTAINER_SCROLLBAR_PRESSED COMPONENT_VOID : VALUE_VOID MOUSE
1242223608051 RESISTOR_CONTAINER_SCROLLBAR_RELEASED COMPONENT_VOID : VALUE_VOID MOUSE
1242223608518 COMPONENT_ICON_CLICKED Resistor : 470.0? MOUSE
1242223609750 FIRST_CONNECTION_POINT_CLICKED Resistor : 470.0? MOUSE
1242223612508 SECOND_CONNECTION_POINT_CLICKED Resistor : 470.0? MOUSE
1242223615937 FUNCTION_GENERATOR_SIGNAL_TERMINAL_CLICKED AC_Signal : VALUE_VOID MOUSE
1242223617247 FIRST_CONNECTION_POINT_CLICKED AC_Signal : 1.0V MOUSE
1242223618712 SECOND_CONNECTION_POINT_CLICKED AC_Signal : 1.0V MOUSE
1242223628562 CLICK_ON_WORKBENCH COMPONENT_VOID : VALUE_VOID MOUSE
1242223629450 CLICK_ON_WORKBENCH COMPONENT_VOID : VALUE_VOID MOUSE
1242223629965 CLICK_ON_WORKBENCH COMPONENT_VOID : VALUE_VOID MOUSE
1242223630479 CLICK_ON_WORKBENCH COMPONENT_VOID : VALUE_VOID MOUSE
1242223630697 CLICK_ON_WORKBENCH COMPONENT_VOID : VALUE_VOID MOUSE
1242223630915 CLICK_ON_WORKBENCH COMPONENT_VOID : VALUE_VOID MOUSE
1242223634095 VIEW_PLOTTED_GRAPHS_BUTTON_HOVER COMPONENT_VOID : 140ms MOUSE
1242223641514 RESISTOR_COLOUR_CODE_CALCULATOR_BUTTON_CLICKED COMPONENT_VOID : VALUE_VOID MOUSE

```

Figure 3.3: Sample listing of events log

3.2.2 The Back-end (The Server-side)

The sequence of events that take place on the server-side, on invocation of the circuit simulator, are highlighted graphically in the upper part of Figure 3.1. At the backend, on the server machine, the server program and other applications software, including the simulation software package, SpiceOpus [131], and the graph plot software, Gnuplot [132], run. Also running on the server is the DataBase Management System (DBMS), MySQL, and the web server, Apache, under XAMPP [133], an integrated server package. SpiceOpus, Gnuplot, and XAMPP are freeware software tools.

On completion of circuit construction, and receipt of a circuit simulation invocation call from the client, the schematic of the constructed circuit and the simulation commands are captured and sent to the server where they are used to generate a netlist for input to the simulator. When the simulation of the circuit is complete, the results are returned to the client for display in the Simulation Results panel. The Simulation Results panel displays the numeric simulation results. Any generated graphs are displayed on the workbench. The breadboarded circuit on the workbench is removed and the graphic displayed on the workbench. After the graphic display, the breadboarded circuit is returned to the workbench and the user can resume working on it as desired.

3.2.2 Netlist Generation

A captured circuit schematic contains the list of components used to construct the circuit and their contact positions on the breadboard. Essentially, a schematic is a set of vectors of the form (<Component Name>, <Component Value>, <x1>, <y1>, <x2>, <y2>), where the x1, y1 and x2, y2 pairs indicate the first and second contact positions (coordinates) on the breadboard respectively. One aim of the netlist generation process is to convert the x1, y1 and x2, y2 pairs to actual node positions on the circuit. Three main steps are involved in the netlist generation process:

- wire removal (circuit collapse).
- nodes identification and node number assignment.
- components' left and right nodes number assignment based on their contact positions.

First, a temporary netlist is generated and passed to a verifier which checks the file for errors and ensures that it meets the netlist requirements for Spice Opus before the final netlist file is created and passed to the simulator, which simulates the circuit and returns results to the server. The Netlist holds the circuit information (circuit topology, components types and values) required by Spice Opus for the simulation process. A netlist is a set of vectors of the form: (<Component Name> <Node left> <Node right> <Value> <Unit>). Figure 3.4 shows a sample circuit schematic and Figure 3.5 shows the same circuit after wire removal and node number assignment. The nodes are numbered 0 to 7 with thick black dots indicating the nodes. The netlist file description of the circuit is listed in Table 3.1.

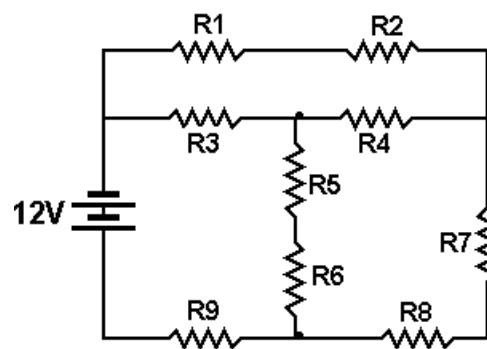


Figure 3.4: Sample circuit schematic
(Adapted from [Harb, 2000])

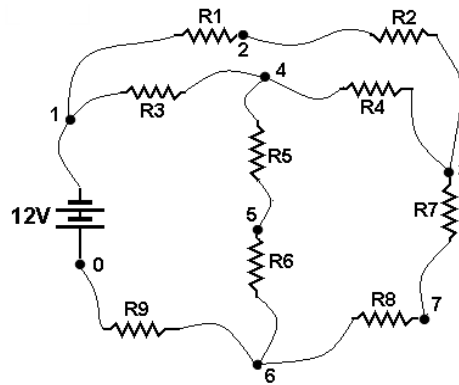


Figure 3.5: The sample circuit schematic after wire removal and node number assignment

Table 3.1: Netlist file description of the circuit of Figure 3.4.

vcc	1	0	dc	12V
r1	1	2	2.2k	
r2	2	3	2.2k	
r3	1	4	4.7k	
r4	4	3	4.7k	
r5	4	5	2k	
r6	5	6	2k	
r7	3	7	2k	
r8	7	6	2k	
r9	6	0	2k	

3.3 EVALUATION

3.3.1 Methodology and Context

Carefully designed laboratory activities are required for the effective use of the VEL. They can make invaluable contributions to students' learning [26]. The onus is on the instructor to design laboratory activities that promote students' understanding of concepts and to ensure that each practical exercise fits in with the overall aims of its focus course unit. Ideas and practical guides for this have been given by [2].

For instance, the instructor can methodically group the contents of a course unit into knowledge goals. Each knowledge goal can have a set of outcome and pre-requisite concepts. The latter are the concepts learners should have prior knowledge of, as "stepping stones" for learning the outcome concepts. The instructor can design concept-based laboratory activities and group them into sessions. Each laboratory session can address a specific knowledge goal while one or more laboratory activities in a session can address aspects of an outcome

concept, with clear learning objectives. Learning objectives are essential in applying an effective system of assessment [10]. This approach ensures that students carry out concept-based laboratory activities aimed at enhancing their understanding of the concept. An example of a knowledge goal is Filters whose outcome concepts include notch, low-, high-, and band-pass passive RC filters, at the fundamental level (see Table 3.2).

Table 3.2: A knowledge goal and its outcome and pre-requisite concepts

S/No.	Knowledge Goal: Passive RC Filters	
	Outcome Concept	Pre-requisite
1	Low pass filter	R, L, C circuit components characteristics. RC, RL, RCL circuits
2	High pass filter	
3	Band pass filter	
4	All pass filter	
5	Band reject (Notch) filter	

Along these lines, a set of laboratory activities aimed at enhancing students' understanding of passive RC filters were designed, to be undertaken by students in the VEL environment. The activities were designed around two related course units in two different engineering faculties at two separate Universities:

1. a 2nd year unit on Analogue Electronics and Circuits at the ECE department of University of Portsmouth (UoP). An aspect of the unit involves the design, building and analysis of passive RC filter circuits.
2. a 3rd year unit on Engineering Studio and Network Theory, at the EEE department of UTON, which addresses passive RC filters as part of its contents.

This was to facilitate the evaluation of the VEL at both Universities, using the same set of laboratory activities, Pre-Test and Post-Test (PTPT) and feedback form, albeit at different time periods.

3.3.2 Sample Student Cohorts

At UoP, only 30% of the class that took the analogue electronics and circuits unit volunteered and participated in the evaluation exercise ($N_P = 9$). At UTON, most of the students that took the course unit on engineering studio and network theory were willing and eager to participate in the evaluation exercise, but participation was limited to only 13% of the students ($N_F = 52$) for cost and logistic reasons. The UTON students were selected on a first-come-first-join

basis. The UoP and UTON student cohorts make up the total sample ($N = 62$). The gender distribution of the student cohorts were: 11% female ($n = 1$) and 89% ($n = 8$) male in the UoP cohort; and 10% ($n = 5$) female and 90% ($n = 47$) male in the UTON cohort. The samples did not include physics students and students from other fields of engineering who took the units as electives or compulsory units. All were undergraduate EEE (at UTON) and ECE (at UoP) students.

3.3.3 Evaluation Process

The aim of the evaluation process was to verify the effectiveness of the VEL as a tool for enhancing students' understanding of fundamental engineering concepts taught in lectures. PTPT observation technique was used to generate data for statistical analysis. PTPT is a widely used evaluation/assessment technique and is accepted as a reliable means of assessing learning and changes in knowledge [134][135]. Its use cuts across numerous disciplines. PTPT assessment is objective and offers a valuable set of data for evidencing increased knowledge with respect to specific concepts [135].

The PTPT consisted of Multiple Choice Items (MCIs). The items had the correct answer and three distracters for answer choices. The distracters were equally plausible, but incorrect and students were not timed for their responses to the items. The items were instruments designed as part of this work, in consultation with domain experts, course syllabi, proposed goals and course objectives. Despite the contested suggested limitations of MCIs, they are being increasingly used in higher education, as a result of growing student numbers, reduced resources and increasing use of new technologies [136]. Moreover, MCIs afford opportunities for rapid feedback and savings in marking time. Also, there was need to take advantage of the fact that the feedback provided by MCIs can be predetermined during test construction. [17], in addition to many other researchers, used the MCI-based PTPT technique for evaluating students' learning before and after an educational intervention.

Both student cohorts had taken lectures on passive RC filters, having learnt the pre-requisite concepts earlier. A pre-test was used to evaluate the students' knowledge of the addressed concept(s) before undertaking the laboratory activities in the VEL environment. Pre-testing was done after the addressed concepts had been taught in lectures, but before the students were exposed to the VEL environment. After taking the pre-test, the students then went ahead to undertake a set of laboratory activities in the VEL environment. The laboratory activities were designed to address four types of passive RC filter concepts: low-pass, high-pass, band-

pass, and notch. The PTPT were constructed to elicit students' knowledge of these filter concepts. The students undertook the laboratory activities at their own pace. The researcher and other persons were available to offer students necessary support. In addition, the students were provided tips and procedure on the usage of the VEL.

The students were aware that their work processes were being unobtrusively "observed" and logged. As sample screenshots, Figure 3.6 shows one of the laboratory activities undertaken by students and Figures 3.7 and 3.8 show a student's built circuit and graphical output, respectively, while undertaking the laboratory activity of Figure 3.6 in the VEL environment. The graphical output of Figure 3.8 was produced for the frequency range of 1000 (start frequency) to 10000 (end frequency), at 50 points per decade. On completion of the laboratory activities, in addition to taking a post-test, the students anonymously completed a usability feedback instrument.

The post-test was used to assess the impact of undertaking the laboratory activities in the VEL environment, on the students' learning of the addressed concept(s). The aim was not only to assess whether the educational intervention had had, possibly, a positive impact on the students' understanding of the addressed concepts, but also, to be able to quantify the impact. The post-test was administered immediately, on completion of the laboratory activities. The post-test items contained nearly twice as many items as the pre-test. All the items in the pre-test were included in the post-test, in addition to other items that assessed the same expected learning outcomes.

ACTIVITY-504: *Passive RC Filters: Notch filter.***AIM:** *To study the filtering characteristics of Notch RC filter*

Design a twin-T passive notch filter (RC only) capable of rejecting frequencies in the 3KHz to 3.5KHz frequency band (see the figure below). Assume $C = 1\mu\text{F}$.

1. Build the filter you have designed on a breadboard. Using a sinusoidal signal source of 1V_{rms} value, generate the frequency response curve for the circuit (without loading the circuit).
2. Load the circuit you have built with a $1\text{K}\Omega$ resistor and generate the frequency response curve of the loaded circuit.
 - a.) Use a single word to describe the shape of the curve generated.
 - b.) Is the frequency response of the filter affected by the load?
 - c.) If the answer to question b.) is yes, state how.
3. Reading the curve of the frequency response, what is the maximum rejection frequency?

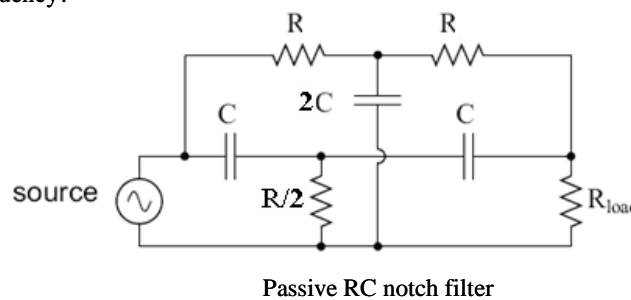


Figure 3.6: One of the laboratory activities undertaken by students in the VEL environment

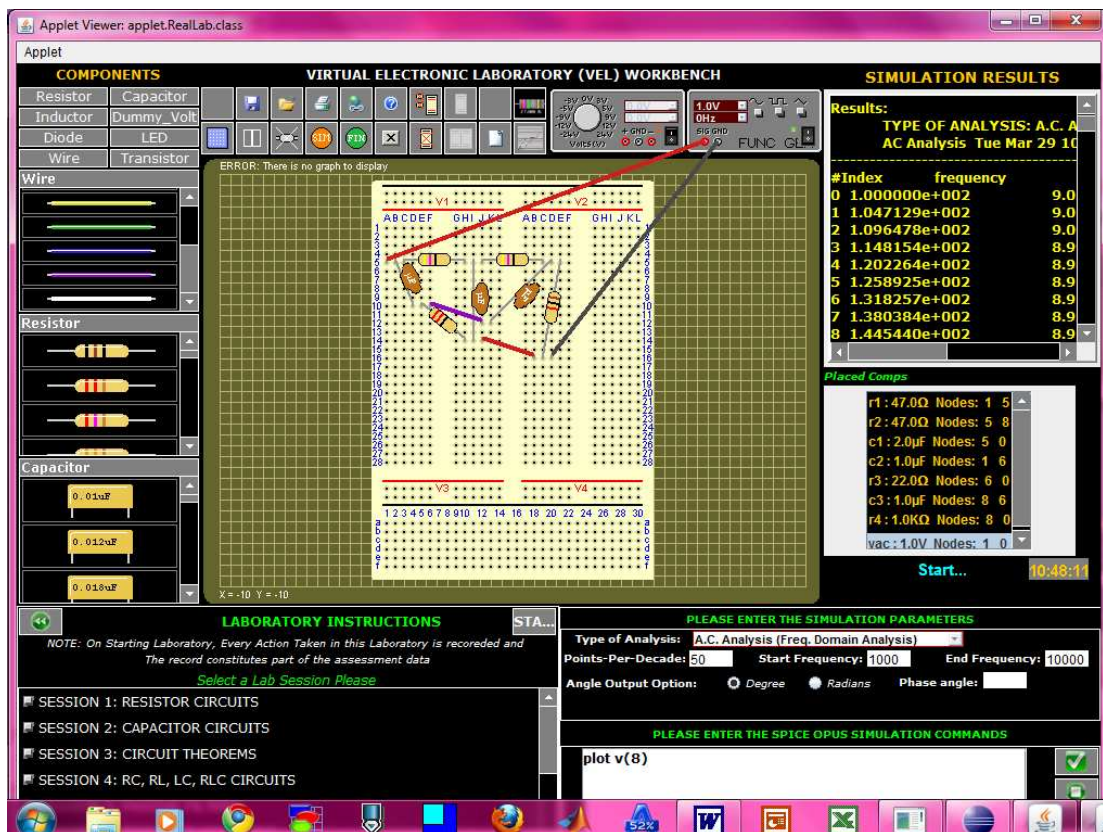


Figure 3.7: A student's built circuit for the laboratory activity of Figure 3.6, in the VEL environment

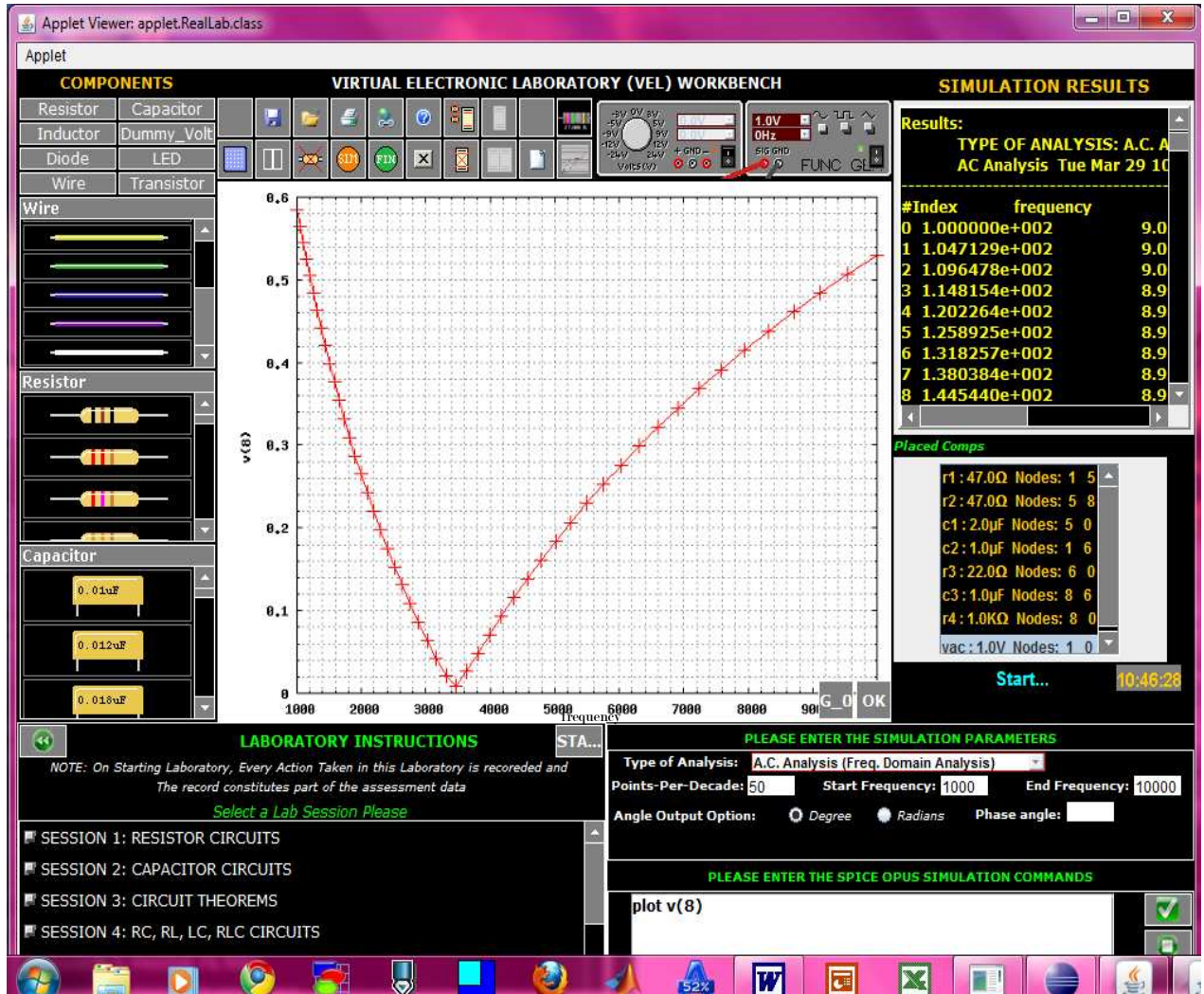


Figure 3.8: A student's graphical output for the laboratory activity of Figure 3.6

3.4 DATA ANALYSIS

The UoP and UTON student cohorts' PTPT scores are shown in Appendix A, Tables A1 and A2, respectively. The Statistical Analysis Method (SAM) and the Concentration Analysis Method (CAM) were undertaken with the PTPT scores and responses, respectively. The aim of the analyses was to highlight the impact of the educational intervention on students' learning, individually (using SAM) and collectively as a cohort (using CAM). Both analysis methods have been used by researchers in various contexts. The SAM needs no further introduction because of its commonality and wide disciplinary application. The CAM, more commonly used by physics researchers, was used, by [137], for the analysis of students' responses to the Force

Concept Inventory, a multiple-choice test designed to categorize students' understanding of basic Newtonian physics concepts. [138] and [139] used the CAM to investigate students' grasp of some of the basic concepts of force and motion. It was used, in this context, to investigate the collective impact of the use of the VEL, on each of the two student cohorts, by analysing the students' responses to the PTPT items designed to elicit their knowledge of passive RC filters.

The CAM analysis is important because, students' responses to multiple choice tests contain valuable information about their knowledge which may not be highlighted by the traditional SAM analyses of multiple-choice tests that focuses only on students' scores. The distribution of their responses can yield information on the collective knowledge state of the cohort and can be used to assess their collective learning [140]. In the realm of cognitive psychology, learning is an active and constructive mental process. Individuals perceive, process, store, and respond to information based on their internal mental states related to that piece of information [141]. Hence, individuals come to a given cognitive task with differing backgrounds of mental states resulting from learning [121]. The set of an individual's mental states, related to a particular concept, collectively constitute the individuals' knowledge state for the concept. The different pieces of mental states in the set constitute the fragments of the knowledge state for the concept and inform the individual's responses to tasks and tests items related to the concept. [137] asserted that students with similar knowledge states, with respect to a concept, are expected to respond to the same set of test items, designed to elicit their knowledge of the concept, in a relatively uniform manner; otherwise, their responses are expected to be "logically inconsistent, guided by a multiplicity of fragmented interpretations or ideas of the concept".

3.4.1 Statistical Analysis

3.4.1.1 *Gain Analysis*

Numerical summaries and Figures 3.9, 3.10, 3.11 and 3.12, on the face value, indicate improvement in the students' knowledge of the addressed concepts. Figures 3.9 and 3.10 diagrammatically represent the UoP cohort's PTPT scores and percentage gain, respectively, while Figures 3.11 and 3.12 represent the UTON cohort's PTPT scores and percentage gain, respectively. Percentage gain, in this context, refers to the difference between the pre-test and post-test percentage scores of a student. The P05, P06, P07, P09, P10, P11, P12, P13, and P14 labels in Figure 3.9 represent the user identities assigned to the UoP participating students. The

user identities do not necessarily need to be in a serial order as is the case for the UTON participating students (Figure 3.11). The requirement is that each user identity is unique.

Collectively, 79% ($n = 48$) showed enhancement in their knowledge and understanding of the addressed concepts. 10% ($n = 6$) of the total student sample actually showed “decline” in knowledge, while 11% ($n = 7$) seem to have failed to benefit from the laboratory activities and use of the VEL, which may be attributed to lack of interest and commitment, except for the one UoP student with PTPT scores of 91%.

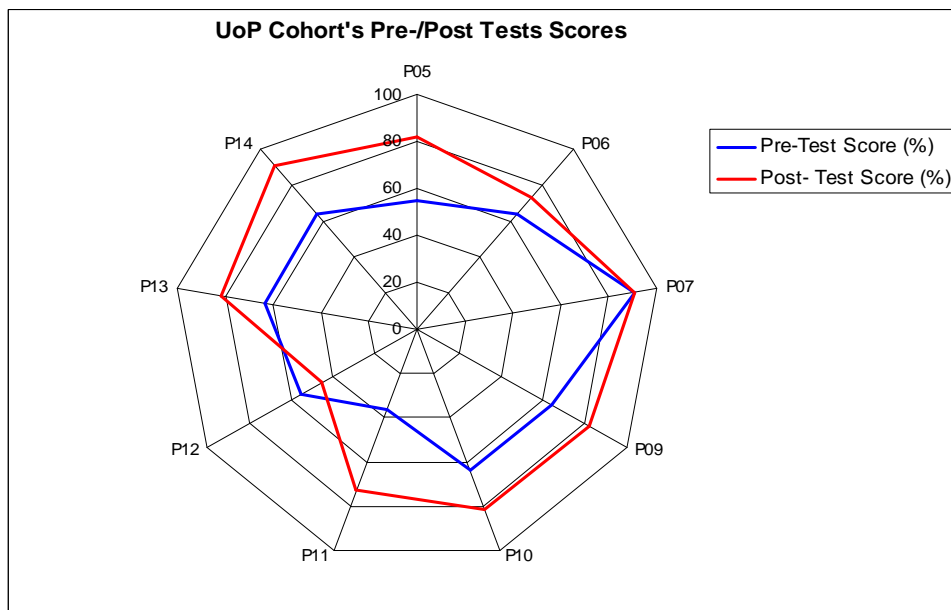


Figure 3.9: UoP cohort's pre-/post-tests scores

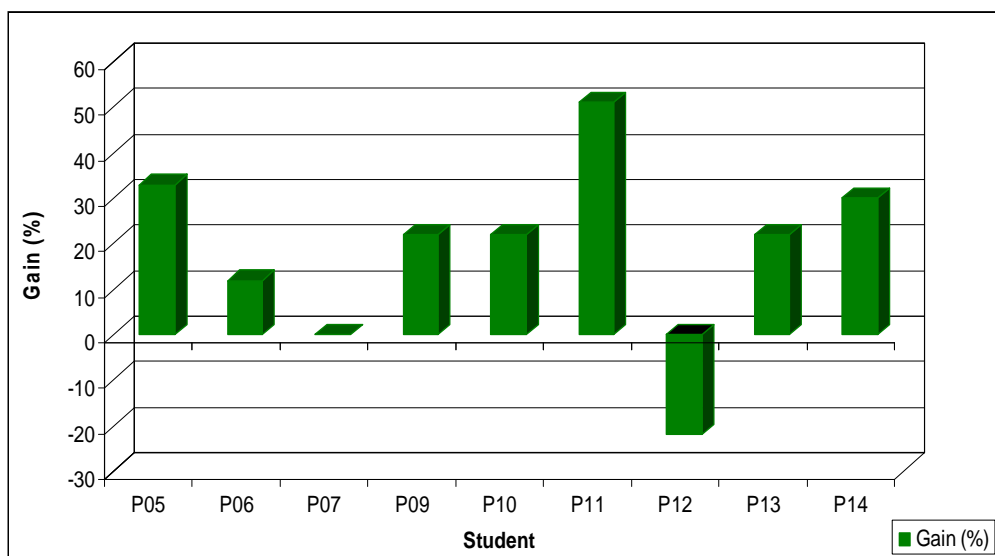


Figure 3.10: UoP cohort's percentage knowledge gain scattergram

The reasons behind the “decline” in knowledge were not investigated as part of this work. However, there is the possibility that the affected students may have undertaken the laboratory activities or responded to the test items randomly, without commitment. This raises the issue of the realness of students’ engagement in using a VLE. If they engage on an artificial, contrived level rather than on a deep, authentic level, then their learning experience will not be meaningful [143]. This is to be expected and probably representative of normal student performance. The mean percentage gains of the student cohorts, 18.89% (for UoP) and 24% (for UTON) are encouraging, but tests for statistical significance are necessary.

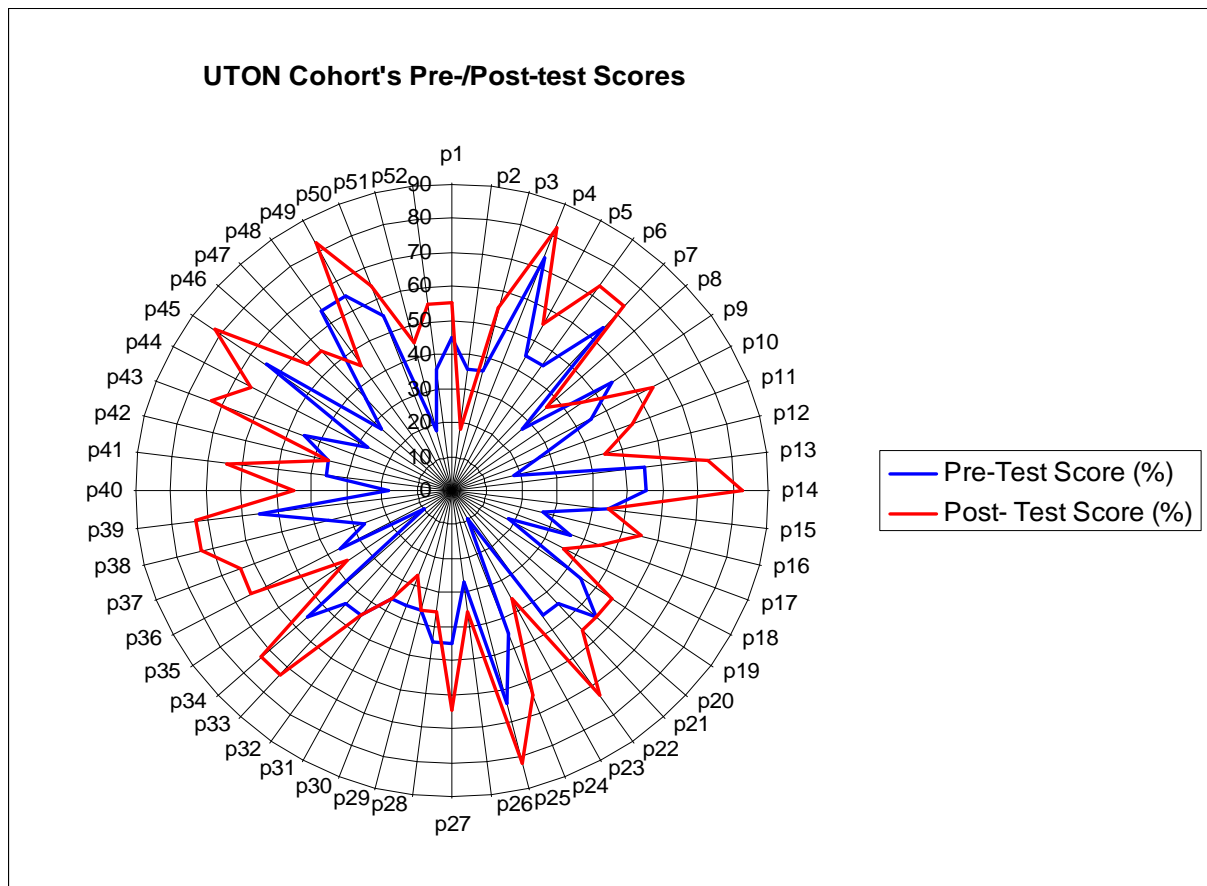


Figure 3.11: UTON cohort's pre-/post-tests scores

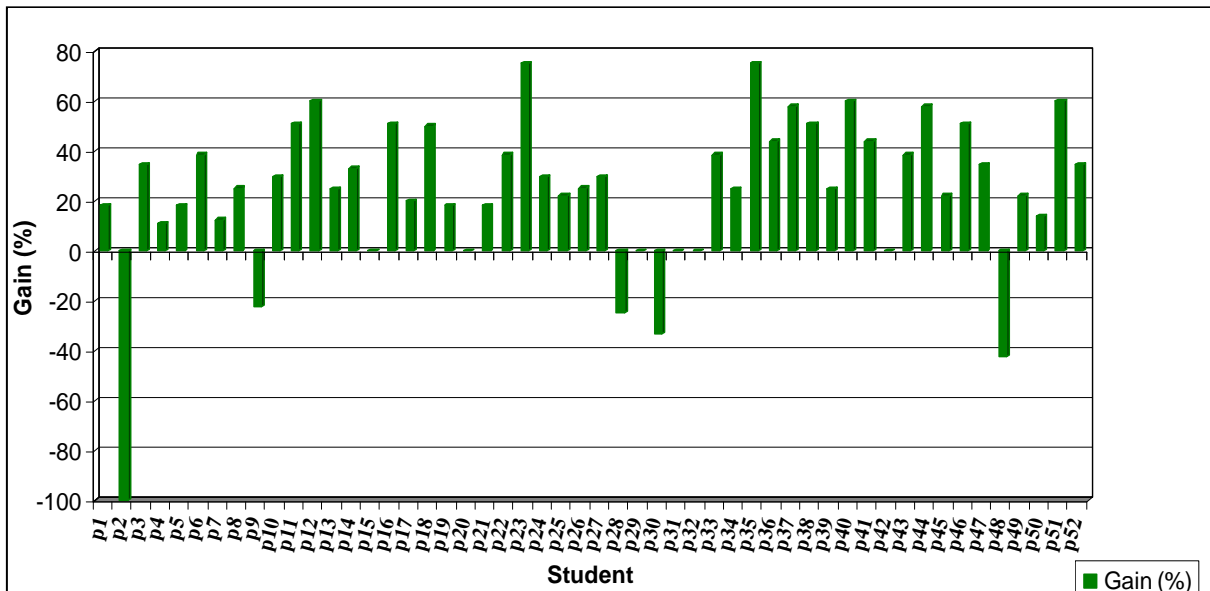


Figure 3.12: UTON cohort's percentage knowledge gain scattergram

3.4.1.2 Tests for Statistical Significance

The data was subjected to two-tailed t-test, with $\alpha = 0.05$. The aim was to statistically compare the post-test mean scores, μ_{P2} and μ_{F2} , to the pre-test mean scores, μ_{P1} and μ_{F1} . μ_{P1} and μ_{P2} are the UoP pre-test and post-test mean scores, respectively, while μ_{F1} and μ_{F2} are the UTON pre-test and post-test mean scores respectively. The aim of the test was to determine the value of the probability associated with the differences between the mean scores of the two student cohorts, in order to find out whether there exists statistically significant difference between them. In order to do this, it was hypothesized as follows:

1. *UoP Hypothesis:*

$$H_0: \mu_{P2} - \mu_{P1} = 0 \quad (\text{difference in mean scores is insignificant})$$

$$H_1: \mu_{P2} - \mu_{P1} \neq 0 \quad (\text{difference in mean scores is significant})$$

2. *UTON Hypothesis:*

$$H_0: \mu_{F2} - \mu_{F1} = 0 \quad (\text{difference in mean scores is insignificant})$$

$$H_1: \mu_{F2} - \mu_{F1} \neq 0 \quad (\text{difference in mean scores is significant})$$

The descriptive statistics for the pre-test and post-test scores, and the results of the tests for statistical significance are shown in Table 3.3. The UoP results, $t(8) = 3.459$ and $p = 0.00859$, compels the rejection of the UoP null hypothesis, indicating that the difference between the UoP pre-/post-test mean scores is statistically significant. Similarly, $t(51) = 6.804$ and $p = 0.000137$, indicate that the difference between the UTON cohort's pre-/post-test mean scores

is also statistically significant. This implies that the VEL has the capacity to enhance learning in a non-traditional manner. It can therefore, be suggested that the VEL can be effectively integrated in the laboratory education of students for the enhancement of their understanding of taught concepts.

Table 3.3: Statistical significance test results

Cohort	Measure	Pre-test	Post-test
UoP	Mean (μ)	61.89	77.89
	STD	14.30	13.88
	Count (N_p)	9	9
	t	3.4586	
	t critical two-tailed	2.306	
	P two-tailed	0.00859	
UTON	Mean	40.75	56.02
	STD	14.75	16.18
	Count (N_p)	52	52
	t	6.8038	
	t critical two-tailed	2.00758	
	P two-tailed	0.000137	

The Welch-Aspin procedure [144], based on pooled variance estimate, was used to investigate the performance differences between the two student cohorts. The test showed that the difference between μ_{P1} and μ_{F1} was statistically significant ($t^*(9, 52) = 4.08$; $p = 0.002$). So also was the difference between μ_{P2} and μ_{F2} ($t^*(9, 52) = 4.25$; $p = 0.001$). It was also sought to find the association between the percentage knowledge gain and total time on the task, using bivariate analysis, in the form of Pearson product-moment correlation measure. There was a weak positive correlation between them ($r = 0.0897$) which was not significantly different from zero ($p = 0.486$), which implies that percentage knowledge gain is independent of the time on task.

3.4.2 Concentration Analysis

As stated at the beginning of section 3.4, the CAM, by [140], was used in this context, to analyse the students' responses to the PTPT items in order to determine the collective impact of the use of the VEL, on each of the two student cohorts. The analysis yielded useful results, highlighting interesting themes.

Students' responses to each item in a test are considered to be the results of the application of their knowledge states for the addressed concept. Each item has associated with it a fragment of the knowledge state required to answer that question correctly. If a multiple-choice item is designed with the knowledge state fragment required to answer it correctly as one of the choices for the item, while the distracters are equally plausible but incorrect and constitute misconceptions, student responses should be concentrated on the correct choice if they are applying the correct knowledge state fragment in responding to the item. However, if students have little knowledge of the concept, their responses may be close to a random distribution among all the choices. As an example, Table 3.4 lists an arbitrary sample set of 100 students' response distributions to a set of 5-choice multiple-choice test item.

Table 3.4: An arbitrary sample set of students' response distributions for test items
(Adopted from [Digironimo, 2007])

Responses Distribution Type	Item Choice				
	A	B	C	D	E
I	20	20	20	20	20
II	50	10	30	5	5
III	100	0	0	0	0

The way in which the students' responses are distributed can yield information on the students' knowledge states. If the students get low scores, their responses are typically either evenly distributed among the different distracters or concentrated on one or two of the distracters. Highly concentrated responses to an item imply that many students are applying a common fragment of the knowledge state associated with the concept; randomly distributed responses indicate less uniformity in reasoning, highlighting possible lack of clue and guessing on the part of the students [137]. The formula and details of the concentration analysis method are given in Appendix B2, while the student cohorts' response distributions to the pre- and post-test items are given in Appendix B3.

The CAM uses a synergy of the *Score (S)*, and the *Concentration Factor (C)*, both of which are functions of the students' responses, to derive the Students' Response Patterns (SRPs) for each item in a test. It requires that the pre-test and post-test instruments contain the same items, in the same order. In this context, the post-test contained nearly twice as many items as the pre-test, where all the items in the pre-test were included in the post-test, in addition to other items that assessed the expected learning outcomes. For the purposes of the analysis, pre-test items component of the post-test were extracted, in order, and students' responses to the items matched against their responses to the pre-test items. The results of the CAM analysis, in terms of the students' *S*, *C*, and SRP distributions for the pre-test items and their corresponding post-test items, are highlighted and discussed in Appendix B 3. Table 3.5 presents a summarized view of the results, with respect to the defined SRPs (see [140]), what the SRPs indicate, the number of the SRPs in the UoP and UTON cohorts' responses to the pre- and post-test test items, and remarks. The remarks relate the indications for an SRP to each cohort's performance for the items with the SRP.

Table 3.5: Concentration analysis of students' responses to the pre- and post-test items.

Students' Response Pattern (SRP) with respect to a test item	Indication	UoP		UTON		Remarks
		Tests Outcome		Tests Outcome		
		No of the SRP in the Pre-Test	No of the SRP in the Post-Test	No of the SRP in the Pre-Test	No of the SRP in the Post-Test	
HH	Most of the students generally have one dominant knowledge state fragment with respect to the item. This one dominant knowledge fragment the students have is the correct knowledge fragment required to respond correctly to the item	4	7	0	4	<p>UoP: For 4 of the items, most of the students had the one correct knowledge fragment, before the educational intervention. After the education intervention, the number of items for which most of the students had the one correct knowledge fragment increased from 4 to 7, indicating <i>improved learning</i>.</p> <p>UTON: Before the educational intervention, there was no item for which most of the students had the one correct knowledge fragment for the item. After the education intervention, the number of items for which most of the students had the one correct knowledge fragment increased from 0 to 4, indicating <i>improved learning</i>.</p>
MM	Most of the students generally have either of two dominant knowledge state fragments. Of these two knowledge state fragments that most of the students have, one is the correct knowledge state fragment required to respond correctly to the item, while one is the wrong knowledge state fragment and will lead to incorrect responses to the item.	5	3	4	1	<p>UoP: The number of items for which over 30% of the students had an incorrect knowledge fragment was reduced from 5 to 3, indicating <i>improved learning</i>.</p> <p>UTON: The number of items for which over 30% of the students had an incorrect knowledge fragment was reduced from 4 to 1, indicating <i>improved learning</i>.</p>
LH	Most of the students generally have one dominant knowledge state fragment with respect to the item. This one dominant knowledge fragment the students have is incorrect and will lead to incorrect responses to the item.	1	0	0	1	<p>UoP: The one dominant incorrect knowledge fragments most of the students had for one of the test items, before the educational intervention, was corrected after the education intervention, indicating <i>improved learning</i>.</p> <p>UTON: For one of the test items, the students had no dominant incorrect knowledge fragment, before the educational intervention, but the students acquired an incorrect knowledge fragment for the item after the education intervention, indicating <i>reverse learning</i>.</p>
LM	Students have two dominant incorrect knowledge state fragments Most of the students generally have one of two wrong ideas	1	1	1	0	<p>UoP: The two different incorrect knowledge fragments most of the students had for one of the items before the educational intervention was not corrected after the education intervention, indicating <i>no learning</i>.</p> <p>UTON: The two different incorrect knowledge fragments most of the students had for one of the items before the educational intervention was corrected after the education intervention, indicating <i>improved learning</i>.</p>
LL	Most students generally have one of three or more different knowledge state fragments, somewhat evenly represented among the student cohort. These three or more different knowledge state fragments may or may not include the correct knowledge state fragment required to respond correctly to the item. This often implies that the students generally have no clue and are guessing the answer to the item.	0	0	6	1	<p>UoP: No guessing of responses to any of the test items before and after the educational intervention.</p> <p>UTON: The number of items for which most of the students were guessing without a clue was reduced from 6 to 1, indicating <i>improved learning</i>. Students became surer of their responses to the items,</p>

Table 3.5, indicates that the usage of the VEL had improved learning impact on both student cohorts, and more especially on the UTON cohort, which reinforces the results of the statistical analysis method. The outcome of the analysis has implications for instruction in that the SRPs inform the instructor of aspects of the addressed concept that require further instructional attention. For example, in Table 3.5, the SRP, LH, did not occur at all in the pre-test response patterns of the UTON student cohort, but occurred once in their post-test response patterns (for the item, Question Q6, in Table B3.2, Appendix B3). This indicates that the instructor may need to further address aspects of the concept represented by that item, for the UTON cohort. Also, the SRP, LM, occurred once (for the item, Question Q6, in Table B3.1, Appendix B3) in the pre-test response patterns of the UoP student cohort, and occurred again (for the same item, Question Q6, in Table B3.1, Appendix B3) in their post-test response patterns. This indicates that the instructor may need to further address issues related to that item, for the UoP student cohort, because of the implication of its repeated occurrence in both their pre- and post-test response patterns, for the same item. The implication is that most of the UoP students had one of two incorrect knowledge fragments about the aspect of the addressed concept represented by the item (Question Q6), and undertaking laboratory activities designed to enhance students understanding of the concept failed to correct the students' incorrect knowledge fragment for the aspect of the concept.

3.5 INDIVIDUAL STUDENT ADVANTAGE IN THE VEL ENVIRONMENT

Students learn in many different ways [86]. How much a student learns depends on the compatibility of the student's Learning Style (LS) with the learning environment [85]. A student may find a learning environment more favourable, in comparison to others, in terms of learning gains, based on his/her LS preferences, which will impact on his/her performance [145] [146]. Hence, LS preferences form a basis upon which to assess the relevance and impact of the VEL usage. The student cohorts completed the Index of Learning Styles (ILS) instrument by [86]. The ILS was specifically proposed for engineering students and is used to classify a student on a 4-dimension LS preference scale. The dimensions represent the [85][86] [147]:

1. type of information readily perceived by a student (the preferred information type): *sensory and/or intuitive*.
2. way a student best receives information (input or communication channel): *visual and/or verbal*.
3. way a student processes perceived information (processing): *active and/or reflective*.

4. way a student progresses/works towards understanding perceived information (progression towards understanding): *sequential* and/or *global*.

The results of the analysis of the ILS instrument highlighted the following distribution for students' distinct LS preferences for the scales: Sensory 33%, Intuitive 20%; Visual 57%, Verbal 13%; Active 21%, Reflective 16%; Sequential 23%, Global 13%. It was found that the results corroborates the findings, by [148] [149] and [150], that engineering students typically have preferences towards Sensory, Visual, and Active learning styles.

Positive relationships were found between percentage knowledge gain and degrees of preference of the following LS dimensions: sensory ($r = 0.143$, $p = 0.547$); intuitive ($r = 0.214$, $p = 0.504$); verbal ($r = 0.20$, $p = 0.635$); and sequential ($r = 0.157$, $p = 0.591$), but none of the relationships was statistically significantly different from zero. The relationship between students' ease of use and enjoyment of use of the VEL (based on their responses to the feedback questions relating to these issues) and the LS scales was also investigated. There was a statistically significant correlation between quantitative responses to the question on ease of use and scores on: Sensory LS scale ($r = 0.358$, $p = 0.132$), and Visual LS scale ($r = 0.188$, $p = 0.279$). Also, there was a statistically significant correlation between quantitative responses to the question on whether the students enjoyed using the VEL and scores on the Sensory LS scale ($r = 0.325$, $p = 0.175$). This is not surprising because, according to [85], sensory learners tend to be practical oriented and are more likely to enjoy hands-on sessions. The correlation between ease of use and scores on the Visual LS scale corroborates the statistically significant correlation finding, between scores on the Visual LS scale and responses to a question concerning student comfortableness using the Universal Virtual Laboratory (UVL) [55] ($r = 0.227$ and $p = 0.019$). This implies that students who expressed preference for diagrammatic, pictorial and demonstration information sources find it easier to use virtual laboratories.

3.6 FEEDBACK FROM STUDENTS

Usability feedback was elicited from students, using a 1-5 Likert instrument. [151] stated that Likert scale "is the most commonly used question format for assessing participants' opinions of usability". This is probably because they are much easier to construct and more reliable than other scales with the same number of items [152]. According to [153], "the simplicity and ease of use of the Likert scale is its real strength".

The collective responses of the students ($N = 61$) indicate an overall positive reaction. 90% ($n = 55$) said that the use of the VEL enhanced their understanding of the concepts addressed by the laboratory activities. This is reflected in the pre-/post-test scores. 85% ($n = 52$) of the students said the VEL was easy to use and 97% ($n = 59$) said they enjoyed participating in the evaluation exercise and using VEL. 98% ($n = 60$) agreed that they would recommend the VEL for use by students. The overall rating of the VEL was encouraging.

Students' comments were uniformly and overwhelmingly positive, eager and enthusiastic, especially at UTON, where the VEL was well received. Comments about the best aspects of the VEL reflected the ability to construct and simulate circuits realistically without the need for real equipment. The students' opinions and comments indicated that the VEL, in the words of the students:

1. is user friendly, does not waste components, is quick to plot graphs and see results in real-time. "It is easier to use than fiddling with real components".
2. is realistic, demonstrates real-life circuit building, and "makes practicals more real-life than any other electronic software I know".
3. "will enhance our performance in laboratory" and "if implemented in our school will really enhance our understanding of electrical networks".
4. "makes room for classes with numerous students like ours to participate individually and actively in laboratory exercises", "a good way to understand basic principles before exploring it further in detail". It is "a very good learning tool and helps to develop students' knowledge before and after lectures".

The participants suggested that the VEL should include:

1. fault detection and correction: ability to point out errors in a breadboarded circuit and suggest possible solutions about the errors, with sound alert for wrong connections.
2. clickable buttons for simulation commands instead of having to enter the simulation commands.
3. Integrated Circuit (IC) components and micro-controllers.
4. the ability to remember previously selected/placed components and their values, and the facility for students to ask questions and/or be asked questions, and be given suggestions and answers to questions.

It is worth noting that the VEL is a laboratory environment and not an Intelligent Tutoring System (ITS) as some of the students recommendations are wont to suggest. In future, the

VEL can be improved upon to incorporate ICs for building circuits, among other possible improvements.

3.6 SUMMARY

The VEL has been described and subjected to evaluation studies on two separate occasions with different student cohorts. The aim of the study was to evaluate the VEL as a tool for facilitating students' enhanced understanding of fundamental engineering concepts taught in lectures, using carefully planned laboratory learning scenarios. PTPTs were used to assess and quantify the students' knowledge states before and after undertaking concept-based laboratory activities in the VEL environment. The analysis of the evaluation data showed that the remarkable positive differences in the students' knowledge states, before and after the educational intervention, did not occur by chance. This implies that the VEL has the capacity to enhance learning, in the light of appropriate learning scenarios. Furthermore, the observed performance differences between the student cohorts and the greater improvement in learning by the UTON cohort combine to reinforce the highlighted need for complements and/or alternatives to the traditional engineering laboratory in faculties challenged by resource constraints and/or large class sizes, which is often the case for faculties in developing countries.

The use of the same set of laboratory activities to evaluate the VEL in two different Universities reinforces the uniformity of fundamental engineering concepts and indicates that the manner in which the VEL can be integrated into laboratory learning is not faculty or unit specific and does not require adaptation. The VEL could also be useful in secondary schools' science laboratories, and other engineering laboratories colleges and polytechnics without adaptation. The feedback responses suggest a generally positive and enthusiastic attitude to the VEL. It was well received and acknowledged as a tool with which to realistically undertake curriculum-based laboratory activities. The next chapter details the construction of the assessment model, for the performance assessment of students' laboratory work in the VEL environment.



CHAPTER 4

BAYESIAN NETWORK-BASED INTELLIGENT ASSESSMENT MODEL: STRUCTURE REALISATION AND CALIBRATION

Intelligence is what you use when you don't know what to do.

Jean Piaget,
1896 - 1980

4.0 INTRODUCTION

The intelligent assessment model is aimed at providing the instructor a tool that facilitates the timely, consistent, and fair performance assessment of students' laboratory work. The model is Bayesian Network (BN) –based and its construction entails two main tasks: definition of the model structure, and the calibration (parameterization) of the model. This chapter details the construction of the assessment model.

First, the derivation of the foundational framework is detailed in section 4.1. Section 4.2 describes the process of realization of the model structure, while section 4.3 details its calibration process. The chapter is summarized in section 4.4.

4.1 THE MODEL FOUNDATIONAL FRAMEWORK

4.1.1 Psychology of Learning: the Basis for Derivation of the Foundational Framework

Psychology of learning (focused on understanding how the learning process works and the effect of learning on behaviour) provides the base framework for the assessment model. There is a large amount of literature on learning, learning theories, and learning models. Learning theories and models are ideas about how learning may happen (conceptualization of the learning process), and are meant to be applied in the instructional process, in order to facilitate learning by instruction and assessment [153]. Assessment drives learning [40]. Researchers give different definitions of learning, each emphasizing different aspects [155]. These differences in emphasis reflect the multidimensional nature of learning and the difficulty of having a single definition or model that will satisfy different purposes and learning situations [155]. The goal herein is not a detailed discussion of learning theories and models, as that is outside the scope of this thesis. It is instead, to integrate aspects of literature on learning in order to derive a foundational framework for realization of the BN-based assessment model structure.

Hence, the foundational framework of the assessment model is based on a model of learning, formulated as part of this work, from various definitions and theories of learning. The framework constitutes the basis for Cognitive Task Analysis (CTA -- discussed in Chapter 2 (section 2.8)), in order to identify and factorize the cognitive abilities/skills necessary to successfully undertake laboratory activities as part of an undergraduate electronic engineering course. Factorisation, in this context, implies the decomposition of a cognitive ability/skill

into its constituent atomic (representative) abilities/skills. That is, complex abilities/skills are broken down into constituent abilities/skills and their interrelationships identified. There is the tendency to differentiate between abilities and skills as internal general and specific potentials, respectively. In this context, abilities and skills are used jointly to mean context specific potentials and are often written together as abilities/skills.

4.1.2 Derivation of the Foundational Framework

It is possible to account for the various theories and definitions of learning with a model of learning that incorporates the common factors (emphasised either explicitly or implicitly) in most of the theories and definitions, in such a way that learning can be addressed by instruction and assessment. This implies the need for concern with the basic components of learning that are assessable. Thus the adopted model of learning, in this context, is constrained to the basic framework of learning as consisting of:

- *Knowledge*-- acquiring and storing facts, rules and principles about a concept;
- *Understanding* (meaning making)-- understanding of expressed or implied interconnections and relationships between facts, rules and principles about concepts, in relation to specific tasks or problem situations;
- *Abilities/Skills*-- mastering the procedures and techniques required to apply knowledge and understanding in specific task or problem situations, in a domain context.

This framework, based on the theories of learning by [155][156][157] and definitions of learning by [158][158][160], is consistent with the definition of assessment as a generic term for a set of processes that measure the outcomes of learning, in terms of knowledge acquired, understanding developed, abilities/skills gained by [40] and [161]. It is also consistent with [11] and [162] definition of engineering education outcome as consisting of knowledge and understanding and abilities/skills, and the assertion by cognitive psychologists [121], that the acquisition of knowledge, understanding, and abilities/skills results in improved performance which reflects learning. The basic framework of learning, represented graphically in Figure 4.1, highlights the main learning related variables or factors that directly influence students' performance of a given task.

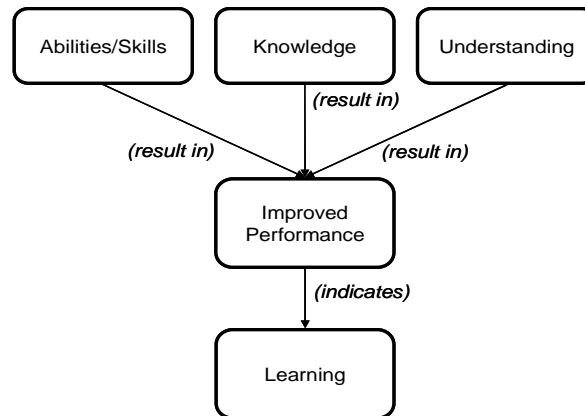


Figure 4.1: Basic framework of learning

Learning is viewed from three schools of thought: *behaviourism*, *cognitivism* and *constructivism*. The Behaviourist position on learning is that learning takes place as the result of a repeated Stimulus (S) - Response (R) cycle which conditions an individual into repeating that particular response whenever the same stimulus is present (experience acquisition) [163]. *Behaviourism* is concerned with external behaviour only, viewing learning as controllable through behaviour modification by experience, rejecting the role of mental processes/activities in the learning process. *Cognitivism* acknowledges the need to examine the internal mental processes that take place between the stimulus and the response, viewing learning as an active process that involves the internal traits/states and processes/activities of the learner. Learners do not merely respond to stimuli, rather their responses are consequences of their internal traits/states and processes/activities which play a key role in determining the responses. Learning would be impossible without internal mental traits/states and processes/activities [164]. *Constructivism* also acknowledges the role of traits/states and processes/activities in learning, but asserts that learners construct their own reality (knowledge formation) based on their experiences. New knowledge is formed from the learner's previous experience based on existing internal structures and beliefs which are used to interpret learning events. Furthermore, constructivism advocates that learning should take place in realistic settings to enable the learner construct knowledge better adapted to the context in which it will be applied. Essentially, cognitivists/constructivists agree that learning is an active and constructive process that involves the use of internal mental traits/states and processes by the learner.

Laboratory work, like most learning tasks, is limited by the cognitive function. Individuals come to a given cognitive task with differing backgrounds of internal traits/states resulting from learning which support and determine the limit of individual performances [121]. According to [124], internal traits/states, including abilities/skills, may combine to produce instances of observable behaviour. Abilities/skills, according to [121], exist from their effects in terms of performance and are measurable or inferable from observable behaviour. [164] asserted that in order to measure or make inferences about the internal traits that cannot be observed or measured directly, experimental situations that will elicit or reflect the traits in the learner's behaviour have to be created. These assertions further extend the framework of Figure 4.1, resulting in the graphical model of Figure 4.2, the foundational framework of the **LA**boratory **P**erformance (LAP) assessment model.

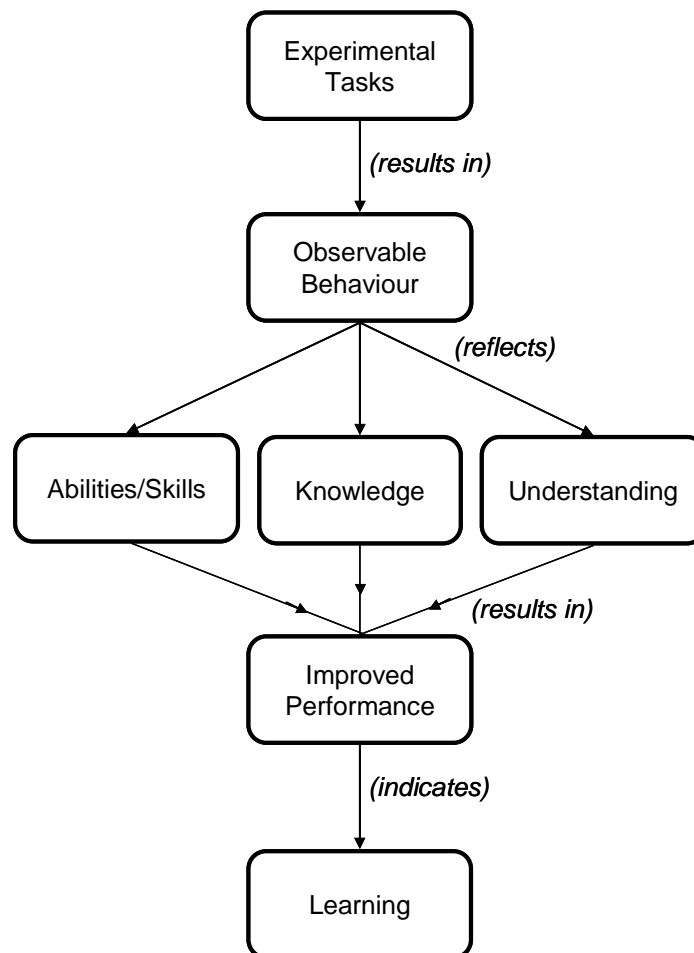


Figure 4.2: The foundational framework of the LAP model

The framework is consistent with the behaviour-observation-trait-inference assessment perspective of [89]. It facilitates the realization of the BN-based assessment model for students' laboratory work performance assessment, from a holistic perspective and constitutes the basis on which to derive a set of students' laboratory performance indicators and rubrics.

4.2 STRUCTURE REALIZATION

Defining the graph structure syntax for a BN model is still a major challenge [165]. This challenge is multifaceted: identifying the set of, $X = \{x_i\}$, variables that will constitute the nodes of the model and choosing their possible values (states); deciding which nodes should be connected by edges; and choosing the direction of the edges for each pair of connected nodes. The first aspect is a general BN modelling problem, but the later two are especially a challenge for non-causal models, which, in addition, demand semantics identification. Hence, in defining the structure of the LAP model, the following issues are taken into consideration [165]:

- when there are many states for a node in a BN model, and the number of its parents become greater than two, it becomes increasingly difficult to elicit its conditional probabilities exhaustively. As the number of states increases to seven and the number of parents become greater than two, exhaustive elicitation becomes infeasible. As highlighted in section 2.6, there is, an exponential relationship between the size of the CPT of a node, the number of its parent nodes, and the number of possible states of its parent nodes. Generally, the number of entries in the CPT of a child node, A , that has k number of parents, where each parent has n possible states, is n^k . For example, the node, A , has k parents and each parent has two possible states, then the CPT for node A will contain, 2^k entries ($k = 2, k = 8, \text{ and } k = 16$, will result in 4, 256, and 65536 entries respectively). If each of the parents has three possible values, 3^k entries are required ($k = 2, k = 8, \text{ and } k = 16$, will result in 9, 6561, and 43046721 entries respectively).
- a BN model need not necessarily model the same type of relationship throughout the network. Different sections or fragments can model different relationships.

First, the cognitive abilities/skills necessary to successfully undertake laboratory activities as part of an undergraduate electronic engineering course had to be identified, and where necessary and possible factorised, using the CTA framework (discussed in Chapter 2, section 2.8). In this context, the CTA framework assumed that simpler cognitive abilities/skills form the basis for higher level abilities/skills so that the start of the CTA was the identification of the key cognitive abilities/skills (the high level or critical cognitive abilities/skills) related to the undertaking laboratory activities within the context of the domain. This entailed the development of a vocabulary associated with the domain, in order to facilitate the CTA. As part of the CTA process, laboratory instruction manuals and task assignment sheets were

reviewed, students undertaking laboratory activities in the traditional laboratory environment were physically observed, and the help of domain experts and Cognitive Psychologists were elicited, in order to identify and factorize the laboratory-related cognitive abilities/skills at the undergraduate level of electronic engineering.

Key laboratory activity tasks were identified and represented with the action verbs: *design*, *analyze*, *construct*, *modify*, *measure*, *plot*, *observe*. In this context, these verbs, referred to as the High-Level Abilities/Skills (HLAS), are taken to mean:

- *design*-- devise the schematic of a circuit to meet specified need(s);
- *construct*-- connect electronic components together to create a circuit whose collective behaviour meets a given specification(s);
- *analyze*-- practically examine a given circuit in order to detect behaviour patterns and draw conclusions;
- *modify*-- alter or adapt a circuit to meet a new purpose, application, or specification;
- *measure*-- determine the size or amount of, express as a number or quantity;
- *plot*-- generate vector graphics using a set of experimental data; and
- *observe*-- look at experimental result with a view to its interpretation and data extraction.

The verbs all represent cognitive abilities/skills. Psychomotor and interpersonal abilities/skills are not considered relevant in this context. Some of the HLAS are amenable to factorization into Basic Ability Components (BACs) because, according to [123], performance activities can be described in terms of BACs most of which can be considered as attributes of behaviour. Table 4.1 highlights a HLAS and its associated BACs. A BAC can further be factored into Lower Level BACs (LLBACs). Table 4.2 highlights a BAC and its associated LLBACs. The LAP model structure is based on such factorizations to create a structure that models relationships that are not necessarily causal. To reduce the complexity of the model, the variables were limited to abilities/skills that have:

- minimum overlap in their BACs and LLBACs factorization.
- potential to impact on performance

Table 4.1: A HLAS and its associated BACs

HLAS	ASSOCIATED BACs
Ability to Construct a Circuit	Ability for Proper Use of Equipment (PUE)
	Ability to Work With Components (WWC)
	Ability to Adapt (ADA)
	Ability to Adhere to Constraints (ADC)

Table 4.2: A BAC and its associated LLBACs

BAC	ASSOCIATED LLBACs
Ability to Work with Components (WWC)	Ability to Use Correct Components (UCC)
	Ability for Correct Placement of Components (CPC)
	Ability to Use Correct Alternative Component (UAC)

Other cognitive abilities considered to have contributory effect on laboratory activity performance are Spatial and Receptive Communication (SRC) abilities. Spatial ability is one of the most important abilities a student should possess to be successful in the engineering profession [165]. It is the ability to understand spatially-based information which can also be received in textual descriptions [38]. Receptive communication is the ability to understand instructions which is key to carrying out a task. It is believed that performance may be compromised without these abilities.

In addition to cognitive abilities/skills, learning-related factors that have the potential to influence students' laboratory activities performance in the VEL environment, such as learning style, were identified. Though [167] suggested that literature fails to provide adequate support for applying LS in educational contexts, LS was considered a possible influential factor, because LS has been shown to serve as a stable indicator of how learners perceive, interact with, and respond to a learning environment [168]. Students learn in many different ways [86]. How much a student learns depends on the compatibility of the student's

learning style with the learning environment [85]. A student may find a learning environment more favourable, in comparison to others, in terms of learning gains, based on his/her LS preferences, which will impact on his/her performance [145] [146]. Student's LS was evaluated, in this context, using the Index of Learning Styles (ILS) instrument [86]. The ILS was chosen for the following reasons:

- it was specifically proposed for engineering students,
- the instrument has been validated [169][170][171],
- the four LS dimensions are numerically coded and easily quantified.

Also, experience using the VEL is considered an influential factor in student performance based on the assertion that a user will develop a more efficient process, over time, with the usage of a system, which will impact on his/her performance [109]. A student's experience using the VEL environment is derived as a function of the total time the student has spent, on tasks within the environment, to date. The time for task completion is said to be a reliable and powerful predictor of learning and performance [172]. [173] used time-on-task to measure experience in a computer aided instruction learning environment.

Furthermore, background knowledge is considered an influential performance factor, because a student's historical data, such as pre-laboratory test score and pre-requisite unit test/exam score, among others, are indicative of prior learning with respect to the concept addressed by the laboratory activities and are therefore expected to impact on their performance. In addition, a memory/feedback node was introduced into the LAP model to facilitate incorporating previous performance estimates into current belief estimates, if desired.

Figure 4.3 represents a graphical view of the LAP assessment model. The variables (nodes) in the model have been named to reflect, as much as possible, what they are intended to represent. The nodes have been numbered (the numbers above the blocks for the nodes in Figure 4.3) for ease of reference. The nodes are categorized as follows:

- *performance* (node 40): the main target node of inference that gives a measure of belief in (a global view of) a student's performance with respect to laboratory activities.
- *memory/feedback* (node 35): creates a feedback loop so that previous performance belief estimates can be taken into consideration in the current estimation for belief in performance;

- *performance factors* (nodes 24, 37, 39): variables constituting the broad framework of learning that directly influence performance.
- *High Level Abilities/Skills (HLAS)* (nodes 15, 16,17,18,22): the identified key laboratory tasks that are targets of inference for a task-level view of a student's abilities/skill;
- *performance indicators* (nodes 4, 7, 10, 11, 14, 17, 18, 21, 22): BACs that are targets of inference for a deeper and more robust view of a student's abilities/skills.
- *other influential factors* (nodes 29, 32, 33): other extraneous factors that can influence performance;
- *intermediate nodes* (nodes 23, 34): used for simplifying the CPTs of affected nodes (nodes 24 and 40) and for the purposes of clarity. For example, nodes 18, 21, and 22 (*APG*, *SRC*, and *IER*, respectively) have been grouped together under the umbrella of the intermediate node, "*GenericAbilitiesAndSkills (GAS)*" (node 23), for reasons of simplifying the CPT of node 24 and also because they cut across a wide range of other disciplines and are therefore generic.
- *evidence variables (performance indices or criteria)* (nodes 1, 2, 3, 5, 6, 8, 9, 12, 13, 17, 18, 19, 20, 22, 25, 26, 27, 28, 30, 31, 33, 36, 38): includes the set of LLBACs to which some of the HLAS could be factored, and other variables, which are the observable attributes of behaviour, in addition to the evidential attributes from students historical data. These serve as evidence (findings) for network update.

These node categories are highlighted graphically in Figure 4.4, which gives a more compact representation of Figure 4.3.

The HLAS that are not amenable to factorization are those that are directly observable from behaviour in which case they also serve as performance indicators (because they contribute to the robust view of a student's abilities/skills) and indices (because they are directly observable from behaviour and do not need to be inferred) namely, nodes 17, 18, and 22.

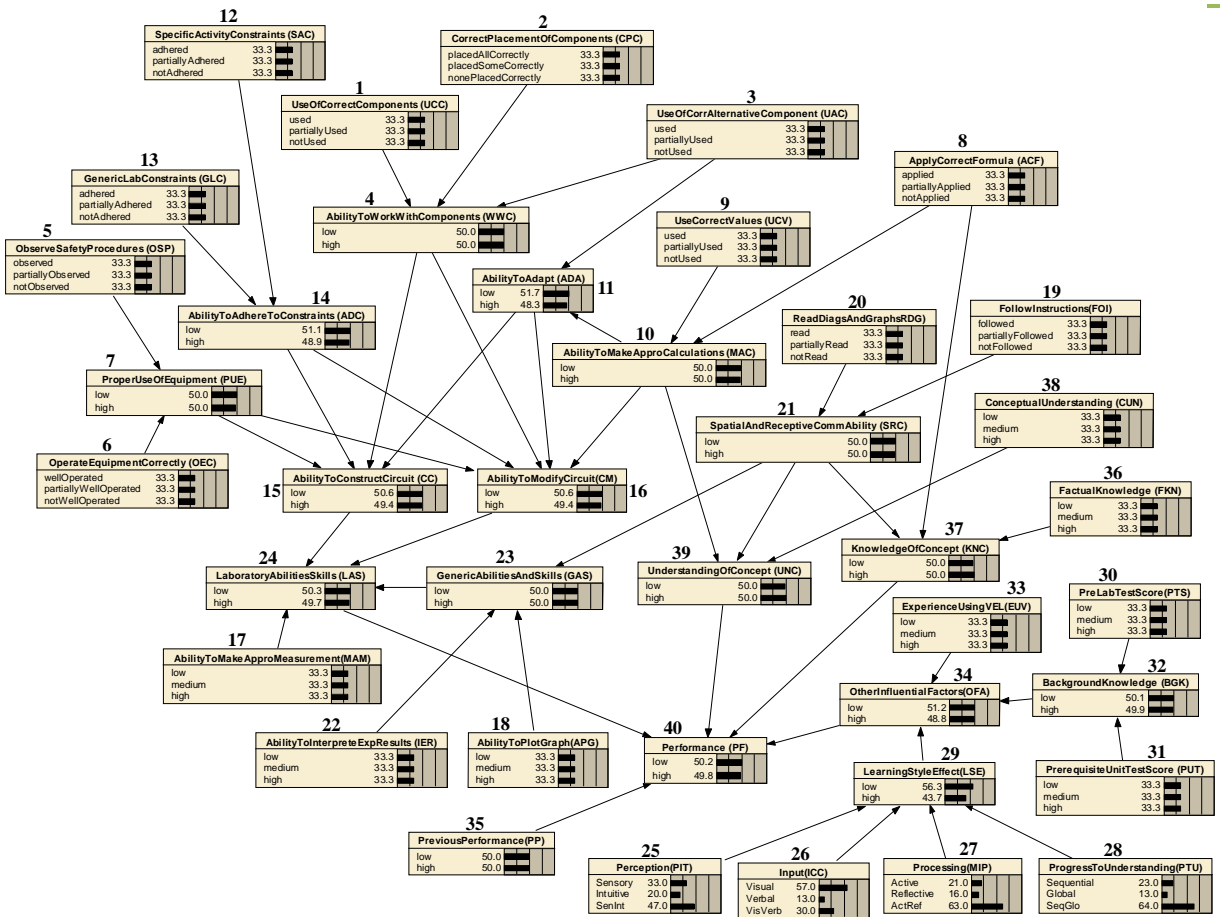


Figure 4.3: Graphical view of the LAP model.

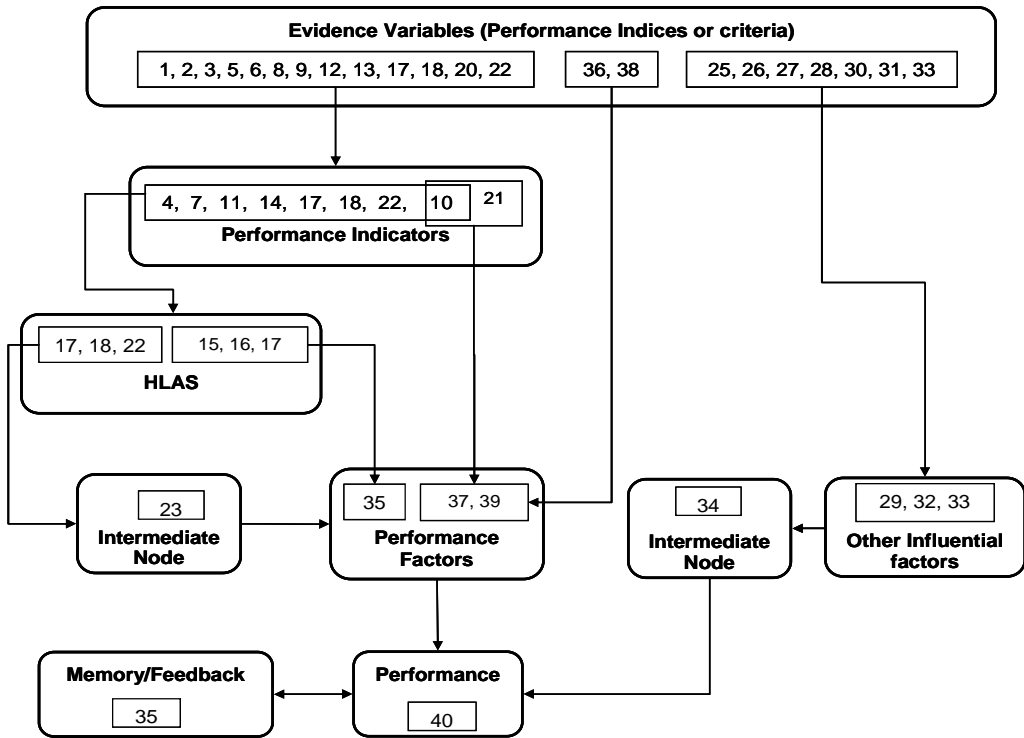


Figure 4.4: Categorization of the LAP model nodes.

4.3 CALIBRATION

After building the model structure, the network has to be calibrated because the model is a measurement tool. Measurement, in this context, is taken to mean a repeatable procedure for assigning metrics to hypotheses or phenomena and calibration (parameterization) is the elicitation of parameter values and quantification of the CPTs and PPTs of the nodes in the model. A model can be parameterised using:

- domain expert knowledge [95],
- data [[174][175],
- a combination of domain expert knowledge and data [89][176].

The data could be historical or empirical. It may be expensive to conduct empirical studies for every student laboratory performance related cognitive ability/skill in order to obtain data for calibrating the model, and lack of existing historical data, cost and time constraints make the use of domain expert knowledge more attractive. [165] assert that domain knowledge is the most suitable alternative in a situation where there are no existing structures, data, and methods, and where empirical means is economically infeasible.

Furthermore, expert knowledge is the best approach to parameterization, in this context, because it is the expert domain knowledge that makes the model an expert system, so eliminating or limiting expert involvement may undermine the overall “expert” quality of the model [177]. Moreover, a BN model for assessment should be expected to rely on calibration by domain experts [177]. There are key benefits to be derived from domain expert calibration, as highlighted in Chapter 5 (section 5.1.1). However, domain expert model calibration is not without challenges, as also highlighted in Chapter 5 (section 5.1.1). One of the main challenges is said to be the difficulty of eliciting expert knowledge. Literature clearly outlines strategies for addressing the challenges which have proven effective. For example, Guidelines for easing the knowledge elicitation process, and elicitation methods (including the Number Line method—a visual scale), have been outlined by [178][179][180].

4.3.1 Knowledge Elicitation

The number line knowledge elicitation instrument, shown in Figure 4.5, was modified as part of this work, and used for knowledge elicitation.

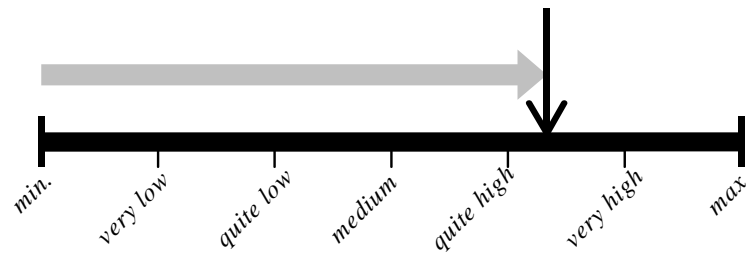


Figure 4.5: Number line knowledge elicitation instrument
(Source: [178])

The linguistic terms were replaced with numbers from 0 to 10. Domain experts were required to rate the importance/relevance of the impact of each parent node to its child node (represented by the directed edge from the parent to the child), on the scale. This is referred to as link rating. A rating of zero meant zero importance/relevance which implies the expert does not believe that a relationship exists between the two nodes. This gave the experts a chance to suggest modifications to the network structure, if they have had a rethink since the last discussion on the design of the network structure.

4.3.2 The Calibration Process

The data for network calibration was obtained from the analysis of the elicited expert knowledge, the link ratings, because most often, expert knowledge does not come in the form of raw probabilities. Also, the values in the CPT of a node reflect the strength of the links with its parent nodes. The ratings were first transformed into percentage link weights from which percentage probability values were derived for the CPTs. For example, given the simple BN of Figure 4.6, if the parent-child links, a, b, c, and d, are given the following importance/relevance ratings by a domain expert: a = 5, b = 3, c = 1 and d = 7 (note: a, b, c, and d do not necessarily have to add up to 10. Rather each can have any value between 1 and 10).

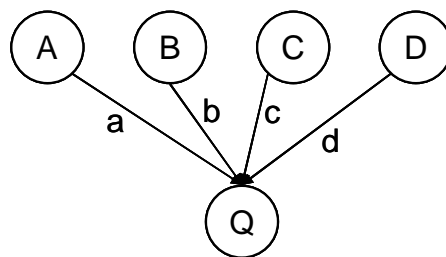


Figure 4.6: A Simple five node BN.

The percentage link weights, w_a , w_b , w_c , and w_d are derived thus: Let $y = a + b + c + d$ and $x = 100/y$. y is the sum of the importance/relevance ratings for all the parent node links to node Q in Figure 4.6. x is the multiplier, derived using y , for the importance/relevance ratings, a , b , c , and d , for the parent node links, A-Q, B-Q, C-Q, and D-Q, respectively (Figure 4.6), such that $x(a+b+c+d)=100$. If, for example, $y = 16$, then $x = 100/16 = 6.25$, then $ax = w_a = 31.25\%$, $bx = w_b = 18.75\%$, $cx = w_c = 6.25\%$, $dx = w_d = 43.75\%$. The requirement is that $w_a + w_b + w_c + w_d = 100\%$ (for percentage probability). Once the percentage link weights are calculated, assuming that the parent nodes A, B, C and D each have two possible states (e.g. High (H) and Low (L)), the CPT for the child node, Q , is parameterized as shown in Table 4.3. Thus, the CPT for each child node is built, with respect to its parents, by summing the contributed percentage link weights of its parents [91].

Table 4.3: Parameter assignment.

A	B	C	D	Q	
				H	L
H	H	H	H	$w_a + w_b + w_c + w_d$	0+0+0+0
H	H	H	L	$w_a + w_b + w_c + 0$	0 + 0 + 0 + w_d
H	H	L	H	$w_a + w_b + 0 + w_d$	0 + 0 + $w_c + 0$
H	H	L	L	$w_a + w_b + 0 + 0$	0 + 0 + $w_c + w_d$
H	L	H	H	$w_a + 0 + w_c + w_d$	0 + $w_b + 0 + 0$
H	L	H	L	$w_a + 0 + w_c + 0$	0 + $w_b + 0 + w_d$
H	L	L	H	$w_a + 0 + 0 + w_d$	0 + $w_b + w_c + 0$
H	L	L	L	$w_a + 0 + 0 + 0$	0 + $w_b + w_c + w_d$
L	H	H	H	0 + $w_b + w_c + w_d$	$w_a + 0 + 0 + 0$
L	H	H	L	0 + $w_b + w_c + 0$	$w_a + 0 + 0 + w_d$
L	H	L	H	0 + $w_b + 0 + w_d$	$w_a + 0 + w_c + 0$
L	H	L	L	0 + $w_b + 0 + 0$	$w_a + 0 + w_c + w_d$
L	L	H	H	0 + 0 + $w_c + w_d$	$w_a + w_b + 0 + 0$
L	L	H	L	0 + 0 + $w_c + 0$	$w_a + w_b + 0 + w_d$
L	L	L	H	0 + 0 + 0 + w_d	$w_a + w_b + w_c + 0$
L	L	L	L	0+0+0+0	$w_a + w_b + w_c + w_d$

Having created a CPT for the child node, Q , based on the percent weights of its parents' links, the leaf nodes, A, B, C and D, have no parents and PPTs have to be created for each of them. The PPTs were set to fair priors. A fair prior reflects the belief that all values of a variable are equally likely and therefore that the most likely values for a two-valued variable, A, are as in Table 4.4.

Table 4.4: Fair Prior values for an Observable variable, A.

Variable	$P(\text{high})$	$P(\text{low})$
A	0.5	0.5

Essentially, the CPT for a child node models the interactions between or the combined effects of its parent nodes by building appropriate constraints in the table [181]. The interaction can be compensatory, conjunctive, disjunctive, or inhibitory (see [181] for details). The interactions between the parents of a child node in the LAP model are not modelled as compensatory because a low level on one of the parents of a child node would not be compensated for with a corresponding high level of one or more of the other parents. For example, in Figure 4.3, depicting the graphical view of the LAP model, a LOW on the ability for “*ProperUseOfEquipment*” (*PUE*), node 7, cannot be compensated for by a HIGH on the “*AbilityToWorkWithComponents*” (*WWC*), node 4, or the “*AbilityToAdapt*” (*ADA*), node 11, and so on. This is because abilities differ from each other and are mutually exclusive as each supports a specific aspect of performance. For example, looking at Figure 4.3 again, a student rated high on circuit construction proficiency, “*AbilityToConstructCircuit*” (*CC*), is not only able to construct a workable circuit whose overall behaviour meets given specifications, but has used equipment correctly; made appropriate measurements; adapted where necessary; adhered to constraints; and worked well with components. Working with components requires the student to select and use the correct components, place the components well in the circuit, and used correct alternative components where necessary.

The interactions in the model are modelled as conjunctive in which a high level status of the child node requires high levels on all its parent nodes. Table 4.5 shows the CPT for node 4, “*AbilityToWorkWithComponents*” (*WWC*). Line 7 in Table 4.5 indicates that if a student, while building a circuit, used the correct components, did not correctly place the components, but used correct alternative components, where necessary, there is a 70% chance that the student is able to work with electronic components. But if the student only partially used correct alternative components, then there is only 55% chance that the student is able to work well with components. The states of the unobservable nodes (the proficiency variables), indicate what a student is actually able to do with respect to laboratory activity performance. The parameter values in the CPTs and PPTs are used for belief estimation.

Table 4.5: The CPT for Node 4 of Figure 4.3: “*AbilityToWorkWithComponents*”

Node: AbilityToWorkWithComp			Apply	Okay
Chance	% Probability		Reset	Close
UseOfCorrectComponents (UCC)	CorrectPlacementOfComponent...	UseOfCorrAlternativeComponen...	low	high
used	placedAllCorrectly	used	10	90
used	placedAllCorrectly	partiallyUsed	13.5	86.5
used	placedAllCorrectly	notUsed	27	73
used	placedSomeCorrectly	used	16.5	83.5
used	placedSomeCorrectly	partiallyUsed	30	70
used	placedSomeCorrectly	notUsed	43.5	56.5
used	nonePlacedCorrectly	used	33	67
used	nonePlacedCorrectly	partiallyUsed	46.5	53.5
used	nonePlacedCorrectly	notUsed	60	40
partiallyUsed	placedAllCorrectly	used	20	80
partiallyUsed	placedAllCorrectly	partiallyUsed	33.5	66.5
partiallyUsed	placedAllCorrectly	notUsed	47	53
partiallyUsed	placedSomeCorrectly	used	36.5	63.5
partiallyUsed	placedSomeCorrectly	partiallyUsed	50	50
partiallyUsed	placedSomeCorrectly	notUsed	63.5	36.5
partiallyUsed	nonePlacedCorrectly	used	53	47
partiallyUsed	nonePlacedCorrectly	partiallyUsed	66.5	33.5
partiallyUsed	nonePlacedCorrectly	notUsed	80	20
notUsed	placedAllCorrectly	used	40	60
notUsed	placedAllCorrectly	partiallyUsed	53.5	46.5
notUsed	placedAllCorrectly	notUsed	67	33
notUsed	placedSomeCorrectly	used	56.5	43.5
notUsed	placedSomeCorrectly	partiallyUsed	70	30
notUsed	placedSomeCorrectly	notUsed	83.5	16.5
notUsed	nonePlacedCorrectly	used	73	27
notUsed	nonePlacedCorrectly	partiallyUsed	86.5	13.5
notUsed	nonePlacedCorrectly	notUsed	90	10

4.4 OVERVIEW OF THE MODEL BELIEF ESTIMATION PROCESS

The main reason for using BNs is to enable reasoning under uncertainty. Uncertainty in this context refers to students’ behaviour with respect to problem scenarios. This implies their knowledge and understanding, demonstrations of relevant abilities/skills, and performances, are uncertain and can only be inferred from observed behaviour. For example, in Figure 4.7 (a subset of the LAP model, where the nodes *UCV*, *ACF*, *MAC*, and *KNC* correspond to nodes 9 (“*UseCorrectValues*” (*UCV*)), 8 (“*ApplyCorrectFormulae*” (*ACF*)), 10 (“*AbilityToMakeAppropriateCalculations*” (*MAC*)), and 28 (“*KnowledgeOfConcept*” (*KNC*)), respectively, in the LAP model (see Figure 4.3), the expressions, $p(KNC = high)$ and $p(MAC = high)$, referred to as marginal probabilities, represent measures of beliefs in the hypothesis represented by the variables. It cannot be stated with certainty that a student has knowledge of the addressed concept(s) or that he/she has applied the correct formulae for a laboratory activity yet to be undertaken. The marginal probabilities can be evaluated using the parameter entries in the CPTs and PPTs of the nodes. For example,

$$p(KNC = high) = p(KNC = high | ACF = applied) \times p(ACF = applied) + p(KNC = high | ACF = partiallyApplied) \times p(ACF = partiallyApplied) +$$

$$p(KNC = high | ACF = notApplied) \times p(ACF = notApplied) \quad (4.1)$$

All the values required to evaluate this expression are derived from the parameter entries in the CPT of the node, *KNC*, and PPT of the node, *ACF*. For the purposes of this example, it is assumed that there is no other node with a link to the node, *KNC*.

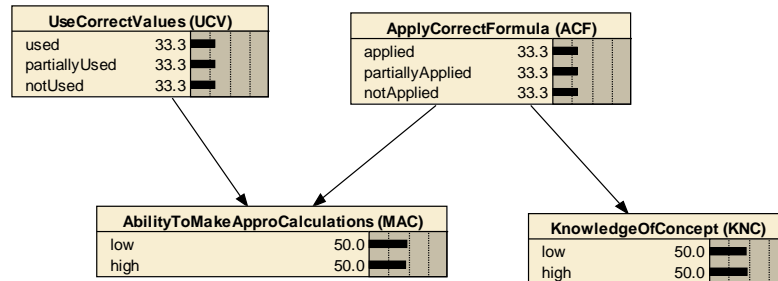


Figure 4.7: A sub-model of the LAP Model

If after undertaking the laboratory activity, it is evidenced that the student had applied the correct formulae ($ACF = applied$) and used the correct values ($UCV = used$) for the specified calculation, the evidence is entered into the model by clamping nodes *ACF* and *UCV* to their evidenced or instantiated states. It still, at this point, does not mean with certainty that the student definitely made appropriate calculations or that he/she definitely has knowledge of the addressed concept(s). However, there is increased probability (belief) that the student made appropriate calculations, and that he/she has knowledge of the addressed concept(s). In this case, the value of the expression, $p(KNC = high | ACF = applied)$, referred to as conditional probability, is directly obtained from the CPT of node *KNC*. That is, the measure of believe in the student's knowledge of the addressed concept is revised after receiving the evidence.

The most important use of BNs is in revising measures of belief in the light of actual observations of events. Suppose that it was not known (not evidenced) that the student applied the correct formulae, probably because the given laboratory activity did not include a task that would elicit the necessary behaviour from the student to facilitate the derivation of the evidence, but the instructor had ascertained, through some other means of assessment (e.g. physical observation), that the student definitely has knowledge of the addressed concept. This information can be entered into the model by clamping the node, *KNC*, to *high*, and then used to determine the revised probability that the student; applied the correct formulae, $p(ACF = applied | KNC = high)$; made appropriate calculations, $p(MAC = high)$; and used correct values, $p(UCV = high)$. In this case, $p(ACF = applied | KNC = high)$ is

calculated using Bayes formula, values from the CPT and PPT of nodes *ACF*, and *KNC*, and the marginal probability of *KNC*, $p(KNC = high)$, as:

$$p(ACF = applied | KNC = high) = \frac{p(KNC=high|ACF=applied) \times p(ACF=applied)}{p(KNC=high)}$$

The calculation of the revised marginal probability of *MAC*, $p(MAC = high)$, will depend on the value derived for $p(ACF = applied | KNC = high)$ and no longer on the value of $p(ACF = applied)$ in the PPT of node *ACF*. Thus, any evidence or information received is used to revise the measures of belief for all affected hypotheses. This is referred to as network update or evidence/information propagation.

4.5 SUMMARY

The construction process of the LAP model, with the assistance of domain experts, has been described from structure realization to calibration. The model belief estimation process has also been highlighted. Though the assistance of domain experts were used for the construction of the model, BN-based models can also be constructed from data. The next chapter investigates the optimal approach for the construction of the assessment model for the performance assessment of students' laboratory work in the VEL environment, with respect to expert-centred and data-centred BN construction approaches. That is, to empirically investigate if the expert-centred approach is the best approach for constructing the LAP model, or if it is possible to derive an improved or better model using the data-centred BN model construction approach.



CHAPTER 5

INVESTIGATION OF THE OPTIMAL CONSTRUCTION APPROACH FOR THE BAYESIAN NETWORK-BASED ASSESSMENT MODEL

*The whole difference between construction and creation is exactly this: that a thing
constructed can only be loved after it is constructed; but a thing created is loved
before it exists.*

Charles Dickens

1812 - 1870

5.0 INTRODUCTION

A BN -based model (the LAP model) for performance assessment of students' laboratory work, in the VEL environment, was constructed with the assistance of domain experts. The details of its construction process and evaluation are given in Chapter 4. Though BN model constructors (researchers working with BN models), ab initio, relied only on domain experts to define both the structure and parameters of a model (which line was toed in constructing the LAP model), currently algorithms exist to construct BN models from data. Hence, BN model construction can be categorized under two approaches: expert- and data- centred. Consequently, there are three techniques to BN model (structure + parameters) construction:

- *total expert-centred (tecen)*
- *total data-centred (todacen)*
- *semi data-centred (sedacen).*

In *tecen* approach, the BN model is a product of domain analysis, whereby domain expert(s) completely specify both the qualitative and quantitative components of the model. The *todacen* approach uses algorithms to generate both the qualitative and quantitative components of a BN model from data. The generation of the qualitative component from data is referred to as *structure learning*, and the generation of the quantitative component referred to as *parameter learning*. The *sedacen* approach is a hybrid framework whereby domain experts assist in the creation of the qualitative component of a BN model, while the quantitative component is learnt from data.

Having constructed the LAP model with the assistance of three domain experts, it was necessary to find out if the LAP model (a *tecen* model) is the best or optimal model for the assessment of students' laboratory work in the VEL environment. That is, to empirically investigate if the expert-centred approach is the best approach for constructing the LAP model, or if it is possible to derive an improved or better model using the data-centred BN model construction approach. This required the construction of *sedacen* and *todacen* models from data, based on the same set of domain variables as the *tecen* model. The aim is to compare the performances of the *sedacen* and *todacen* models to the performance of the *tecen* LAP model (the reference model), based on a set of performance metrics.

There are two possible sources of sample datasets that can be used for the construction of the *sedacen* and *todacen* models. First, sample domain historical sample dataset(s) on students' laboratory work performance assessment, with respect to a set of performance indicators and their respective criteria, could be used if available. This option was not possible as there are

no existing historical sample datasets, for the domain of engineering students' laboratory education (with respect to performance-based assessment of students' laboratory work), to the best of the author's knowledge. The second option is the use of simulated sample datasets. Often, researchers needing to undertake empirical investigations, with respect to structure and/or parameter learning, create frameworks that would allow them to generate the required sample datasets from a reference model. This approach has been adopted by a number of researchers including [181][183][184][185]. The procedure starts with an existing model (the reference model), generates datasets from the Joint Probability Distribution (JPD) represented by the model, and then uses learning algorithms to attempt to retrieve the reference model from the datasets. The retrieved (learnt) model is then compared with the original model that generated the dataset [185]. This procedure is commonly used for evaluating learning algorithms, and was deemed appropriate for this empirical investigation because of lack of existing historical sample datasets. The idea is that if, using the above procedure, the algorithms fail to retrieve (induce) the reference model, a comparable model, or a better model (in terms of performance) from the sample datasets generated using the JPD encoded by its structure, then it may imply that the algorithms will also fail to induce a comparable model, or a better model from sample datasets generated from other sources. That is, failing to retrieve the reference model, the algorithms should at least learn a model whose performance is comparable to or better than that of the reference model. It is assumed that the reference model is the existing optimal model. If the algorithms learn a model whose performance is significantly better than that of the reference model, then the learnt model is taken to be the optimal model, else the reference model is taken to be the optimal model. Optimal, in this context, refers to the model which is best in terms of the adopted optimality criteria [186].

First, section 5.1 details the different BN model construction approaches. Section 5.2 describes the procedure for the empirical investigation, highlighting how the different *sedacen* and *todacen* models used in the investigation were constructed, and the model test process, while section 5.3 highlights the criteria and comparative tools used to compare the models. The results of the investigation are presented in section 5.4, while the observations and attendant discussion are given in section 5.5. The chapter is summarised in section 5.6.

5.1 BAYESIAN NETWORK MODEL CONSTRUCTION APPROACHES

5.1.1 Expert-Centred Approach

“Manually” building a BN model involves three ordered tasks: identification of the network variables and their possible values (states), definition of the relationships between the variables for graphical representation, and model calibration (parameterization). There are no formal foundations for “manual” BN model construction, and the process is still essentially an art [187]. It depends on the model constructor to use suitable techniques and tools for undertaking the knowledge elicitation tasks. Expert-centred BN model construction approach offers a number of benefits:

- the model embeds reasonably accurate domain knowledge because it is built interactively with, often, more than one expert. The, model variables, their states, and relationships are fully appreciated. Hence, the reasoning and rationale behind the BN model can be clearly articulated and communicated.
- model creation is often based on the consensus or average of information and opinions of more than one domain expert, thereby enabling the capture of uncommon or rare scenarios and knowledge.
- the technicalities of the domain represented by the model can be verified/discussed in details at each stage of the development cycle.
- expert probability elicitation codifies knowledge so that the knowledge is available in the future for other projects and systems thereby promoting reliability in assessment of a family of systems that change within a changing usage environment [187].

However, knowledge elicitation is said to be a major challenge of the expert-centred BN model construction approach because it is difficult to elicit expert knowledge, which is often biased, and experts rarely agree. Guidelines for easing the elicitation process and elicitation methods have been outlined by [178][179][180]. Experts’ opinion disagreement is generally acknowledged [189]. Methods for resolving expert opinion conflicts and how to obtain composite or consensus opinion are addressed by [189]. Also, [165] argued that the issue of bias no longer holds as a range of techniques and tools that minimize the effort required for probability elicitation have been developed. In addition, the issue of bias is often addressed through the involvement of more than one domain expert and the knowledge elicitation process often goes through review stages, after which the model is subjected to sensitivity analysis. Moreover, BNs have been found not to be too sensitive to inaccuracies in their

parameters [190], so determining good parameter values in many application areas is quite feasible [106].

5.1.2 Data-Centred Approach

Let $B = \langle G, \theta \rangle$ be a discrete BN model, where G is the network structure with nodes corresponding to the set of random variables, $X = (X_1, \dots, X_m)$, in the focus domain, and θ represents the set of parameters for the network. B encodes the JPD $p(X_1, \dots, X_m) = \prod_{i=1}^m p(x_i | pa(x_i))$, where $pa(x_i)$ represents the parent set of node x_i . The probability distribution, $P(x_i | pa(x_i))$, for each discrete node, X_i , is represented as a CPT at node X_i in B . The data-centred approach entails learning the structure, G , and/or the parameters, θ , from a given sample dataset.

A dataset, D , is a table (matrix) consisting of records of observations (findings) for the network variables, such that $D = [d_1, d_2, \dots, d_N]$, where $N =$ total number of records in D , and $d_l = \{x_1[l], x_2[l], \dots, x_m[l]\} \in D$, $l = 1$ to N , represents a record of observation (a single set of findings or instantiation of all the variables, X). A form of a dataset is shown in Figure 5.1

IDnum	Ability To Make Appro Measurements	Ability To Work With Components	Use Of Cor Alternative Component	Ability To Construct Circuit	Ability To Modify Circuit	Ability To Interpret Exp Results	Ability To Adapt	Ability To Make Appro Calculations	Use Correct Values	Apply Correct
493	low	high	used	high	high	medium	high	high	partiallyUsed	applied
494	high	low	notUsed	low	low	high	low	high	notUsed	applied
495	low	high	used	high	low	medium	low	low	notUsed	partiallyUsed
496	medium	high	partiallyUsed	low	high	medium	high	high	notUsed	applied
497	medium	low	notUsed	high	high	low	high	high	used	notApplied
498	low	high	notUsed	high	high	low	low	low	notUsed	notApplied
499	high	high	notUsed	high	high	medium	high	low	notUsed	notApplied
500	high	high	used	high	low	high	low	low	notUsed	notApplied
501	medium	high	used	low	low	high	low	low	partiallyUsed	notApplied
502	medium	high	partiallyUsed	low	high	low	low	low	partiallyUsed	applied
503	high	low	partiallyUsed	low	low	low	low	low	notUsed	notApplied
504	high	high	used	high	low	low	high	low	notUsed	applied
505	high	high	used	high	high	medium	low	low	used	partiallyUsed
506	low	low	used	high	high	low	high	high	partiallyUsed	partiallyUsed
507	high	high	used	high	high	high	low	high	notUsed	applied
508	high	low	notUsed	high	low	high	high	high	used	applied
509	high	high	partiallyUsed	low	high	high	low	high	partiallyUsed	applied
510	high	high	used	high	high	high	high	high	used	notApplied
511	medium	high	partiallyUsed	low	low	low	low	low	partiallyUsed	notApplied
512	medium	high	used	low	low	low	low	low	notUsed	partiallyUsed
513	high	high	partiallyUsed	high	high	medium	high	high	used	partiallyUsed
514	low	low	partiallyUsed	low	high	high	low	low	partiallyUsed	partiallyUsed
515	medium	low	partiallyUsed	low	high	high	low	low	used	notApplied
516	medium	low	partiallyUsed	low	low	high	low	low	partiallyUsed	notApplied
517	low	high	partiallyUsed	high	high	medium	high	high	partiallyUsed	partiallyUsed
518	low	high	used	high	high	high	high	high	partiallyUsed	applied
519	high	high	notUsed	high	high	medium	high	high	used	applied
520	low	low	used	low	low	low	low	low	notUsed	notApplied
521	high	low	notUsed	low	low	low	low	high	partiallyUsed	applied
522	low	high	used	high	high	medium	high	high	partiallyUsed	applied
523	high	low	used	low	low	low	low	low	used	partiallyUsed
524	high	low	partiallyUsed	low	high	high	low	low	notUsed	notApplied
525	high	low	partiallyUsed	low	low	medium	high	low	used	notApplied
526	medium	low	partiallyUsed	low	high	high	high	low	notUsed	applied
527	medium	low	partiallyUsed	low	high	low	low	low	notUsed	notApplied
528	medium	high	used	high	high	high	high	low	notUsed	applied
529	low	high	partiallyUsed	high	high	medium	high	high	used	notApplied
530	medium	low	partiallyUsed	low	low	medium	high	low	notUsed	notApplied
531	high	high	used	low	high	medium	high	high	used	partiallyUsed

Figure 5.1: A form of a dataset

A dataset can be complete or incomplete because incompleteness (missingness) is sometimes a feature of datasets. A complete dataset contains No Missing Values (NMV) (no missing instantiation or finding) for any of the variables, implying full observability. An incomplete dataset contains missing values for some variables in some or in all the cases in the dataset, implying partial observability or presence of latent (hidden) variables, respectively. Percentage missingness refers to the percentage of data values that are missing in a dataset.

5.1.2.1 *Parameter Learning*

In parameter learning, the structure, G , is known and the problem is to learn the parameter, θ , from the given dataset, D (a sample set of records of observations of X). That is, the estimation of $\theta = \{\theta_i\}_{i=1,\dots,m}$, where θ_i is the set of numerical entries in the CPT of node X_i from D , given G . θ is the complete set of parameters that can best explain the set of observations, D [191]. Parameter learning could be single or multiple (multinomial) parameters learning.

Single Parameter Learning implies that the variable, X_i , whose parameters are to be estimated, has only two possible mutually exclusive states denoted, x_i and \bar{x}_i , such that the probability mass function $p(X_i)$ is defined by: $p(X_i = x_i) = \theta_i$ and $p(X_i = \bar{x}_i) = 1 - \theta_i$.

Multinomial Parameter Learning implies that X_i is a multinomial variable with $r > 2$ possible states, x_{i1}, \dots, x_{ir} , such that X_i has the set of probabilities, $\theta_i = (\theta_{i1}, \dots, \theta_{ir})$, respectively, where $\sum_{k=1}^r \theta_{ik} = 1$.

5.1.2.2 *Structure Learning*

Given the dataset, D , the structure learning problem is to find, using D , the most probable network structure, G_i , from among the set of possible network structures, $\Phi = (G_1, G_2, \dots, G_\lambda)$, where λ is the cardinality of the search space. That is, discover the BN structure that most likely generated D . That is, G_i is the network that best describes the conditional independences suggested by the given dataset, D [174]. This is often referred to as *model selection* in literature, which term will not be used in this thesis because, in this

context, BN model refers to a complete Bayesian network (structure + parameters). Once G is found, its parameters, θ , are derived as described in section 5.1.2.1.

Structure learning algorithms are either based on Conditional Independence (CI) tests or Search and Score (SaS). The CI approach uses *constraint-based* algorithms to find the structure whose implied independence constraints “match” those found in the data by performing CI tests on tuples of variables, using *statistical tests* or *information theoretic* measures [192]. CI-based algorithms include the PC algorithm by [193]. The SaS approach consists of three components: the search space, the scoring function, and the search engine. The *search space* consists of the set of all possible BN structures, Φ , given the domain variables. The main operation in the search space is the modification of one structure to produce another structure with the operators “add an edge”, “delete an edge”, and “reverse an edge” [194].

The *score metric* takes the dataset and a possible structure and returns a score reflecting the goodness-of-fit of the data to the structure [195]. There are two categories of scoring functions: Bayesian and information-theoretic. The information-theoretic score functions include: the Log-likelihood (LL) [196], Minimal Description Length (MDL) [197], Akaike Information Criterion (AIC) [198], and the Bayesian Information Criterion (BIC) [199]. The MDL is said to be equivalent to the BIC function; hence they are often written as MDL/BIC. The Bayesian scoring metrics include: Bayesian Dirichlet (BD) [200], likelihood-equivalence Bayesian Dirichlet (BDe) [200], the uniform joint distribution Bayesian Dirichlet (BDeu) [201], and the K2 [182]. The K2 has been described as one of the most successful scoring metrics [202].

The *search engine* (*search algorithm*) works to identify structures with high scores by exploring the search space. It makes comparisons of network structures as it searches heuristically for the most likely structure [203]. Essentially, the dataset D , the scoring function, and the search space constitute the inputs to the search algorithm while the output is a network that maximizes the score, $P(D|G_i)$, the probability of the most probable structure, G_i , given the dataset, D [204].

One of the main challenges of the data-centred approach is that structure learning is NP-hard [205]. Researchers have attempted to reduce the complexity of BN structure learning by various algorithmic means, but the problem remains complex and hard, without exact and

exhaustive solution [192]. Consequently, heuristic algorithms are often employed for the learning process. The latter help produce an acceptable solution to a problem in many practical scenarios, though it is not certain to arrive at an optimal solution. It begins with an approximate method of solving the problem within the context of the goal, and then uses feedback from the solution to improve its performance, searching for a satisfactory solution rather than optimal solution.

The complexity of the search space is another major challenge of structure learning because the number of possible structures grows super-exponentially with the number of variables, n , in the problem domain [192]. For, n variables, the cardinality of the search space is given by [206] as the recursive function:

$$f(n) = \sum_{k=1}^n (-1)^{k+1} \binom{n}{k} 2^{k(n-k)} f(n-k), \text{ where } f(1) = 1 \quad (5.1)$$

Table 5.1 lists the possible number of BN structures for some values of n .

Table 5.1: Number of possible BN structures for various numbers of variables.
(source: [192])

No of variables, n	No of possible BN structures
1	1
2	3
3	25
4	543
5	29,281
6	3,781,503
7	1,138,779,265
8	78,370,2329,343
9	1,213,442,454,842,881
10	4,175,098,976,430,598,100

This super-exponential relationship between the number of variables and the number of possible structures is a major source of computational complexity [207].

5.2 INVESTIGATION PROCEDURE

First, sample datasets were generated with the reference model in such a way that the investigation was undertaken from two different approaches referred to, in this context, as *non-parameteric* and *parameteric* approaches. The *non-parameteric* approach entailed the use of the unparameterised version of the reference model (structure without the elicited parameters) to generate some of the sample datasets for the investigation. The *parameteric*

approach entailed the use of the parameterised version of the reference model (structure with elicited parameters) to generate some of the sample datasets for the investigation. A set of four training (learning) datasets were generated using each of the two approaches. Hence, there were a total of eight training datasets, $D_{np} = (D1_{np}, D2_{np}, D3_{np}, D4_{np})$ and $D_p = (D1_p, D2_p, D3_p, D4_p)$. The np and p at the end of the dataset names indicate *non-parameteric* and *parameteric* approaches to the generation of the datasets, respectively. Models constructed with training datasets generated using the non-parameteric approach are referred, in this context, as *non-parameteric models*, while models constructed with training datasets generated using the *parameteric* approach are referred as *parameteric models*. Also, a set of four test datasets, $T = (T1, T2, T3, T4)$, different from the training datasets, were generated for the evaluation of the models. Several training datasets were used to facilitate the acquisition of meaningful data for the intended analysis.

The BN software tools, Genie [208], Netica [209], SamIam [210], and WinMine Toolkit [211] were used for the investigation. Genie is the graphical interface to SMILE [208], a Bayesian inference engine. Netica is a complete program for working with belief networks. SamIam (Sensitivity Analysis Modeling Inference And More) is a tool for modelling and reasoning with Bayesian networks, developed in Java, and includes a GUI for editing Bayesian networks.

5.2.1 Construction of the Data-centred Models used for the Investigation

A total of twenty four (24) *sedacen* models were constructed, consisting of three sets of four *non-parameteric* models, $PND_{np} = (PND1_{np}, PND2_{np}, PND3_{np}, PND4_{np})$, $PGD_{np} = (PGD1_{np}, PGD2_{np}, PGD3_{np}, PGD4_{np})$, and $PSD_{np} = (PSD1_{np}, PSD2_{np}, PSD3_{np}, PSD4_{np})$; and three sets of four *parameteric* models, $PND_p = (PND1_p, PND2_p, PND3_p, PND4_p)$, $PGD_p = (PGD1_p, PGD2_p, PGD3_p, PGD4_p)$, and $PSD_p = (PSD1_p, PSD2_p, PSD3_p, PSD4_p)$. 24 *sedacen* models were constructed because there were 8 training datasets (4 *parameteric* and 4 *non-parameteric*) and 3 software tools, whereby each software tool is used construct *sedacen* models from each of the 8 datasets. The model names indicate the type of learning undertaken, and the software tool and the dataset used for the construction of the model. For example, $PND1_{np}$ indicates that the model was constructed by parameterizing, using Netica, a known structure from the training dataset, $D1_{np}$. Netica, Genie and SamIam were used for the construction of the *sedacen* models, based on the Expectation Maximization (EM) algorithm,

implemented by all three software tools. The plan to also undertake parameter learning using Gibbs sampler, for comparative purposes, was jettisoned because, according to [212], EM and Gibbs sampler are substantially equivalent in their parameter estimates.

It was possible to use the same sets of datasets with the three different software tools, Netica, Genie, and SamIam, because they all support text data file formats, albeit with different file extensions (.cas, .txt/.dat, and .dat, respectively). All that was required was to save the data file with the appropriate file extension for each software tool and edit as may be necessary (for example, Netica requires the inclusion of the row number for each record of the data file, while Genie and SamIam do not). Also, all the software tools used similar network file formats (.dne, .dnet, and .net) so it was possible to open any of the networks with any of the tools.

It was aimed to construct a total of thirty two (32) *todacen* models from the two sets of four training datasets, $D_{np} = (D1_{np}, D2_{np}, D3_{np}, D4_{np})$ and $D_p = (D1_p, D2_p, D3_p, D4_p)$, using Genie and WinMine, but only twelve (12) complete *todacen* models were realised due to the limitations of the software tools which will become highlighted in section 5.4. WinMine supports SaS-based structure learning approach, while Genie supports both CI- and SaS-based. Exploiting this, in order to exact more investigative effort at inducing a better model than the reference model, the *todacen* model construction using Genie was done as per the following three steps:

- a. first, eight *todacen* models were constructed based on the CI approach with the PC algorithm, using each training datasets in the two sets of training datasets. Search time limit was not imposed. This step generated the $SPGaD_{np} = (SPGaD1_{np}, SPGaD2_{np}, SPGaD3_{np}, SPGaD4_{np})$ and $SPGaD_p = (SPGaD1_p, SPGaD2_p, SPGaD3_p, SPGaD4_p)$ groups of models. The model name, $SPGaD1_{np}$, for example, indicates that the model was constructed by learning both structure and parameters, using Genie, this step, a, and the training dataset, $D1_{np}$.
- b. next, another eight models were constructed with the Greedy Thick Thinning (GTT) search algorithm and the K2 score metric, based on the SaS approach, using each training datasets in the two sets of training datasets to get the $SPGbD_{np} = (SPGbD1_{np}, SPGbD2_{np}, SPGbD3_{np}, SPGbD4_{np})$ and $SPGbD_p = (SPGbD1_p, SPGbD2_p, SPGbD3_p, SPGbD4_p)$ groups of models. The model name,

$SPGbD1np$, for example, indicates that the model was constructed by learning both structure and parameters, using Genie, this step, b, and the training dataset, $D1np$.

- c. finally, eight models were constructed with the Greedy Thick Thinning (GTT) search algorithm and the BDeu score metric, based on the SaS approach using each training datasets in the two sets of training datasets to get $SPGcDp = (SPGcD1p, SPGcD2p, SPGcD3p, SPGcD4p)$ groups of models. The model name, $SPGcD1np$, for example, indicates that the model was constructed by learning both structure and parameters, using Genie, this step, c, and the training dataset, $D1np$.

Factors related to the learning algorithms and score metrics, such as maximum adjacency size (for the PC algorithm) and number of parents constraints (for the K2 and BDeu score metrics), were varied, aimed at increasing the chances of inducing a model structure that will yield a better model than the reference model.

Furthermore, two groups of models, $SPWDnp = (SPWND1np, SPWD2np, SPWD3np, SPWD4np)$ and $SPWDp = (SPWD1p, SPWD2p, SPWD3p, SPWD4p)$, were constructed from the two sets of four training datasets, using WinMine Toolkit. The letters, SP , in the model names imply structure and parameter learning, W implies WinMine Toolkit, and $D1np$ and $D1p$ highlight the particular training dataset used for the construction of the model.

5.2.2 Evaluation Method

The reference, *sedacen*, and *todacen* models were evaluated using the set of four test datasets. The test procedure consisted of entering findings at selected evidence nodes of a model, and querying one or more target nodes. The nodes of the model are divided into two sets: evidence and target nodes. Any node can belong to any one of the two sets, for the purposes of the test. It is often preferable to choose as target, the node that in the real context would be target of inference. The values in each record of the test dataset are split into two sets: values for the chosen evidence nodes, and values for the chosen target nodes. The values for the evidence nodes are entered as findings into the network, the network is updated, and inference made at the target nodes. This process is repeated for each record in the test dataset. For each network update, the probability distribution of the target node is recorded and its prediction determined. That is, after each network update, the state with higher belief value (the most likely or maximum likelihood state), based on a cut-off threshold probability, is taken to be the prediction for the target node. For example, for a 50% cut-off threshold probability, of the two states of the target node, the state which belief level is higher than 50% is taken to be the

prediction. The predictions are then compared with the observations (the set of values for the target node in the test dataset taken as the actual “observations”), for each record of the test dataset. If a prediction corresponds to the observation, it is recorded as a success otherwise it is recorded as a failure. The statistics are then collected and used to assess the performance of the model, generating values for the various performance metrics that constitute the optimality criteria.

5.3 OPTIMALITY CRITERIA AND COMPARATIVE TOOLS

5.3.1 Structure Comparison

Structural difference measures are often used to compare the structural differences between an induced model structure and a reference model structure [185]. The comparison may sometimes not take into consideration the orientation of the edges. If the focus is causality, then the orientation of the edges becomes extremely important. Otherwise, the orientation of the edges can be deemphasized as the exact topology of the network may be immaterial and the emphasis is on performance [185]. A causal model is a “Bayesian network with added property that the parents of each node are its direct causes” [213]. This implies an asymmetric relationship between parent and child nodes, such that in the case edge of reversal, the resulting network will not be equivalent in terms of representational ability. In induced non-causal model structures, it is possible to ignore the direction of the reversible edges but not those of the compelled edges. The reversible edges are the edges that occur in the opposite direction in some other DAG that is equivalent (in terms of representational ability) to the current DAG [214]. “If two DAGs encode the same conditional independencies, they are called Markov equivalent. The set of all DAGs can be partitioned into Markov equivalence classes. Graphs within the same class can have the direction of some of their arcs reversed without changing any of the CI relationships. Each class can be represented by a PDAG (partially directed acyclic graph) called an essential graph or pattern. This specifies which edges must be oriented in a certain direction (compelled edges), and which may be reversed. When learning graph structure from observational data, the best one can hope to do is to identify the model up to Markov equivalence” [215].

In this context, ignoring some edge directions (those highlighted by the software tool as reversible), the link statistics of the induced structure are categorized as:

- *correct positive (cp)*-- a link is learnt between two nodes where a link exists between the same two nodes in the reference model (correct link)

- *false positive (fp)*-- a link is learnt between two nodes where a link does not exist between the same two nodes in the reference model (extra link)
- *correct negative (cn)*-- no link is learnt between two nodes where a link does not exist between the same two nodes in the reference model (correct nolink)
- *false negative (fn)*-- no link was learnt between two nodes where a link exists between the same two nodes in the reference model (missing link).

5.3.2 Performance Comparison Metrics

Scoring functions or rules are appropriate for evaluating the performance of probabilistic predictive models [216]. Mathematically convenient scoring rules are most commonly used [217]. These include, *error rate*, *logarithmic (logloss) score*, and *Briers score* [192][216][218][219]. *Sensitivity* is also, often used as a model performance measure [192][219].

The error rate, based on the maximum likelihood state of the target node [209], is a way to analyze model predictions by dividing the number of predictive errors by the number of test cases in the test dataset. It gives the percentage failure rate. It identifies the percentage of the cases in a test dataset for which the network predicted a wrong value for the query node. For example, an error rate of 24% implies that in 24% of the cases for which the test dataset contains a value for the target node, the predictions did not match the observed values.

The logarithmic (logloss) score was suggested by [209], and is defined as follows: let X denote a discrete random variable, with m (mutually exclusive) possible states, $(x_1, x_2, \dots, x_i, \dots, x_m)$, which is to be observed for a sequence of cases, $i = 1, \dots, N$.

Let $p(x_i)$ denote the estimated probability (referred to as the predicted value for the purposes of the test) for the i^{th} state. Suppose the j^{th} state is actually observed, then the particular observation is associated with a logloss score for the j^{th} state given by [174][221]

as: $\ell_j = \log \frac{1}{p(x_j)} = -\log p(x_j)$. Then, by accumulating the scores for the N cases, a total

penalty for the N observations is obtained by: $\ell = \sum_{j=1}^N \ell_j$, and the average logloss score for

the N cases is: $\ell_{avg} = \frac{1}{N} \sum_{j=1}^N \ell_j = \frac{1}{N} \sum_{j=1}^N -\log p(x_j)$. The logloss value lies in the range $[0, \infty]$,

where smaller (lower) values of the score imply better model performance.

The Brier score (b), also referred to as Quadratic Loss (QL) or Mean Squared Error of Prediction (MSEP), measures the accuracy of a set of probability assessments. The Brier score function, as used in BN model performance comparison, is given by [174][192][219][221] as:

$$b = \frac{1}{N} \left[\sum_{i=1}^N \left((1 - 2 \times p(y = c | x_i)) + \sum_{j=1}^k p(y = j | x_i)^2 \right) \right] \quad (5.1)$$

where $p(y = c | x_i)$ is the probability predicted for the actual (observed) state, c , of the target variable, y (the state of y in the particular record of the test dataset), given the evidence variables, x_i ; $p(y = j | x_i)$ is the probability predicted for the j^{th} state of y , given the evidence variables; k is the number of states of the target variable, y ; N is the number of records in the test dataset. The QL is a measure of the average quadratic loss that occurred on each instance in the test dataset. It is averaged over all the records in the test dataset and not only accounts for the probability assigned to the actual (observed) state, but also the probabilities assigned to the other possible states of y . The value of Brier score lies in the range $[0, 1]$, with $b = 0$ indicating higher prediction accuracy, thus better performance.

Sensitivity (also referred to as the recall rate) is a statistical measure of model performance. It measures the proportion (in percentage) of actual values (observations) which are correctly predicted. A sensitivity of 100% means that the model correctly predicted all actual observations for the target variables (100% actual or true positives).

These metrics are often used together, in any one investigation, by researchers [192][219], in order to facilitate the drawing of more robust conclusions. The different metrics, though not complementary, evaluate performance from different perspectives, thereby collectively giving a more robust picture of the performance of a model. Error rate informs on the percentage failure rate of a model, the Brier score gives a measure of the accuracy of the probability estimates made by the model, and sensitivity informs on the percentage success rate of the model. The logloss score is similar to Brier score, however, the logloss score is *local* in that it only depends upon the probability assigned to the particular state and not on any of the probabilities assigned to the other states.

5.4 RESULTS

As stated earlier (in section 5.2), the models constructed for the purposes of this investigation are broadly grouped as *parameteric* and *non-parameteric* models (based on the training

dataset used for the construction model), and the type of model (*sedacen* and *todacen*). Table 5.2 highlights the different sample training and test datasets and their respective sizes.

Table 5.2: Training and Test Datasets

Type of Dataset	Name of set of Datasets	Member Datasets	Sizes of the Datasets
Training	Dp	D1p, D2p, D3p, D4p	24000, 72000, 142000, 240000, respectively
	Dnp	D1np, D2np, D3np, D4np	24000, 72000, 142000, 240000, respectively
Test	T	T1, T2, T3, T4	270, 5000, 35000, 70000, respectively

The sizes of the training datasets were chosen to represent an increasing reasonably spaced size range, for the purposes of the investigation. Full observability was assumed, for the purposes of the investigation. It was also assumed that the data samples, generated using a BN software tool, are representative of the larger set of baselines samples. As highlighted in Table 5.2, four different sizes of test datasets were used for evaluating the models. One of the reasons was to investigate the relationship between model performance and the size of the test dataset. The second reason was for repeatability of the test for evaluating the performances of the models in order to facilitate the drawing of more robust conclusions from the investigation.

The results of the empirical investigation are hereby presented with respect to the models' performance indices. Rather than use the performance metric values (error rate, logloss, Brier score, and sensitivity) individually, for each of the four test instances, to compare the models, a single performance index was derived for each model, based on the metrics. The function, ψ , for calculating the performance index of a model is defined, in this context, is defined as: $\psi = [(100 - e) + (1 - b) * 100 + s + (1 - l) * 100]_{normalized}$, where e = error rate, b = brier score, l = logloss, and s = sensitivity. The function, ψ , takes the values of the performance metrics for a model as input, and yields a performance index for the model, for a test instance. The function assumes equal importance for all the performance metrics. Table 5.3 lists the performance indices for the *parameteric* models, with respect to the four test instances, while Figure 5.2 graphically highlights the average performance indices of the *parameteric* models and that of the Reference model. Also, Table 5.4 lists the performance indices for the *non-parameteric* models, with respect to the four test instances, while Figure 5.3 graphically highlights the average performance indices of the *non-parameteric* models and that of the Reference model.

Table 5.3: Parametric Models: Performance indices

PARAMETERIC MODELS									
Type	Model Group and (Software Tool)	Model	Structure Learning Approach	No of links learnt	Performance Index				
					T1	T2	T3	T4	Average Performance Index
<i>tecen</i>	-	REF	-	-	0.638	0.623	0.633	0.629	0.630
<i>Sedacen</i> (known structure, parameter learning)	PNDp (Netica)	PND1p	-	-	0.633	0.629	0.628	0.630	0.630
		PND2p	-	-	0.633	0.629	0.629	0.630	0.630
		PND3p	-	-	0.633	0.629	0.629	0.630	0.630
		PND4p	-	-	0.638	0.623	0.633	0.629	0.631
	PGDp (Genie)	PGD1p	-	-	0.633	0.629	0.629	0.630	0.630
		PGD2p	-	-	0.638	0.623	0.633	0.629	0.630
		PGD3p	-	-	0.638	0.623	0.633	0.629	0.630
		PGD4p	-	-	0.638	0.622	0.633	0.629	0.630
	PSDp (SamIam)	PSD1p	-	-	0.461	0.479	0.489	0.490	0.480
		PSD2p	-	-	0.328	0.326	0.327	0.326	0.326
		PSD3p	-	-	0.328	0.326	0.327	0.326	0.326
		PSD4p	-	-	0.328	0.326	0.327	0.326	0.326
<i>todacen</i> (structure and parameter learning)	SPGaDp (Genie)	SPGaD1p	CI-Test (PC)	35	0.638	0.622	0.633	0.629	0.630
		SPGaD2p	CI-Test (PC)	35	0.643	0.623	0.633	0.629	0.632
		SPGaD3p	CI-Test (PC)	35	0.638	0.623	0.633	0.629	0.630
		SPGaD4p	CI-Test (PC)	35	0.637	0.622	0.632	0.628	0.630
	SPGbDp (Genie)	SPGbD1p	SaS (GTT/K2)	35	0.637	0.622	0.632	0.628	0.630
		SPGbD2p	SaS (GTT/K2)	35	0.637	0.623	0.633	0.629	0.630
		SPGbD3p	SaS (GTT/K2)	35	0.638	0.623	0.633	0.629	0.630
		SPGbD4p	SaS (GTT/K2)	35	0.637	0.622	0.632	0.628	0.630
	SPGcDp (Genie)	SPGcD1p	SaS (GTT/BDu)	35	0.637	0.622	0.632	0.628	0.630
		SPGcD2p	SaS (GTT/BDu)	35	0.637	0.621	0.632	0.628	0.629
		SPGcD3p	SaS (GTT/BDu)	35	0.636	0.622	0.632	0.628	0.630
		SPGcD4p	SaS (GTT/BDu)	35	0.638	0.623	0.633	0.629	0.631
	SPWDp (WinMine Toolkit)	SPWD1p	SaS (?/BDu)	35	*	*	*	*	*
		SPWD1p	SaS (?/BDu)	35	*	*	*	*	*
		SPWD1p	SaS (?/BDu)	35	*	*	*	*	*
		SPWD1p	SaS (?/BDu)	35	*	*	*	*	*

* Could not evaluate the models based on the performance metrics, using the WinMine toolkit.

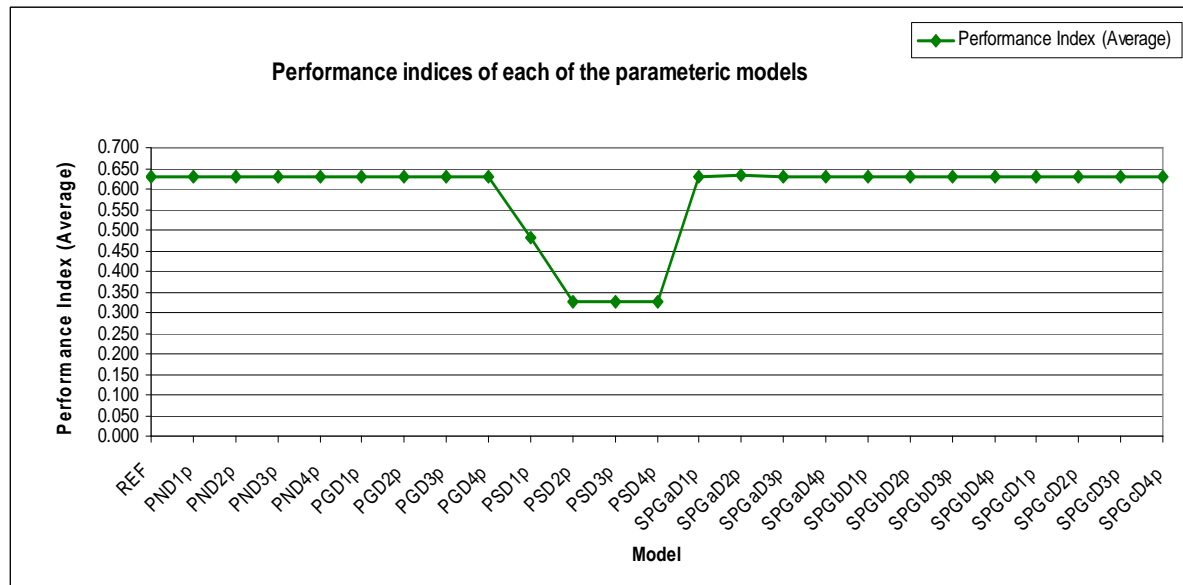


Figure 5.2: Average performance indices of the *parameteric* models and the Reference model

Tables 5.3 and 5.4, and Figures 5.2 and 5.3, highlight performance differences between the *parameteric* and *non-parameteric* models, relative to the performance of the Reference model. The performance indices of the SPWDp group of models were not listed in Table 5.3 because they could not be evaluated with respect to the performance metrics, using WinMine toolkit. The WinMine BN network file format (.xmod) did not allow for its conversion to network file formats supported by other software tools that facilitate model evaluation with respect to the performance metrics. Also, the performance indices of the *non-parameteric todacen* groups of models (with the exception of one) were not listed in Table 5.4 because no structures or meaningful structures were learnt, hence no model to evaluate.

Table 5.4: Non-parameteric models: Performance indices

NON-PARAMETERIC MODELS									
Type	Model Group and (Software Tool)	Model	Structure Learning Approach	No of links learnt	Performance Index				
					T1	T2	T3	T4	Average Performance Index
<i>tecen</i>	-	REF	-	-	0.638	0.623	0.633	0.629	0.630
<i>Sedacen</i> (known structure, parameter learning)	PNDnp (Netica)	PND1np	-	-	0.480	0.469	0.469	0.471	0.472
		PND2np	-	-	0.436	0.438	0.433	0.431	0.435
		PND3np	-	-	0.484	0.487	0.488	0.490	0.487
		PND4np	-	-	0.436	0.456	0.453	0.454	0.450
	PGDnp (Genie)	PGD1np	-	-	0.480	0.469	0.469	0.469	0.472
		PGD2np	-	-	0.436	0.455	0.453	0.454	0.449
		PGD3np	-	-	0.436	0.438	0.433	0.431	0.435
		PGD4np	-	-	0.484	0.487	0.488	0.490	0.487
	PSDnp (SamIam)	PSD1np	-	-	0.328	0.326	0.327	0.326	0.326
		PSD2np	-	-	0.328	0.326	0.327	0.326	0.326
		PSD3np	-	-	0.328	0.326	0.327	0.326	0.326
		PSD4np	-	-	0.328	0.326	0.327	0.326	0.326
<i>todacen</i> (structure and parameter learning)	SPGadnp (Genie)	SPGad1np	CI-Test (PC)	6 (4 nets)	*	*	*	*	*
		SPGad2np	CI-Test (PC)	9 (7 nets)	*	*	*	*	*
		SPGad3np	CI-Test (PC)	10 (8 nets)	*	*	*	*	*
		SPGad4np	CI-Test (PC)	25 (3 nets)	0.495	0.488	0.493	0.493	0.493
	SPGbdnp (Genie)	SPGbd1np	SaS (GTT/K2)	0	*	*	*	*	*
		SPGbd2np	SaS (GTT/K2)	0	*	*	*	*	*
		SPGbd3np	SaS (GTT/K2)	0	*	*	*	*	*
		SPGbd4np	SaS (GTT/K2)	0	*	*	*	*	*
	SPGcdnp (Genie)	SPGcd1np	SaS (GTT/BDu)	0	*	*	*	*	*
		SPGcd2np	SaS (GTT/BDu)	0	*	*	*	*	*
		SPGcd3np	SaS (GTT/BDu)	0	*	*	*	*	*
		SPGcd4np	SaS (GTT/BDu)	0	*	*	*	*	*
	SPWDnp (WinMine Toolkit)	SPWD1np	SaS (?/BDu)	0	*	*	*	*	*
		SPWD2np	SaS (?/BDu)	0	*	*	*	*	*
		SPWD3np	SaS (?/BDu)	0	*	*	*	*	*
		SPWD4np	SaS (?/BDu)	0	*	*	*	*	*

* No structure or meaningful structure was learnt, hence no model to evaluate

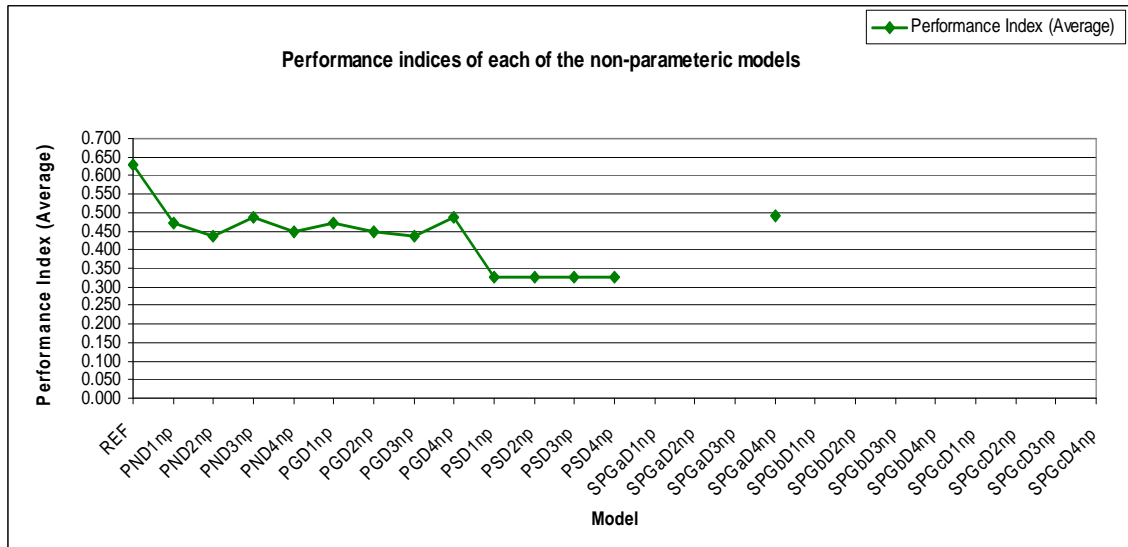


Figure 5.3: Average performance indices of the *non-parameteric* models and the Reference model

It was observed, as highlighted and indicated by Tables 5.3 and 5.4, and Figures 5.2 and 5.3, that the:

- best performing model is the SPGaD2p with an average performance index of 0.632, as against the Reference model's average performance index of 0.630.
- performances of the *parameteric* groups of model (with the exception of one) are comparable (equivalent) to the performance of the Reference model. That is, the performances of 67% of the *parameteric sedacen* and 100% of the *parameteric todacen* models were comparable (equivalent) to that of the reference model.
- The performances of the *non-parameteric* models were relatively poor compared to the performance of the Reference model. That is, taking 0.500 as the threshold between comparable and poor performance, the performances of 100% of both the *non-parameteric sedacen* and *todacen* models were relatively poor. The learnt CPT entries were more or less inconclusive.
- In 15 (94%) of the 16 cases in which the *Dnp* set of datasets (*non-parameteric* datasets) were used for the construction of *todacen* models (structure and parameter learning), no structures or meaningful structures were learnt.

In addition, a weak negative correlation ($r = -0.0734$) was found between the size of the training dataset and model performance. Also, a weak negative correlation ($r = -0.0569$, $p = 0.94400$) was found between the size of the test dataset and model performance.

5.6 COMMENTS

The results of the empirical investigation are encouraging. The results provide additional insight for BN model constructors and contribute to the literature providing supportive evidence for the conceptual feasibility and efficiency of structure and parameter learning algorithms and approaches. The induction of models which performances were comparable (equivalent) to, and in one case better than (albeit marginally by 0.002 (0.32%)) the performance of the Reference model is significant. However, the results show that it may not have been possible to construct the *sedacen* and *todacen* models with performances that are comparable to or better than the performance of the Reference model without first constructing the complete Reference model (structure + parameters), with assistance of domain experts.

Though the performances of all the models (*parameteric* and *non-parameteric*, *sedacen* and *todacen*) constructed using the SamIam software tool were relatively poor, it is assumed to imply inefficiency of the software tool, which however, may be due to some uncontrolled factor(s) and therefore may require further investigation. The results also indicate that the Conditional Independence (CI) test based PC structure algorithm is equivalent in its learning outcomes to the Score and Search (SaS) based GTT/K2 and GTT/BDeu structure learning algorithms, with respect to the *parameteric* models. The PC algorithm performed better than the SaS-based algorithms with respect to the *non-parameteric* models. It was able to learn some links (in 100% of the cases), though the links did not yield meaningful models in 75% of the cases. Furthermore, the sizes of the training and test datasets did not seem to have any relationship with model performance. However, the results showed that in 100% of the 12 *non-parameteric sedacen* models, almost equal probability values were assigned to all possible parent combinations. The parameter values seemed logically and realistically unacceptable for the purpose for which the models are aimed, as highlighted in Figure 5.4, the CPT of one of the nodes in a *non-parameteric sedacen* model.

Node: AbilityToConstructCircuit

Chance: % Probability

Apply Okay

Reset Close

ProperUseOfEquipment (PUE)	AbilityToWorkWithComponents ...	AbilityToAdapt (ADA)	AbilityToAdhereToConstraints (...)	high	low
properlyUsed	high	high	adhered	48.801	51.199
properlyUsed	high	high	partiallyAdhered	49.926	50.074
properlyUsed	high	high	notAdhered	49.338	50.662
properlyUsed	high	low	adhered	50.597	49.403
properlyUsed	high	low	partiallyAdhered	50.711	49.289
properlyUsed	high	low	notAdhered	51.069	48.931
properlyUsed	low	high	adhered	51.084	48.916
properlyUsed	low	high	partiallyAdhered	48.511	51.489
properlyUsed	low	high	notAdhered	50.671	49.329
properlyUsed	low	low	adhered	49.22	50.78
properlyUsed	low	low	partiallyAdhered	51.244	48.756
properlyUsed	low	low	notAdhered	49.949	50.051
partiallyProperly	high	high	adhered	50.131	49.869
partiallyProperly	high	high	partiallyAdhered	48.201	51.799
partiallyProperly	high	high	notAdhered	50.902	49.098
partiallyProperly	high	low	adhered	49.631	50.369
partiallyProperly	high	low	partiallyAdhered	50.964	49.036
partiallyProperly	high	low	notAdhered	50.077	49.923
partiallyProperly	low	high	adhered	49.949	50.051
partiallyProperly	low	high	partiallyAdhered	50.357	49.643
partiallyProperly	low	high	notAdhered	49.876	50.124
partiallyProperly	low	low	adhered	50.83	49.17
partiallyProperly	low	low	partiallyAdhered	50.916	49.084
partiallyProperly	low	low	notAdhered	49.572	50.428
wronglyUsed	high	high	adhered	50.519	49.481
wronglyUsed	high	high	partiallyAdhered	50.348	49.652
wronglyUsed	high	high	notAdhered	50.495	49.505
wronglyUsed	high	low	adhered	49.355	50.645
wronglyUsed	high	low	partiallyAdhered	51.127	48.873
wronglyUsed	high	low	notAdhered	49.874	50.126
wronglyUsed	low	high	adhered	51.421	48.579
wronglyUsed	low	high	partiallyAdhered	48.659	51.341
wronglyUsed	low	high	notAdhered	50.121	49.879
wronglyUsed	low	low	adhered	47.149	52.851
wronglyUsed	low	low	partiallyAdhered	49.581	50.419
wronglyUsed	low	low	notAdhered	51.517	48.483

Figure 5.4: The CPT of a node in a *non-parameteric sedacen* model, PND2np.

Finally, the results have also highlighted an important area that may require further investigation. The results showed that a complete Reference model (that is knowledge of the relationship between the domain variables and their Conditional Probability Distributions) is a requirement for simulating sample datasets for structure and/or parameter that will yield meaningful and comparable models for the domain. This suggests the need for further research in order to investigate the outcome of structure and/or parameter learning using historical sample datasets from the domain that may not have been generated with knowledge of the relationship between the domain variables and their Conditional Probability Distributions. The main challenge to this investigation will be the obtainment of historical sample datasets for the domain.

5.7 SUMMARY

The best approach for the construction of the BN-based model for the performance assessment of students' laboratory work in the VEL environment has been empirically investigated. The optimisation exercise yielded a better model (albeit marginally). This provides reassurance that the procedure followed in the derivation of the assessment model (Chapter 4) was fit for purpose. The results contribute to the literature providing supportive evidence for the conceptual feasibility and efficiency of structure and parameter learning algorithms and approaches. In addition, they also highlighted the need for further investigation with respect to data-centred BN model construction approach for the domain.

Furthermore, from our experience, the data-centred BN model construction approach depends on the availability of appropriate software tools and sample training datasets. Model construction may be limited by the software tools available. Commercial software tools may be inaccessible, in which case freeware tools, which may have limited capabilities, are used. Also, there seems to be no standardized data and network file formats for BN software tools. Different tools support different data and network file formats. For example, some software tools may support only numeric data files, some string data files, while some may require the inclusion of record occurrence frequencies. This may be counterproductive for the data-centred BN construction approach.



CHAPTER 6

ASSESSMENT MODEL: DATA EXTRACTION, EVIDENCE VARIABLE QUANTIZATION AND INSTANTIATION

From time immemorial, man has desired to comprehend the complexity of nature in terms of as few elementary concepts as possible.

Abdus Salam
1936 - 1996

What we see depends on what we look for.

John Lubbock
1834 - 1913

6.0 INTRODUCTION

The performance data required to generate evidence for updating the LAP model is generated from the analysis of students' laboratory work behaviour log. The extraction of the necessary performance data for the quantization of evidence variables and their instantiation processes are detailed in this chapter. In addition, some valuable information derivable from the inspection of students' laboratory work schemas, generated from their behaviour logs, are highlighted.

First, section 6.1 describes the behaviour log analysis and performance data extraction processes. Section 6.2 details the evidential data extraction and evidence variable quantization for some of the evidence variables, while section 6.3 addresses evidence instantiation using a fuzzification engine. Section 6.4 highlights some valuable information derivable from the inspection of student's laboratory work schemas. The chapter is summarized in section 6.5.

6.1 PERFORMANCE DATA EXTRACTION

The application of a BN model requires that evidence variables are instantiated for network update/inference. In this context, the evidence variables are first quantified, using the performance data extracted from the behaviour log, and then instantiated, using a fuzzification engine. Figure 6.1 diagrammatically highlights these processes. A behaviour log consists of the following performance (evidential) data sources:

1. laboratory activity results and answers to laboratory questions
2. built circuit schematics and simulation commands (a set of netlists)
3. responses to pre-/post-laboratory test items.
4. interaction events log (mouse click stream and key presses).

Data sources 1, 2, and 3 are referred to as explicit sources and 4 is referred to as implicit source, which is subjected to transformation and analysis. In addition to extracting performance data, information from the analysis of a student's interaction events log can be visualised in order to gain insight into the student's laboratory work schema, can uncover aspects of a student work process (such as areas where a student had difficulties, and conformance to safety issues) that are difficult to capture otherwise. It is also useful for student characterization and grouping, with respect to laboratory work behaviour.

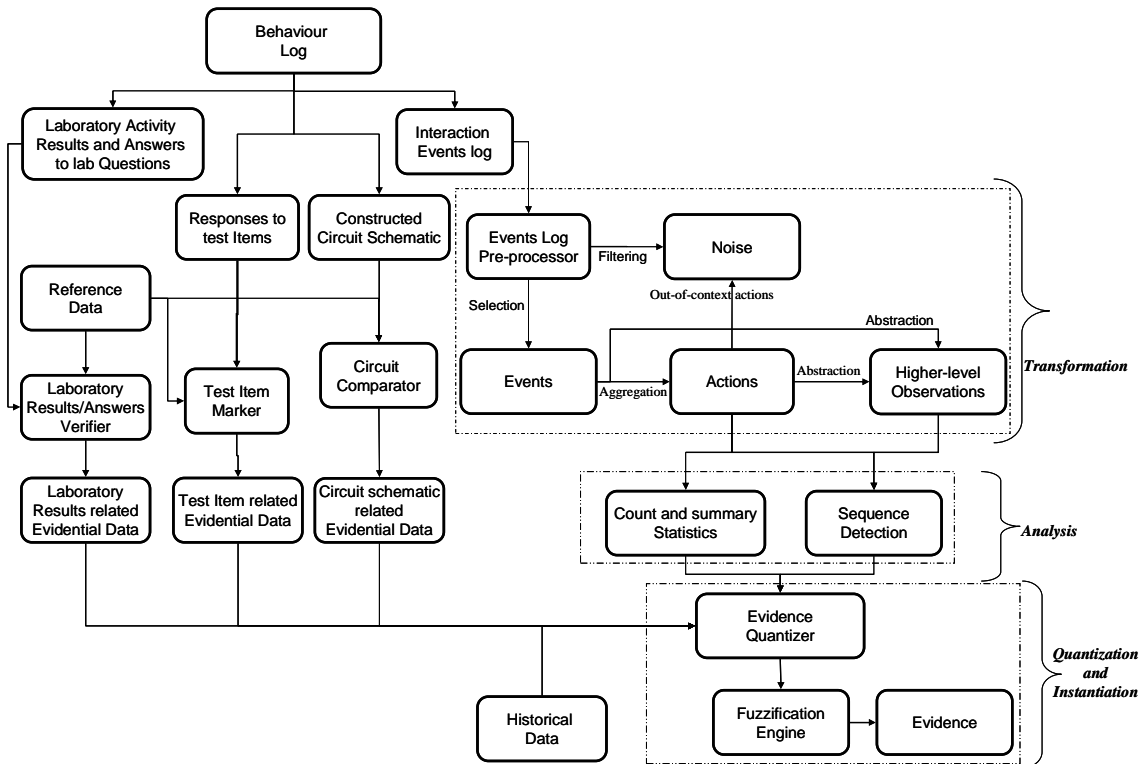


Figure 6.1: Data extraction and evidence quantization and instantiation processes.

The events log is first filtered to remove noise and then subjected to selection and aggregation. Transformation involves filtering, selection, aggregation, and abstraction. The concepts of selection, aggregation, and abstraction are depicted diagrammatically in Figures 6.2 and 6.3.

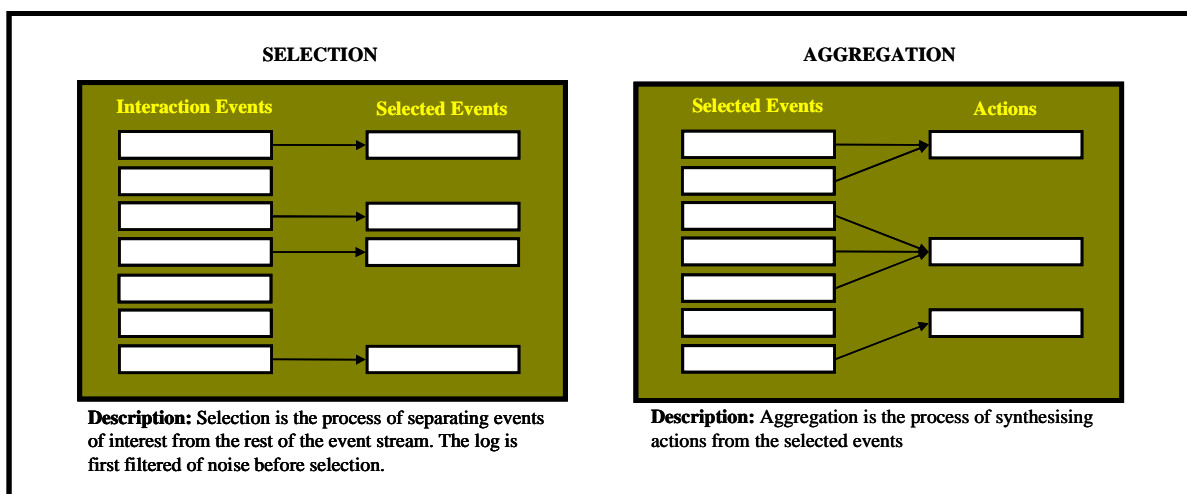


Figure 6.2: Interaction events transformation: Selection and Aggregation
(Adapted from: [108]).

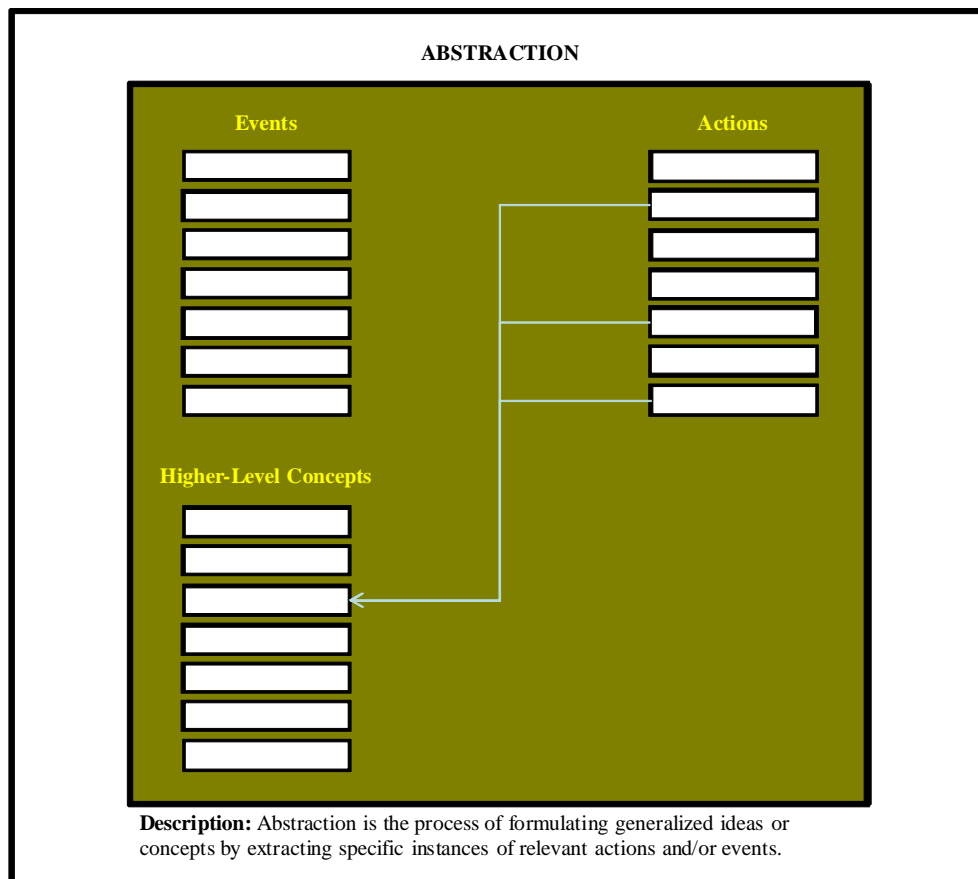


Figure 6.3: Interaction events transformation: Abstraction
(Adapted from: [108]).

Selection involves separating events of interest from other events. For example, it was chosen to disregard all events associated with mouse movements in order to focus analysis on more relevant actions such as circuit component selection, pickup, and placement on the breadboard. Aggregation is the process of synthesizing actions from event stream. Events occur serially and actions are composed of a number of events that occurred in sequence. An event can, in itself, constitute an action. Rules are used to characterize and aggregate events occurring in sequence. For example, the power supply is seen to have been connected to the breadboard, if the following events have occurred in sequence, as highlighted in Figure 6.4.

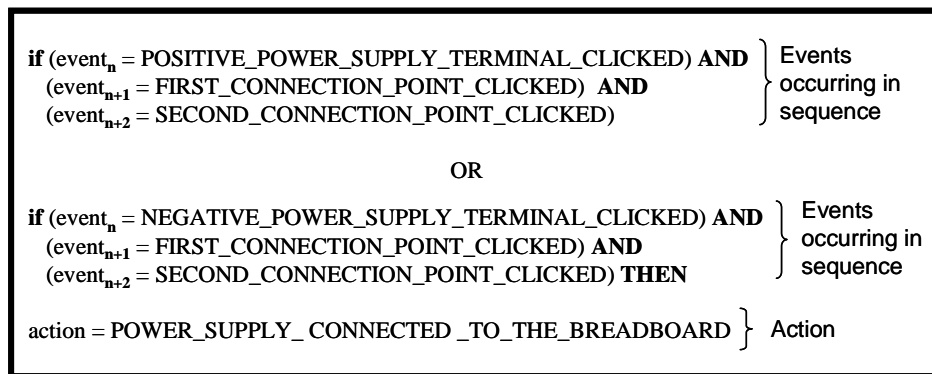


Figure 6.4: Aggregation of atomic events to highlight an action.

First and second connection points refer to contact positions on the breadboard where the student connected the red (live) and black (ground) wires from the terminals of the power supply. Table 6.1 lists some of the actions that can be taken in the VEL environment. Abstraction involves relating actions, and possibly relevant events, to higher-level concepts of interest. For example, it could be abstracted from a student's actions if the student observed an instance of safety procedure, such as first connecting the terminals of the power supply to the breadboarded circuit before switching the power supply on.

Table 6.1: A list of some of the actions that can be taken in the VEL

- SELECT_A_COMPONENT_IN_A_COMPONENT_CONTAINER
- DESELECT_A_COMPONENT_IN_A_COMPONENT_CONTAINER
- PICK_UP_A_COMPONENT_FROM_A_COMPONENT_CONTAINER,
- DROP_A_PICKED_UP_OMPONENT
- PLACE_ONE_BREADBOARD_ON_THE_WORKBENCH
- PLACE_TWO_BREADBOARDS_ON_THE_WORKBENCH
- REMOVE_A_PLACED_COMPONENT_FROM_THE_BREADBOARD
- SET_VOLTAGE_VALUE_ON_THE_POWER_SUPPLY
- REMOVE_BREADBOARD_FROM_THE_WORKBENCH
- CASCADE_TWO_BREADBOARDS
- DECASDE_TWO_BREADBOARDS
- CONNECT_POWER_SUPPLY_TO_THE_BREADBOARD
- DISCONNECT_POWER_SUPPLY_FROM_THE_BREADBOARD
- CONNECT_FUNCTION_GENERATOR_TO_THE_BREADBOARD
- DISCONNECT_FUNCTION_GENRATOR_FROM_THE_BREADBOARD
- SIMULATE_CONSTRUCTED_CIRCUIT
- PLACE_A_COMPONENT_ON_THE_BREADBOARD
- CLEAR_THE_WORKBENCH
- CLEAR_THE_BREADBOARD
- SEEK_INFORMATION
- SWITCH_ON_THE_POWER_SUPPLY
- SWITCH_OFF_THE_POWER_SUPPLY
- SWITCH_ON_THE_FUNCTION_GENERATOR
- SWITCH_OFF_THE_FUNCTION_GENERATOR
- VIEW_PLOTTED_GRAPHS
- SELECT_THE_SIGNAL_TYPE_FOR_AC_ANALYSIS
- SET_SIGNAL_RMS_VALUE_ON_THE_FUNCTION_GENERATOR
- SET_FREQUENCY_RANGE_FOR_FUNCTION_GENERATOR
- SEARCH_FOR_RESISTOR
- SEARCH_FOR_CAPACITOR
- SEARCH_FOR_INDUCTOR
- SEARCH_FOR_WIRE
- REPOSITION_COMPONENT_PIN_ON_BREADBOARD
- SELECT_A_COMPONENT_ON_BREADBOARD
- DESELECT_A_COMPONENT_ON_BREADBOARD

6.2 EVIDENCE QUANTIZATION CASE STUDIES

Each evidence variable is quantized using the relevant data extracted from the performance data sources, a set of rules, sequence detection (see Figure 6.5), and count and summary statistics (see Figure 6.5). The purpose of sequence detection is to recognize sequences of actions that, for example, violate or meet particular expectations, such as safety procedure.

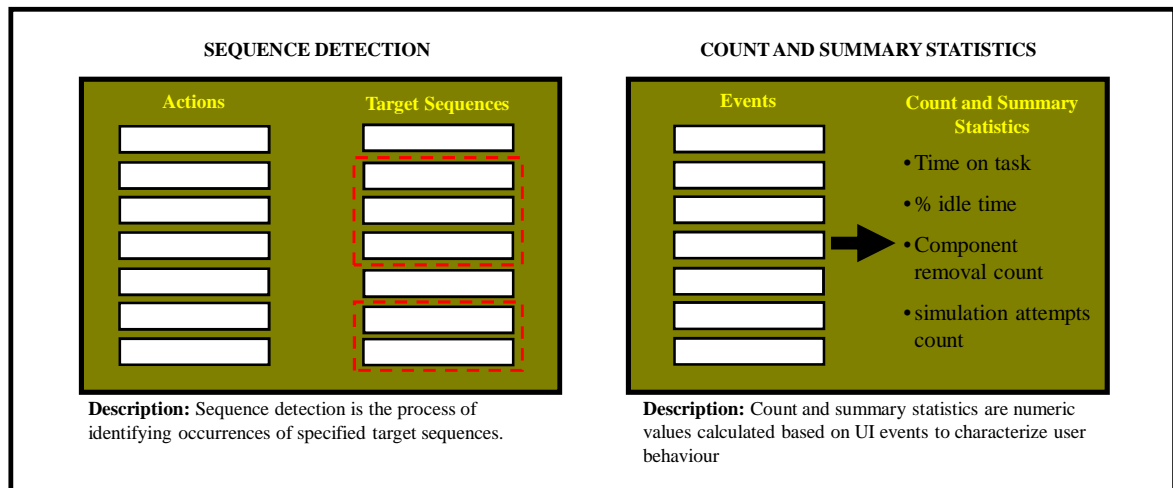


Figure 6.5: Sequence detection and count and summary statistics
(Adapted from: [108]).

The quantization process uses a reference data set, as the “gold standard”, where necessary. The processes of data extraction and quantization of evidence variables, node 1, “*UseOfCorrectComponents*” (*UCC*), node 2, “*CorrectPlacementOfComponents*” (*CPC*), and node 33, “*ExperienceUsingVEL*” (*EUV*), (see Chapter 5, Figure 5.3) are hereby described. The quantization of the remaining evidence variables is described in Appendix C. The quantization of nodes 1 (*UCC*) and 2 (*CPC*) involves the comparison of a student’s built circuit with the reference circuit, for equivalence, by the *CircuitComparator*, which consists of two component parts: *ComponentComparator* and *TopologyComparator*. Two electrical circuit networks are equivalent if they both contain the same set of components in type and value, and the manner in which the components are connected result in the same branch currents and voltages [222]. The circuits are represented in the system as netlists.

6.2.1 Quantization of Node 1: “*UseOfCorrectComponents*”

The *Quantization of UCC* entails checking if a student’s built circuit is component- and value-wise equivalent to the reference circuit, by the *ComponentComparator*. The variable, *UCC*, is then quantized as:

$$UCC = \frac{100}{n} \times q \left(\frac{\sum_{i=1}^n x_i}{m} \right) \quad (6.1)$$

Where, in this case, n = number of different types of reference circuits, m = number of circuits built and simulated by a student in the process of undertaking an activity, x_i = number of circuits built and simulated by student that are component- and value-wise equivalent to reference circuit type i , and q = number of x_i 's not equal to zero (q is an indication that a circuit of a particular type is built with the correct components and values, at least once).

6.2.2 Quantization of Node 2: “CorrectPlacementOfComponents”

The *quantization of CPC* necessitates that a student's built circuit be compared to the reference circuit for topological equivalence, by the *TopologyComparator*. Two circuits are considered to be topologically equivalent if they have the same voltage/current characteristics across all terminals of the networks [222]. That is, the topological comparison of two circuits requires the explicit comparison of the branch voltage/current characteristics of the circuits. For this purpose, a directed graph, G , without self-loops, is used to completely describe the circuit (network) and the KVL and KCL equations of the network are then derived from the graph, using the relevant component of the graph [223]. A graph, G , is a pair (U,E) where U is the set of all the vertices on the graph and E is the set of pairs (u,v) called edges such that $u,v \in U$, $u \neq v$ [224]. The nodes and branches of the network make up the vertices and the edges of the graph, respectively. The orientations of the edges reflect the reference direction of the currents/voltages across the branches of the network. The relevant components of the graph, G , required for the derivation of the matrices/equations necessary for the characterization of the network represented by G include path, subgraph, loop, cutset, tree, and cotree (see [223][225] for details). The matrices include the Incident Matrix (IM), A_a , and Reduced Incident Matrix (RIM), A ; Loop Matrix (LM), B_a , and Fundamental Loop Matrix (FLM), B ; Cutset Matrix (CM), D_a , and Fundamental Cutset Matrix (FCM), D . The IM, A_a , is generated directly from G . For example, the network of Figure 6.6, is described by the graph, G , of Figure 6.7.

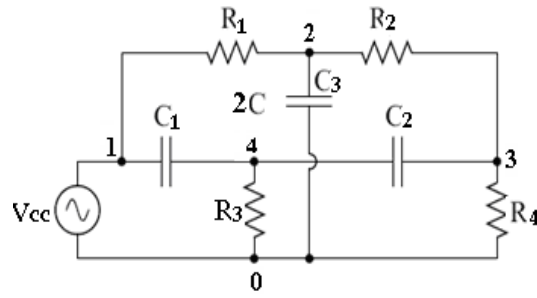


Figure 6.6: Circuit schematic of one of the laboratory activities used for the evaluation of the VEL: *Twin-T Notch filter*.

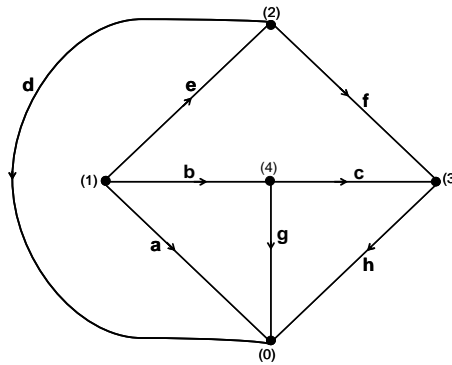


Figure 6.7: The graph, G , describing the network of Figure 6.6.

In G , the edges represent the branches of the network, so that a, b, c, d, e, f, g , and h represent $v_{cc}, c1, c2, c3, r1, r2, r3$, and $r4$, respectively. G has n vertices and b edges corresponding to the n nodes and b branches of the network. The graph, G , is used to generate the KCL and KVL equations describing the network, such that, in this context, the check for topological equivalence between a student's built and the reference circuit reduces to matrix comparison. The details of the generation of the KCL and KVL equations, from G , are given in Appendix C1. The result of the comparison is used to generate the quantized value for the node, CPC , using equation 6.1.

$$CPC = \frac{100}{n} \times q \left(\frac{\sum_{i=1}^n x_i}{m} \right) \quad (6.1)$$

Where, in this case, n = number of different types of reference circuits, m = number of circuits built and simulated by a student, x_i = number of circuits built and simulated by a student that are topologically equivalent to reference circuit type i , and q = number of x_i 's not equal to zero (q is an indication that a circuit of a particular type is built topologically correctly at least once).

6.2.3 Quantization of Node 33: “ExperienceUsingVEL”

In order to quantize the variable, “ExperienceUsingVEL (EUV)”, the total time on task to date is used. The measure of experience is a real number, β , such that $0 \leq \beta \leq 100$. $\beta = 0$ if a student has not previously undertaken any laboratory activity in the VEL environment. Each time a student undertakes a laboratory activity in the VEL environment, the activity counter is increased, and a value for the total time on activity to date, x , is derived, as $x = t_d + t_c$, where t_d is the total time on task to date, before undertaking $activity_{n-1}$, and t_c is the total time on task for $activity_{n-1}$. That is, since $activity_n$ is the activity being assessed, the experience that will impact on the performance of $activity_n$ is the experience up to $activity_{n-1}$. Using x , the measure of experience, β , is evaluated as a percentage of the set maximum total time on activity, $t_m = 18 \times 10^4$. t_m is set based on the assumption/expectation that a student would have mastered the use of the VEL by the time of undertaking up to 50 laboratory activities in the environment, where the average time per activity is 3600secs (1 hr). Hence, using x , the measure of experience, β , is evaluated as follows:

$$\beta = 0, \text{ if } x = 0, \text{ else, } \beta = \left(\frac{x}{t_m} \times 100 \right) \quad (6.2)$$

In this way, a student’s measure of experience, β , increases with each activity. The number of activities and average time per activity could in practical settings be varied to suit specific educational needs.

6.3 EVIDENCE INSTANTIATION

The quantization process generates a numerical value (a score), for an evidence variable, which serves as input into the fuzzification engine. Fuzzification is the process of making a crisp quantity fuzzy [Ross, 1995]. The fuzzification process (shown in Figure 6.8) transforms the score into one of the linguistic values (states) of the evidence variable. The technicalities of the fuzzification process are documented in literature on fuzzy sets, logic, and systems such as [226].

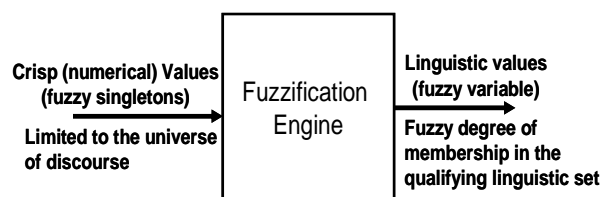


Figure 6.8: Fuzzification Engine.

In this context, the fuzzification process is based on the universe of discourse, $X = \{0, 100\}$, the set of possible numerical scores for the evidence variables. The score, $x_i \in X$, for an evidence variable, A_i , is fuzzified by generating its membership value in $T(A_i) = (t_1, t_2, t_3)$, the linguistic term set (states) of A_i . $\mu(x_i)$ maps each element, $x_i \in X$, to a degree of membership between 0 and 1, in t_i , where t_i is the set of ordered pair: $\{x_i, \mu(x_i) \mid x_i \in X\}$. The trapezoidal function (equation 7.15) is the core of the fuzzification engine used in this work.

$$\mu(x_i) = \begin{cases} 0 & \text{if } x_i < a \\ \frac{x_i - a}{m - a} & \text{if } x_i \in [a, m] \\ 1 & \text{if } x_i \in [m, n] \\ \frac{b - x_i}{b - n} & \text{if } x_i \in [n, b] \\ 0 & \text{if } x_i > b \end{cases} \quad (6.3)$$

The equivalent notation for equation 7.15 is:

$$a_i(x_i, a, m, n, b) = \max\{\min[(x_i - a) / (m - a), 1, (b - x_i) / (b - n)], 0\} \quad (6.4)$$

The trapezoidal membership function is shown graphically in Figure 6.9, for node 1 of Figure 4.3, “*UseCorrectComponents (UCC)*”, with states *used*, *partiallyUsed*, *notUsed*, for example. The partitions, P_1 , P_2 , and P_3 are created using the mark-grade mapping scheme of one of the pilot Universities. The crossover points for the partitions are at $\mu(x_i) = 0.5$.

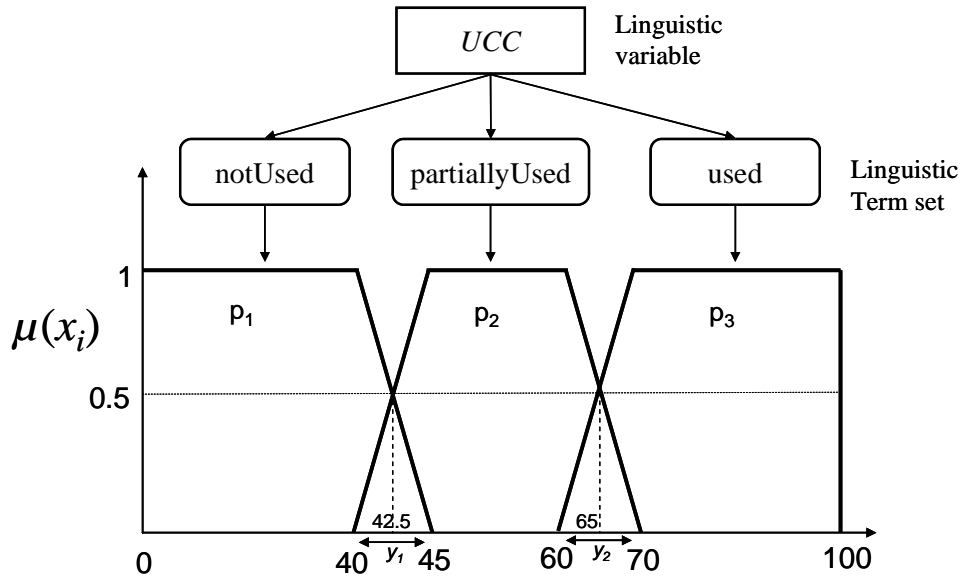


Figure 6.9: The trapezoidal function representation for node 1 of Figure 4.3, “UseCorrectComponents”

If, for example, $UCC = 44$, then $\mu_{used}(44) = 0.2$ and $\mu_{partiallyUsed}(44) = 0.8$, based on which UCC will be instantiated to the state, *partiallyUsed*. A node is instantiated to the state for which the membership value of its score is greater. The fuzzification rules, based on the membership values of x_i , are summarily highlighted in Table 6.2.

Table 6.2: Summary highlight of the fuzzification rules

S/No.	Score (x_i)	Membership Value	Partition segment	Linguistic Term (State)
1	$0 \leq x_i \leq 40$	$\mu(x_i) = 1$	P_1 core	<i>notUsed</i>
2	$40.01 \leq x_i \leq 42.99$	$0.5 < \mu(x_i) < 1$ in P_1 and $0.0 < \mu(x_i) < 0.5$ in P_2	p_1 - p_2 boundary, left	<i>notUsed</i>
3	$x_i = 42.5$	$\mu(x_i) = 0.5$	p_1 - p_2 boundary, centre	<i>notUsed</i>
4	$42.51 \leq x_i \leq 44.99$	$0.0 < \mu(x_i) < 0.5$ in P_1 and $0.5 < \mu(x_i) < 1$ in P_2	p_1 - p_2 boundary, right	<i>partiallyUsed</i>
5	$45 \leq x_i \leq 60$	$\mu(x_i) = 1$	p_2 core	<i>partiallyUsed</i>
6	$60.01 \leq x_i \leq 64.99$	$0.5 < \mu(x_i) < 1$ in P_2 and $0.0 < \mu(x_i) < 0.5$ in P_3	p_2 - p_3 boundary, left	<i>partiallyUsed</i>
7	$x_i = 65$	$\mu(x_i) = 0.5$	p_2 - p_3 boundary, centre	<i>used</i>
8	$65.01 \leq x_i \leq 69.99$	$0.0 < \mu(x_i) < 0.5$ in P_2 and $0.5 < \mu(x_i) < 1$ in P_3	p_2 - p_3 boundary, right	<i>used</i>
9	$70 \leq x_i \leq 100$	$\mu(x_i) = 1$	P_3 core	<i>used</i>

6.4 STUDENT LABORATORY WORK SCHEMA

In addition to the extraction of performance data from students' interaction events logs, inspection of a student's laboratory work schema can divulge valuable information, such as patterns of actions, and highlight areas where a student is possibly experiencing some difficulties. Figures 6.10, 6.11, and 6.12 represent the graphical depiction of the laboratory work schemas of three different students, Stud_A, Stud_B, and Stud_C, for the same laboratory activity.

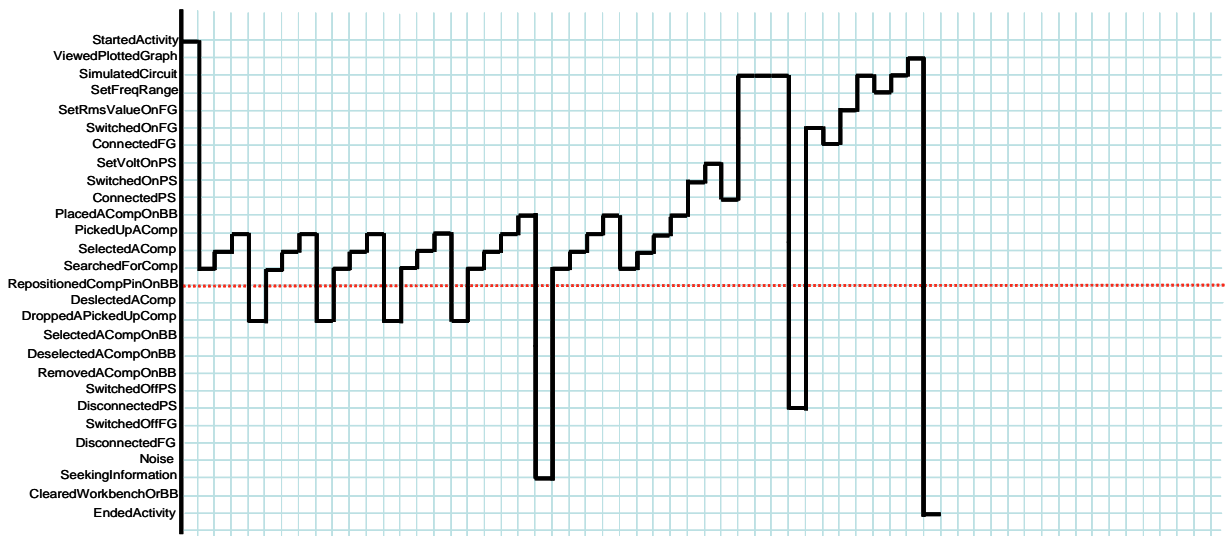


Figure 6.10: Stud_A laboratory work schema

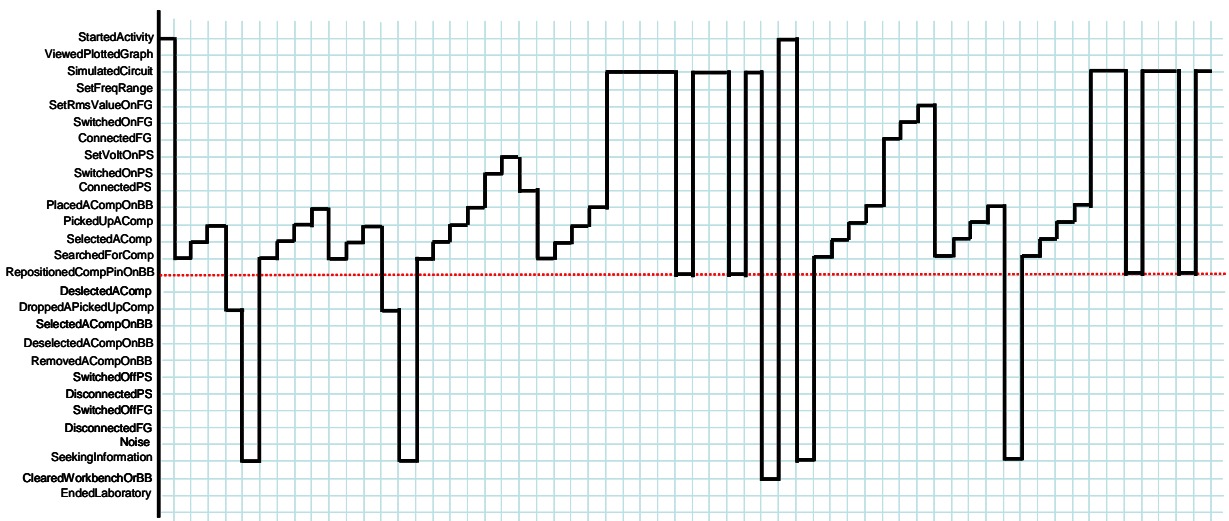


Figure 6.11: Stud_B laboratory work schema

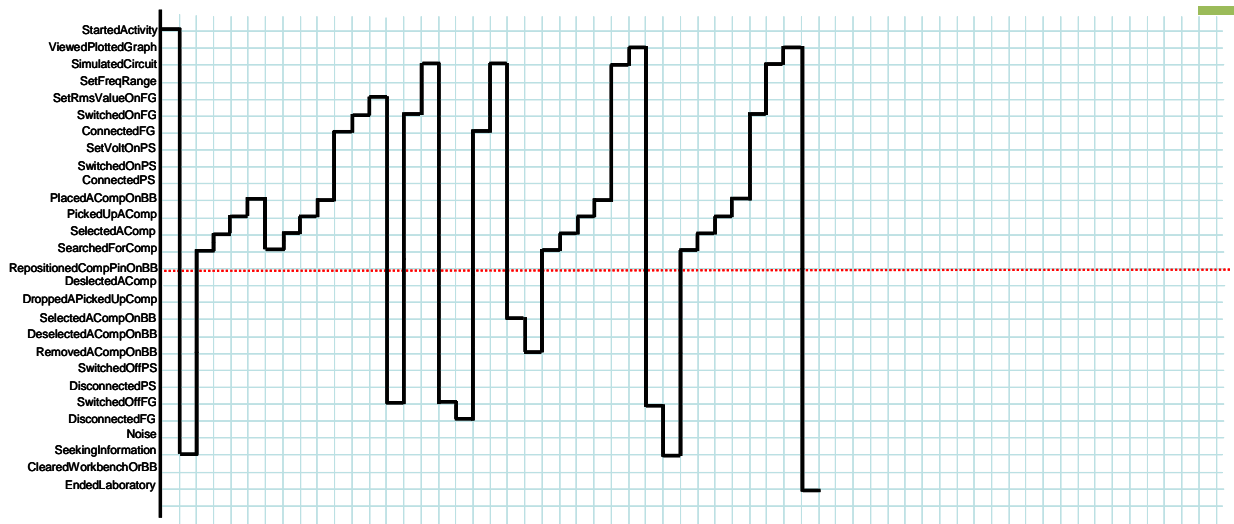


Figure 6.12: Stud_C laboratory work schema

Inspection of these schemas highlights interesting themes. *Stud_A*, on starting the activity, searched for, selected, picked up and dropped components severally before placing the first component on the BreadBoard (BB) and then went on to peruse the resistor colour code calculator (*UsedRColCodeCalc*), after which the student placed two more components on the BB, before simulating the breadboarded circuit. The student, obviously, used the Power Supply (PS) instead of the Function Generator (FG) and probably realized this mistake after three attempts to simulate the circuit. This realization made the student disconnect the PS and connect the FG. It can be seen that the student has not observed safety procedures at all. He first switched on the PS before connecting its terminals to the BB and disconnected the PS from the BB without switching it off first. Again, the student switched on the FG before connecting its terminals to the BB and while the FG was on and its terminals still connected to the BB, the student modified the breadboarded circuit by repositioning the pin of one of the components on the BB.

Stud_B seems to be struggling with the laboratory activity. First, the student picked up a component from a component container, dropped the component, and used the resistor colour code calculator before then picking and placing a component on the BB, a process which the student repeated. In the course of undertaking the activity, the student simulated the breadboarded circuit over twelve times, in-between repositioning component pins on the BB. At a point, the student cleared the entire breadboard, restarting the activity. Also, the student did not observe safety procedure. The student switched on the FG before connecting its terminals to the BB and severally modified the circuit while the FG was on and its terminals still connected to the BB.

Stud_C has a much shorter signature for the activity than students *A* and *B* and seems more precise and partially safety conscious. The student did not go through the process of pick and drop. The student first connected the terminals of the FG to the BB before switching it on. However, it can be seen that the student was randomly flicking the FG on and off, connecting, disconnecting, and reconnecting the FG to the BB which does not indicate proper equipment use. Also, the student modified the breadboarded circuit while the FG was on and its terminals still connected to the BB.

Essentially, visualizing students' laboratory work schema can highlight whether students are behaving consistently or if there is a consistently wrong or erroneous behavioural trend on the part of the students. For example, it can be seen from the work schemas represented in the above three figures that all the three students, *A*, *B*, and *C*, seem to have problems with safety related issues. Also, students can possibly be categorised based on their work schemas, and monitor the transition of a student from one category to another.

6.5 SUMMARY

The use of the LAP assessment model entails the derivation, from students' behaviour logs, data necessary for the generation of evidence to update the model. The processes for extracting performance data from a student's log have been highlighted and the process of quantizing the evidence variables detailed. The process of using a fuzzification engine to instantiate an evidence variable from its quantized value has also been described. At this juncture, the LAP model is ripe for verification and evaluation. Its verification and evaluation methodologies and results are presented in the next chapter.



CHAPTER 7

ASSESSMENT MODEL VERIFICATION, EVALUATION AND APPLICATION

*Not everything that can be counted counts, and not everything
that counts can be counted*

Albert Einstein

1879 - 1955

7.0 INTRODUCTION

Bayesian network models are often subjected to verification and evaluation. Verification is concerned with knowledge elicitation review, functional verification, and sensitivity analysis. Elicitation review involves addressing issues related to the elicitation process and the elicited data. The review process enables the domain expert to reconstruct his/her thought processes, thereby giving the expert a chance to rethink his/her judgements. In this context, elicitation review and model (structure and parameter) refinements were inherent parts of the knowledge elicitation processes and so will not be addressed further. Functional verification is basically a check of the functional correctness of the model, while sensitivity analysis is broadly concerned with investigating the relationship between network input parameters and the inference output of the network. Sensitivity analysis is useful in a number of areas, including model debugging. The evaluation of a model is a systematic assessment of its performance in terms of reliability and validity.

First, the verification of the model (functional verification and sensitivity analysis) is presented in section 7.1, while the evaluation (reliability and validity assessments) processes and results are detailed in section 7.2. Section 7.3 highlights areas of possible application of the model, while section 7.4 summarizes the chapter.

7.1 VERIFICATION

7.1.1 Functional Verification

The model was functionally verified by instantiating evidence nodes and checking whether the evidence was correctly propagated through the network. For example, when information about a student's placement of components is received and it is ascertained that the student has placed all the components correctly, node 2 is instantiated and clamped to the state "*placedAllCorrectly*". Figure 7.1 shows the relevant part of the LAP model before node 2 was instantiated, with the resulting updating of beliefs shown in Figure 7.2. As expected, the hypotheses, nodes 4, 15, and 16, that are influenced by node 2 have become more likely. The belief estimation for node 4, "*AbilityToWorkWithComponents*", rose from 50% to 65.4%. Also, there was a rise in the measure of belief, from 50% to 54.8%, in the ability to construct circuits and an increase in belief in node 16, "*AbilityToModifyCircuit*", from 49.9% to 53.1%. There was a 1.00% increase in the measure of belief for performance.

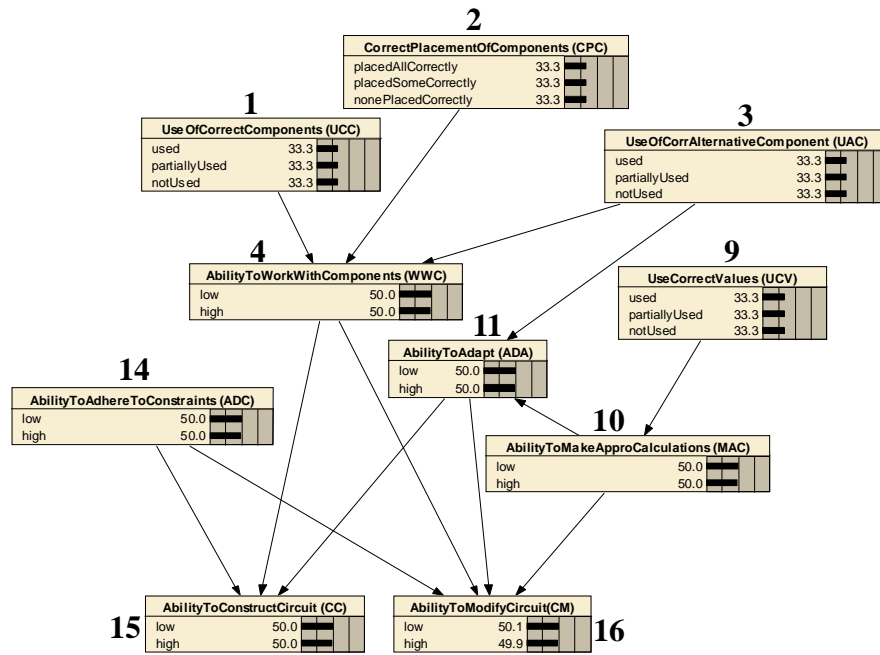


Figure 7.1: Extract from LAP model before node 2 was instantiated.

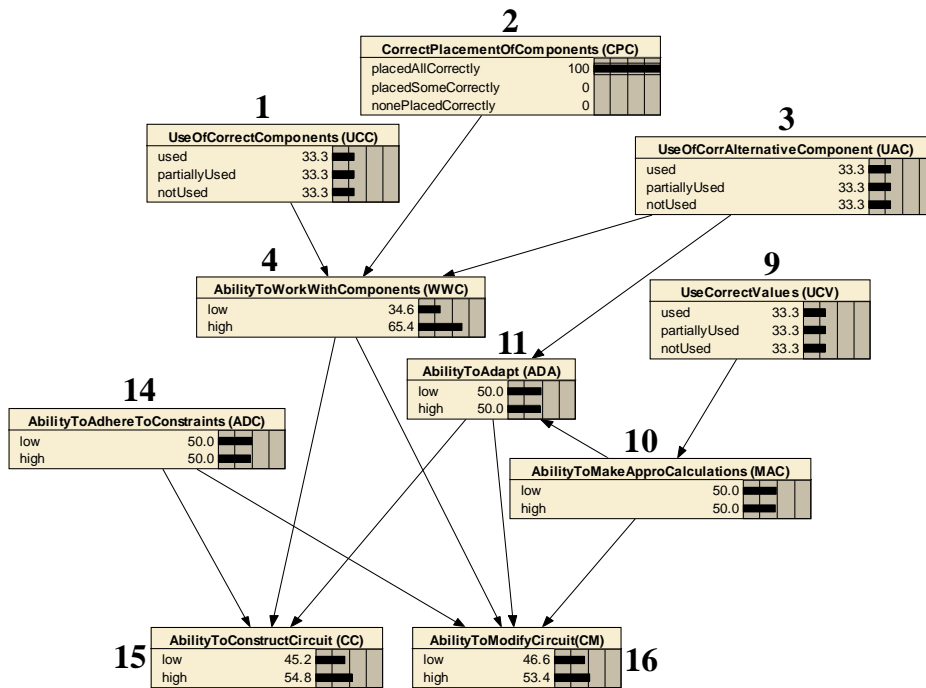


Figure 7.2: Extract from LAP model after node 2 was instantiated.

7.1.2 Sensitivity Analysis

Sensitivity analysis (SA) is a technique for systematically investigating the effects of variations in inputs on a model’s output. That is, analysis of the extent to which imprecision in the input probabilities affect the estimated probabilities. There are two types of SA:

sensitivities oriented to evidence (Evidence-based Sensitivity Analysis (ESA)) and sensitivities oriented to parameters (Parameter-based Sensitivity Analysis (PSA)) [176]. ESA determines the variables that have the highest or lowest impact on the belief estimation of a target variable, X , while PSA measures the impact of changes in the parameters of a node, A , on the probability distribution of a target node, X . Researchers often employ only one of these two types of analysis in any one study, mostly ESA, for instance [227][228][229][230]. ESA is often used to detect and minimize the effects of poorly calibrated network or bias in knowledge elicitation by identifying the excessively influential variables.

ESA estimates the extent to which the posterior probability of a target variable, X , is changed when finding is entered at a node, A , and the network updated. The influence of node A on X , referred to as the sensitivity of X to A , or the contribution of A to the reduction of the uncertainty at X , is measured using *Shannon's measure of mutual information* based on the *entropy* function [231][232]. The total uncertainty reduction potential of A , with respect to X , is expressed as [231]:

$$S = H(X) - H(X | A) = -\sum_A \sum_X P(x, a) \log \frac{P(x, a)}{P(x)P(a)} \quad (7.1)$$

where $H(X)$ is the initial uncertainty in X , before receiving the evidence, A , and updating the network, and $H(X|A)$ is the average residual uncertainty in X , summed over all possible values of A .

ESA was used to check the sensitivity of the performance node to all the evidence nodes in the LAP model. This type of analysis is relevant in the context of the LAP model because it is important that performance estimation is not excessively influenced by a particular performance index. The sensitivity of the target node, node 40, "*performance*" (PF), to all the 23 indices (evidence variables) in the LAP model, and the Memory/Feedback node, node 35, "*PreviousPerformance*" (PP), is depicted graphical in Figure 7.3. Nodes 35, "*PreviousPerformance*" (PP), and 8, "*ApplyCorrectFormula*" (ACF), have excessive influence on performance. The influence of node 8, "*ApplyCorrectFormula*" (ACF), was moderated through the adjustment of its contributing weights in the parameters of its children (nodes 10, "*AbilityToMakeApproCalculations*" (MAC), and 37, "*KnowledgeOfConcept*" (KNC), the result of which is highlighted in Figure 7.4. Its influence was reduced by 72.58%. In the process, the influences of nodes 38, "*ConceptualUnderstanding*" (CUN), and 36, "*FactualKnowledge*" (FKN) were acceptably increased. Experts agreed that the level of

influence of node 35, "*PreviousPerformance*" (*PP*), was acceptable so its influence did not require moderation. Also, the levels of influence of nodes 38, "*ConceptualUnderstanding*" (*CUN*) and 36, "*KnowledgeOfConcept*" (*FKN*), on performance was considered acceptable because it is in line with the domain experts' opinion that they should impact equally and significantly on performance.

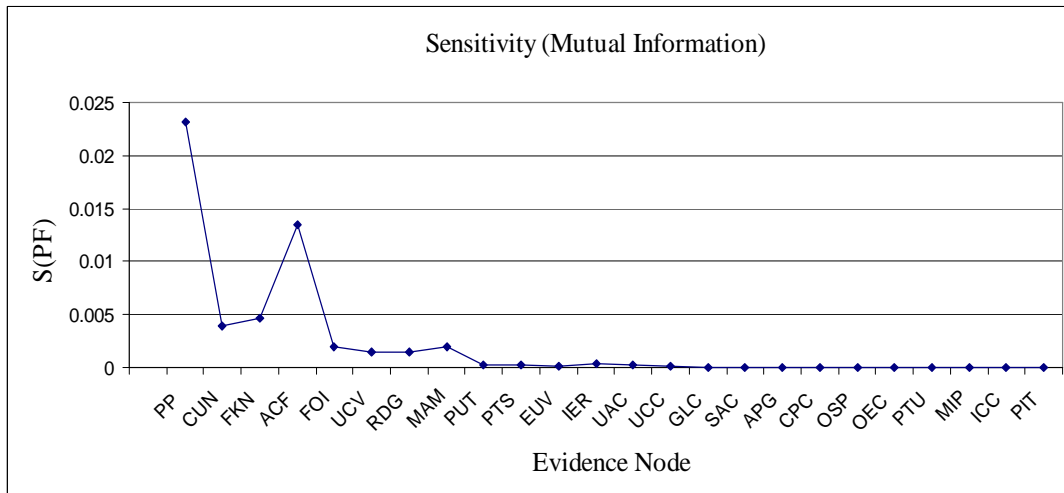


Figure 7.3: Initial sensitivity of the target node to the evidence nodes.

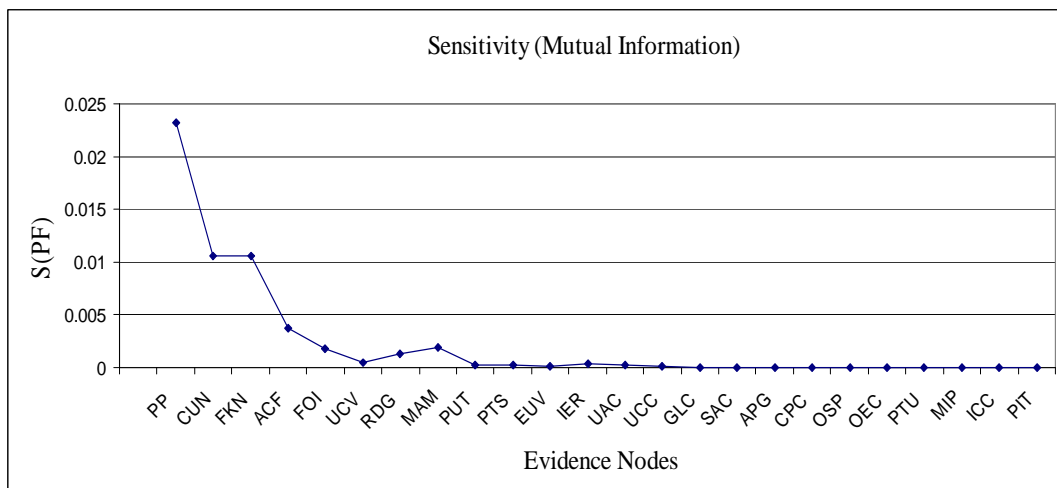


Figure 7.4: Sensitivity of the target node, after parameter adjustment.

The parameters were adjusted in order to reduce the excessive influence of the two nodes, 8 and 35, because it is generally desirable to obtain similar and low sensitivities to the variables, which signifies model robustness. Robustness of a system refers to the quality of the system being able to cope well with abnormalities in input (such as invalid or unexpected inputs), or variations in its operating environment, with minimal alteration or loss of functionality. Consequently, the robustness of a Bayesian network can be defined as the maximization of

operational flexibility and simultaneously the minimization of uncontrolled effects such as disturbances, noise, changes in inputs and/or the state, etc.

Sensitivity (robustness) analysis employs sets of distributions to model perturbations in the parameters of a probability distribution. Robust Bayesian inference is the calculation of bounds on posterior values given such perturbations. This is necessary because, in the real world all the assumptions of a Bayesian model can rarely be met. First, there are the imperfections in the elicited expert beliefs, as it may not be possible to elicit exact probability values. Second, domain experts have the tendency to disagree among themselves, each specifying a differing set of probability distributions. There is some empirical evidence that Bayesian networks are not too sensitive to parameters [233]. This is mostly true in cases where changes in one variable do not affect many variables (such as in sparsely inter-connected networks), and changes are not large relative to the magnitude of the probability values, then it is likely that the changes may not produce significant variations in inferences and robustness may be present [234]. Otherwise, robustness may be an issue, especially in situations where the network is heavily inter-connected, in which case the robustness analysis aspect of inference has to be explored [234].

7.2 EVALUATION

7.2.1 Reliability

Reliability is concerned with the extent to which an assessment, if repeated, would give the same results [235]. It is an estimation of the consistency or repeatability of an assessment. An assessment can be considered reliable if the assessment outcome for two students with similar abilities/aptitude is similar, for the same assessment scenarios. If an assessment is item-based and the assessment outcome a single score, the conventional reliability evaluation method is to use a test-retest method to generate two sets of assessment data or to use a single test and partition the test items into two subsets and check if students scores on one subset correlate with their scores on the second subset. [95] asserted that assessment models for cognitive constructs cannot be appropriately evaluated for reliability using this conventional test-retest method but did not give the reason for the assertion.

Another method of reliability evaluation is the derivation of *predictive accuracy* measure, based on the use of training and test data sets [93][236]. This method is most appropriate for

models induced from data [93]. It is therefore not applicable to LAP model reliability testing. To evaluate the reliability of the LAP model, *simulated students* evaluation approach was adopted. This approach has been used by other researchers for the evaluation of assessment models and systems including [95][237]. It has also been used for other purposes such as improvement of teaching skills [238][239], instructional development testing [239], assessment of document retrieval system [240], and training/testing of neural network-based models [237].

Simulated students are models of human learning and have been categorized based on their granularity: large and fine-grained [239]. The granularity of a simulated student refers to the amount of detail in its representation [239]. Large-grained simulated student models, with known states of performance indices and scores for the performance criteria, and therefore known overall performance scores, were constructed and used for the evaluation. The overall performance score is arrived at by normalizing the sum of the contributing link weights of the known states of the performance indices with respect to the maximum link weights of the nodes. The LAP model was run for each of the simulated students, with the evidence variables instantiated to their known states, in order to obtain the model's performance belief estimates for the students.

There was a strong positive correlation between the students' known performance scores and the estimated beliefs ($r = 0.87685$, $p = 0.0000$). For reliability, "an assessment with a correlation coefficient that is less than 0.70 is generally not considered suitable for individual student evaluations" [95]. A high correlation between scores obtained from two assessments indicates that the assessment is reliable [241]. Where the known performance scores had mean, $\mu = 51.79$ and standard deviation, $\sigma = 18.11$, the estimated measures of belief in performance had $\mu = 50.23$ and standard deviation = 14.46. Figure 7.5 is a graphical depiction of the known performance scores compared to the estimated measures of belief in performance.

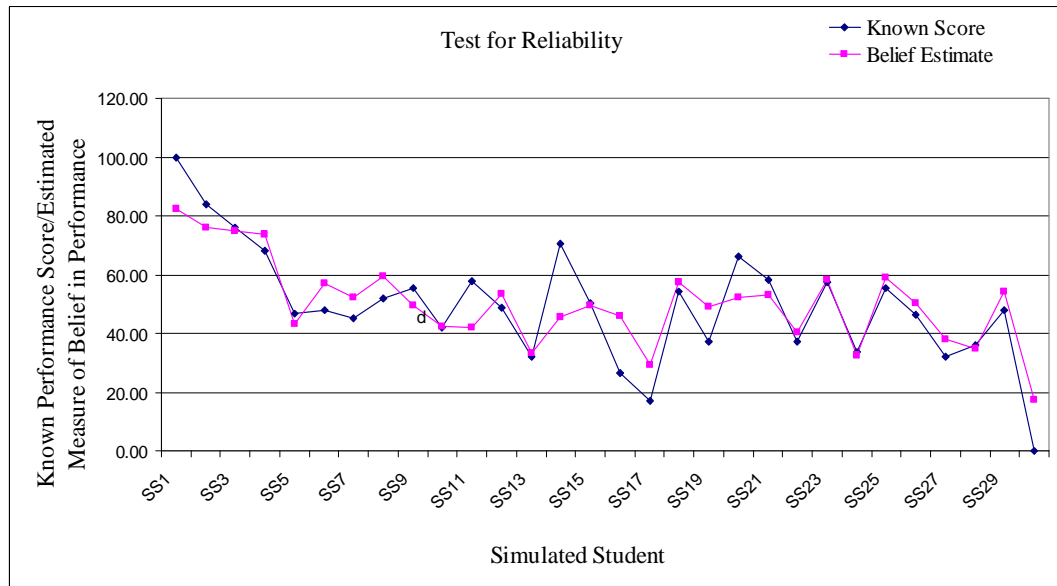


Figure 7.5: Test for reliability.

7.2.2 Validation

The different types of validation checks include face, content, criterion, and construct validity [95]. *Face* validation checks if, on the face value, an assessment process and the outcome appear relevant to the assessed construct. *Content* validation checks the assessment tasks for relevance and representativeness of the domain [242]. Content validity is often established by agreement in the judgements of domain experts [243]. *Criterion* validity is of two types: predictive and concurrent. Predictive validity check is deemed most appropriate for assessment models that predict future performance, such as aptitude tests [95]. Construct validity checks the extent to which an assessment model measures the specific construct for which it is designed [244]. *Concurrent* validation checks the accuracy of assessment outcomes for a set of assessment scenarios.

In this context, domain experts established the face validity of the assessment outcomes of the LAP model, having been key participants in the design and parameterization of the model. Content validity is assumed implicit because domain experts' were actively involved in the design of the assessment tasks. Predictive validation is not relevant to this work because the LAP model is not for use to predict future performance. The focus was on concurrent validity.

Evaluating the concurrent validity of an assessment involves administering another assessment of the target construct, using a similar assessment model and comparing the two assessment outputs. Alternatively, the historical-case technique can be employed. The

procedure is to assess a set of historical scenarios, using the assessment model being evaluated, and then check how well the model's assessment outcomes match the existing assessment outcomes for the scenarios [95]. This is may be feasible where historical data exists and where the assessment methodology and concepts do not represent a paradigm change. In such situations, the assessment model would have been designed to model that historical case or a similar case. Another possible method is the *physical-observation* validation technique which entails that students physically undertake a set of assessment tasks while the instructor physically observes and scores their performance based on the observations. The performance scores then serve as a "gold standard" for the validation of the assessment outcome from the model, based on the same set of assessment tasks. This method is time, space, and assessor intensive, especially for large class sizes, and is logistically demanding.

The LAP model was evaluated using the virtual-observation technique. The validation process involved a team of four assessors. The assessors, rather than physically observe the students themselves, used the "observations" (students' behaviour logs) made and recorded by the events tracker/recorder component of the VEL, while students' were undertaking laboratory activities in the VEL environment, to assess students' laboratory work performance. They used the four performance data source instruments that make up the behaviour log to score a set of performance indicators, based on their respective sets of scoring criteria.

The assessors were familiarised with the research work, the nature and content of the assessment instruments, and the assessment process. Assessors worked individually in the same physical space, with minimal discussions between them to avoid any undue influence on each other. One of the researchers facilitated the assessment process by answering assessors' questions and clarifying any relevant seemingly unclear issues. The assessors each assessed the same laboratory activities for each of 52 students, with respect to a set of performance indicators and their related criteria. The focus was on seven of the nine performance indicators in the LAP model, considering time and cost. The LAP model was then used to assess the same laboratory activities, for the 52 students, using the same set of performance data sources.

As part of the validation process, it was sought to answer the following questions:

- *Q1*-- are the assessors consistent in their assessment (assessor consistency)?
- *Q2*-- to what extent do the assessors agree with each other (inter-assessor correlation)?

- *Q3*-- to what extent do the assessor-assigned students' performance scores agree with the LAP model estimated beliefs in students' performance (assessor-model correlation)?

7.2.2.1 Assessor Consistency

In order to verify assessor consistency, each assessor assessed one of the laboratory activities twice for each student, using the same set of performance data sources. Scores from the two independent assessments were then compared. At $\alpha = 0.05$ and two tailed significance tests, the assessors were statistically significantly consistent in their assessment, with respect to the assigned scores for the performance indicators. Assessors *A1*, *A2*, and *A3* were highly consistent ($0.50878 \leq r \leq 0.84319$, and $0.000000 \leq p \leq 0.000118$). Assessor *A4* was moderately consistent ($0.33141 \leq r \leq 0.67701$, and $0.000000 \leq p \leq 0.024006$), providing a positive answer to *Q1*. Figures 7.6, 7.7, 7.8, and 7.9 highlight the 1st and 2nd assessment scores for the performance factor, *WWC*, by assessors *A1*, *A2*, *A3*, and *A4*, respectively. The remainder of the diagrammatical representations of assessor consistency are presented in Appendix D. Also, the correlation factors and the p-values for the repeated assessment, for all assessors and all the assessed performance indicators, are presented in Appendix D.

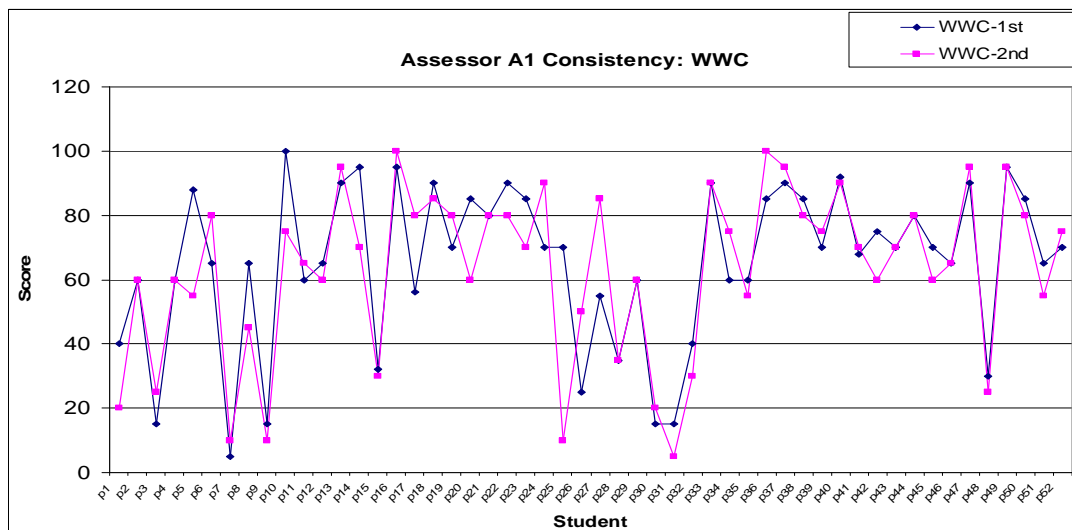


Figure 7.6: Assessor A1 1st and 2nd assessment scores for the performance factor, *WWC*

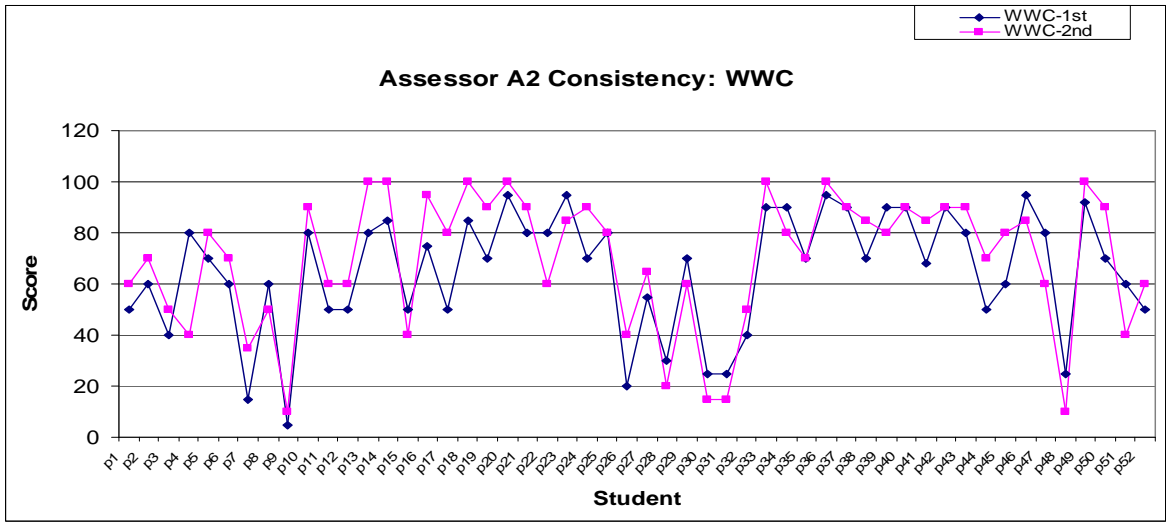


Figure 7.7: Assessor A2 1st and 2nd assessment scores for the performance factor, WWC

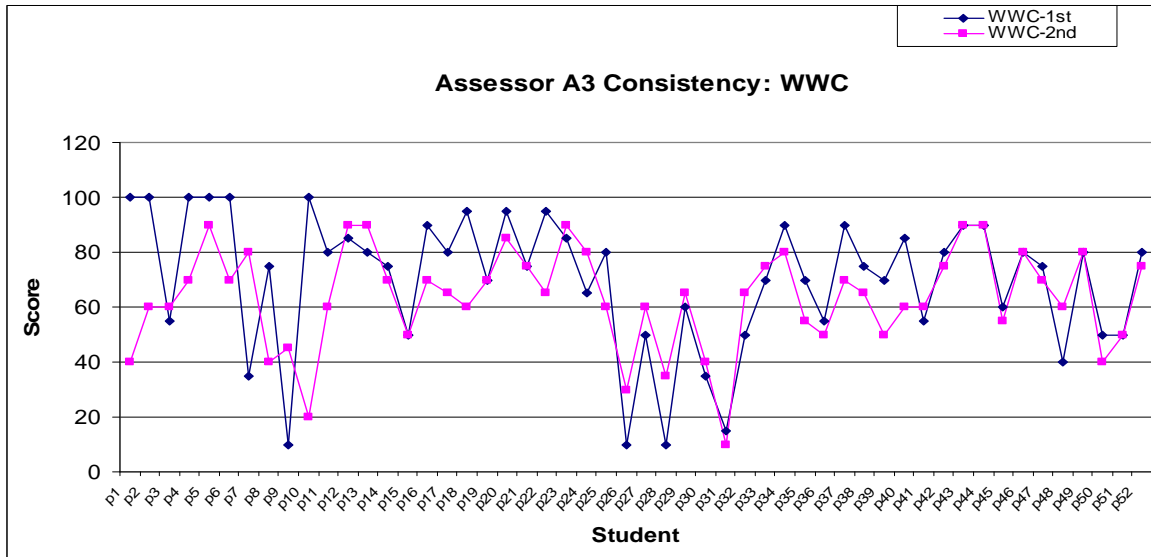


Figure 7.8: Assessor A3 1st and 2nd assessment scores for the performance factor, WWC

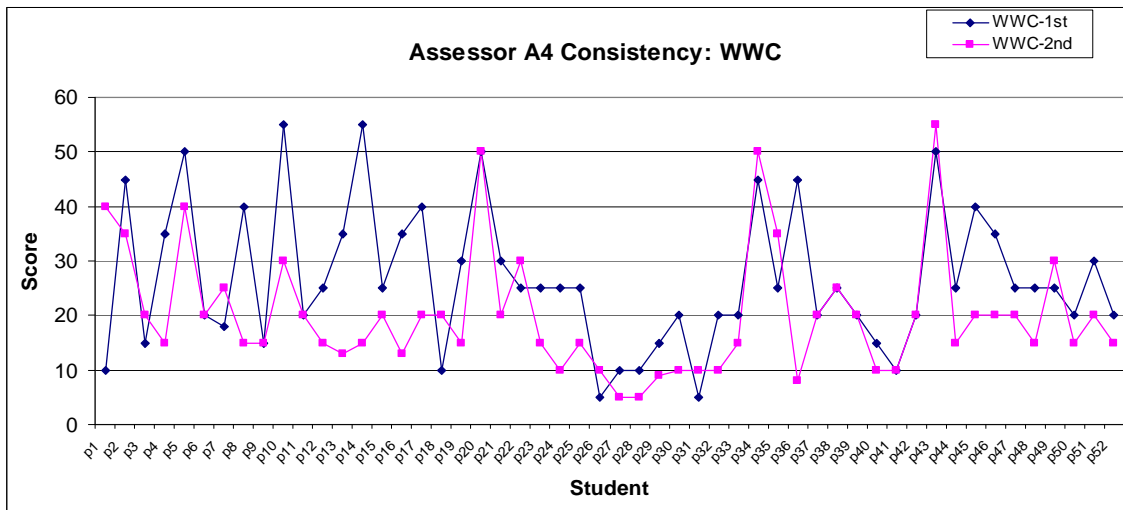


Figure 7.9: Assessor A4 1st and 2nd assessment scores for the performance factor, WWC

7.2.2.2 Inter-Assessor Correlation

Next, inter-assessor scores correlation was verified, in order to answer question *Q2*. Table 7.1 lists the inter-assessor scores correlation factors, r , with respect to a set of performance indicators. *A1* was in agreement with *A2*, on average, 75.63% of the time, with *A3*, 64.88%, and with *A4*, 33.94%. *A2* was in agreement with *A3* 67.86% of the time, and with *A4*, 37.77%. *A3* agreed with *A4* only 41.99% of the time, on average. Assessors *A1*, *A2*, and *A3* agreed with each other more than 64% of the time. Figures 7.10 and 7.11 show the assessors scores for the performance indicators *PUE* and *WWC*, respectively. The remainder of the figures showing the assessors scores for the performance indicators *SRC*, *MAM*, *IER*, *MAC*, and *APG* can be found in Appendix E.

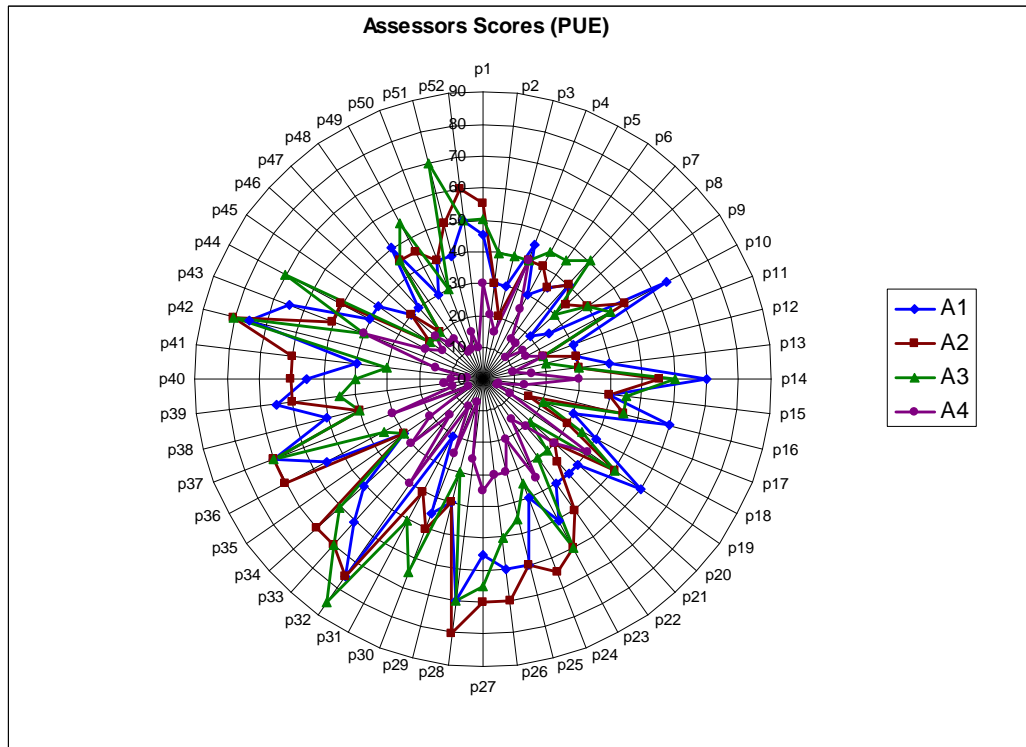
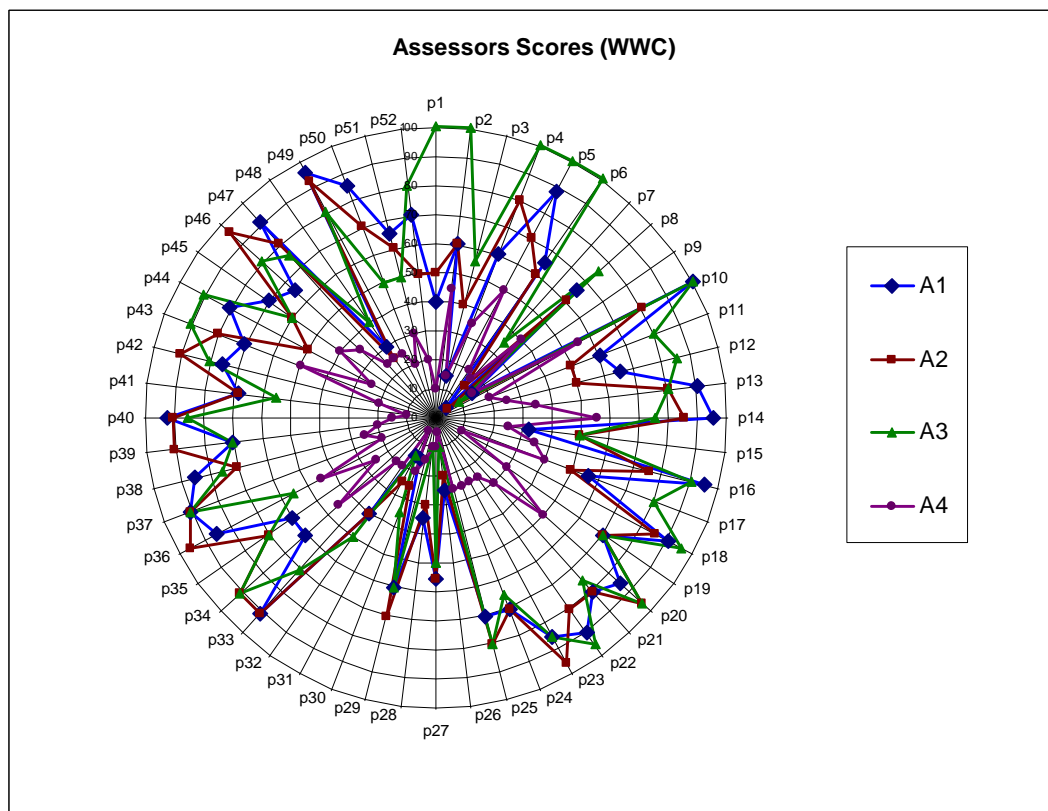
Figure 7.10: Assessors scores for the performance factor, *PUE*Figure 7.11: Assessors' scores for the performance factor, *WWC*

Table 7.1: Inter-Assessor correlation factors, r .

Performance Indicator	A1			A2		A3
	A2	A3	A4	A3	A4	A4
pue	0.75808	0.60385	0.40658	0.74067	0.32004	0.38460
src	0.62128	0.74617	0.31662	0.64266	0.22674	0.42388
wwc	0.85274	0.69625	0.44976	0.69267	0.44050	0.50168
mam	0.76539	0.67283	0.41472	0.74912	0.53092	0.43997
ier	0.77081	0.50905	0.39649	0.61171	0.46481	0.43096
mac	0.73432	0.52202	0.16000	0.60404	0.35095	0.43527
apg	0.79147	0.79111	0.23181	0.70951	0.30988	0.32324

7.2.2.3 Assessors vs. LAP Model

In order to answer question $Q3$, the LAP model was used to assess the same laboratory activities as the assessors and the assessment outcomes were compared to those of the assessors. Performance data were extracted from the same sources used by the human assessors, for the quantization of evidence variables, which were then instantiated into findings or hard evidence. The findings were entered into the network and the model updated. The estimated beliefs for the target nodes were recorded and compared to the scores assigned by the human assessors. The human assessors had access to the same set of assessment instruments and data as the LAP model, and used the same set of criteria for scoring the performance indicators. Hence, the human performance assessment outcomes were used as the “gold standard”, for the validation of the assessment outcome of the LAP model.

Comparison of the human assessor assigned scores and the corresponding belief estimates by the LAP model indicated strong positive, statistically significant, correlations between the human assessment outcomes and the LAP model performance belief estimation outcomes. Figures 7.12, 7.13, 7.14, and 7.15 show assessors A1, A2, A3, and A4 scores, respectively, for the performance factor, PUE , against the corresponding belief estimates by the LAP model. Figure 7.16 shows the average scores by all the assessors, for the performance factor PUE , against the belief estimate for PUE by the LAP model. Similarly, Figures 7.17, 7.18, 7.19, and 7.20 show assessors A1, A2, A3, and A4 scores, respectively, for the performance factor, WWC , against the corresponding belief estimates by the LAP model, while Figure 7.21 shows the average scores by all the assessors, for the performance factor WWC , against the belief estimate for WWC by the LAP model. Other assessors’ scores compared to model belief estimates are presented in Appendix F.

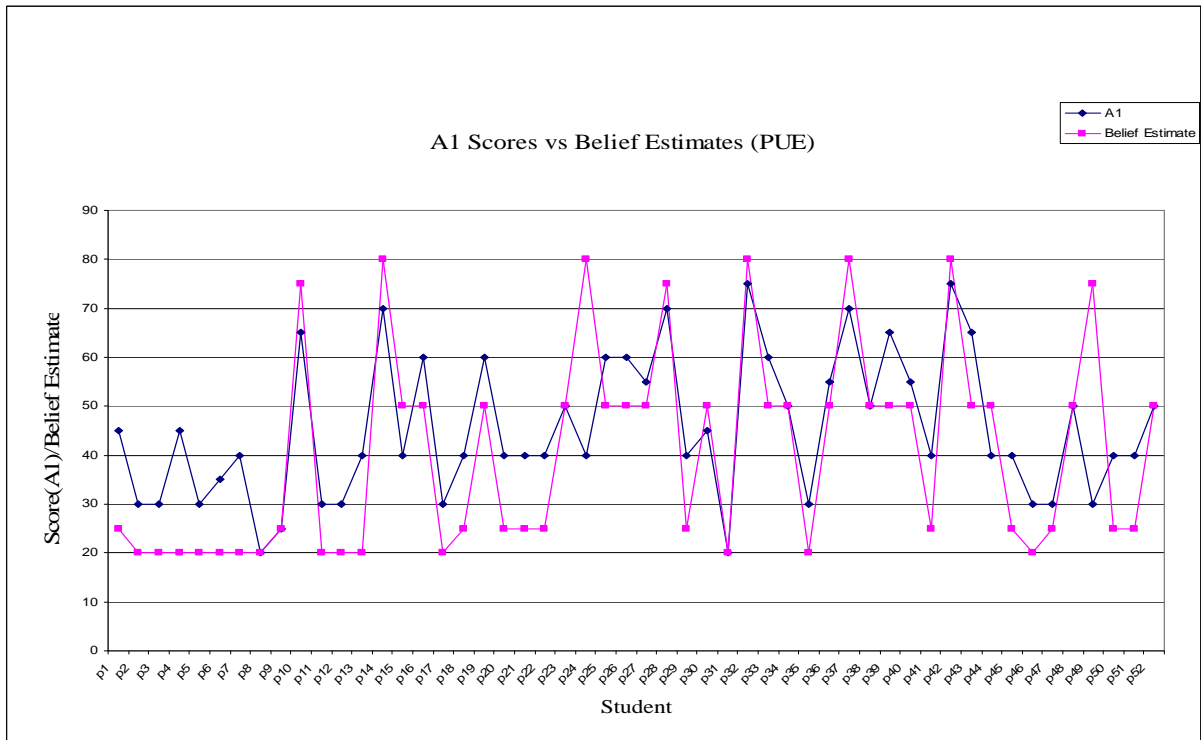


Figure 7.12: Assessor A1 scores vs. model belief estimates for *PUE*.

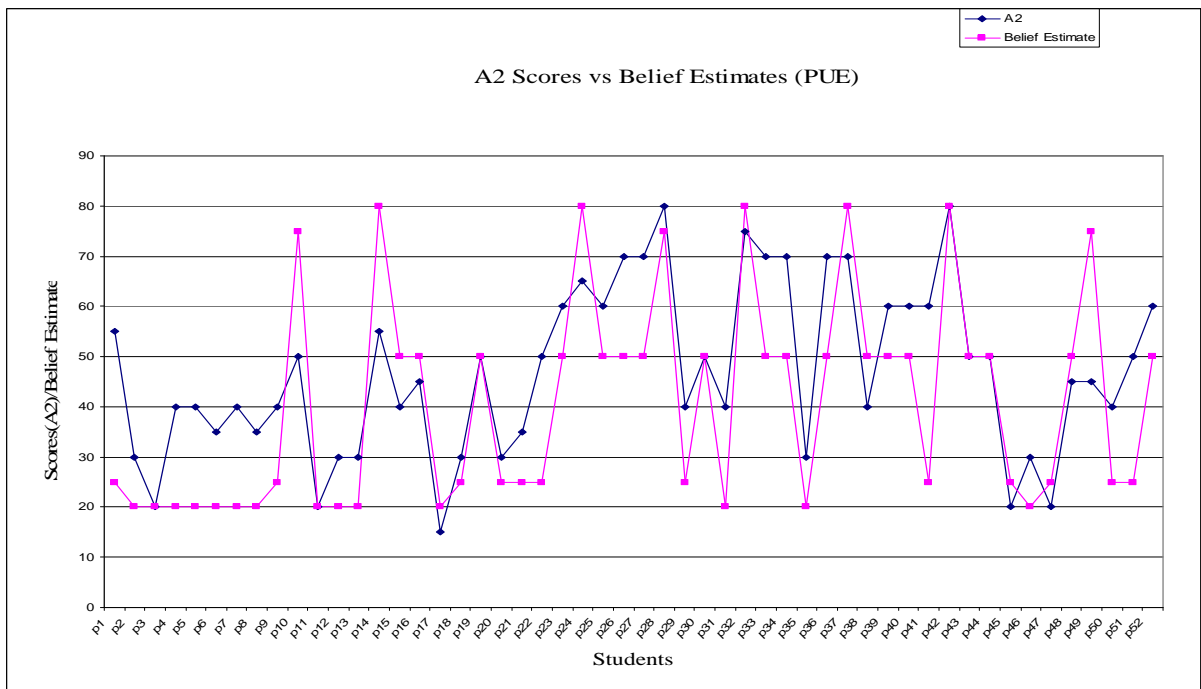


Figure 7.13: Assessor A2 scores vs. model belief estimates for *PUE*.

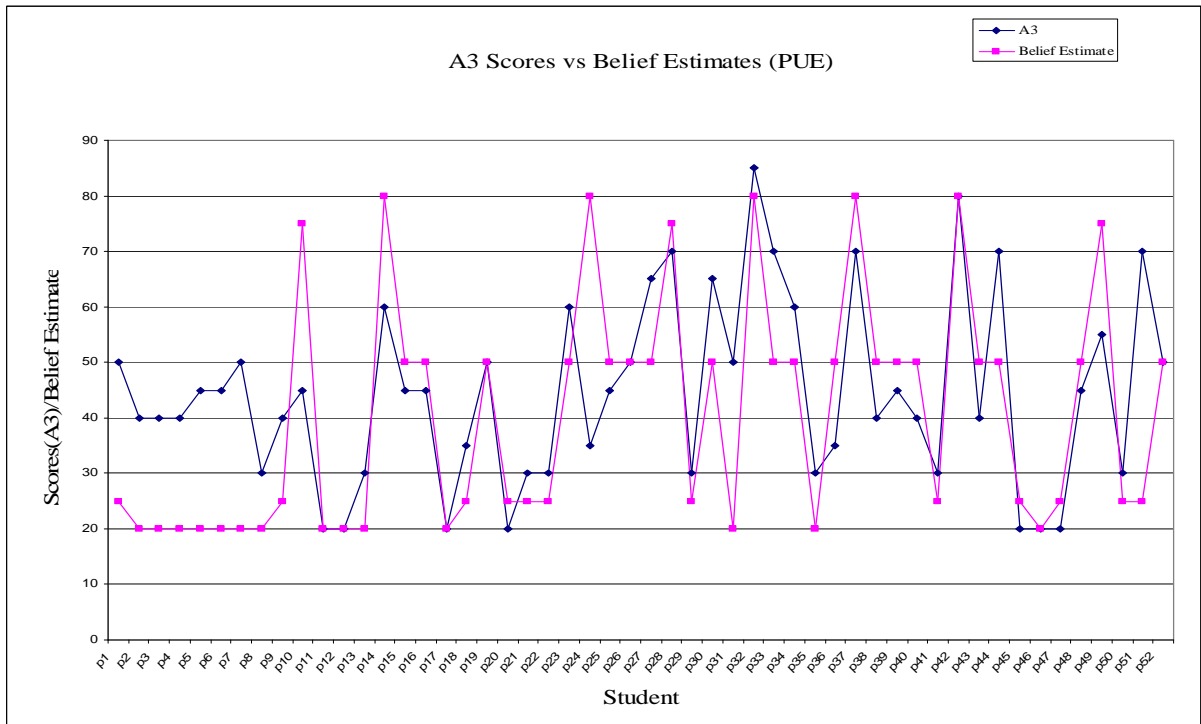


Figure 7.14: Assessor A3 scores vs. model belief estimates for *PUE*.

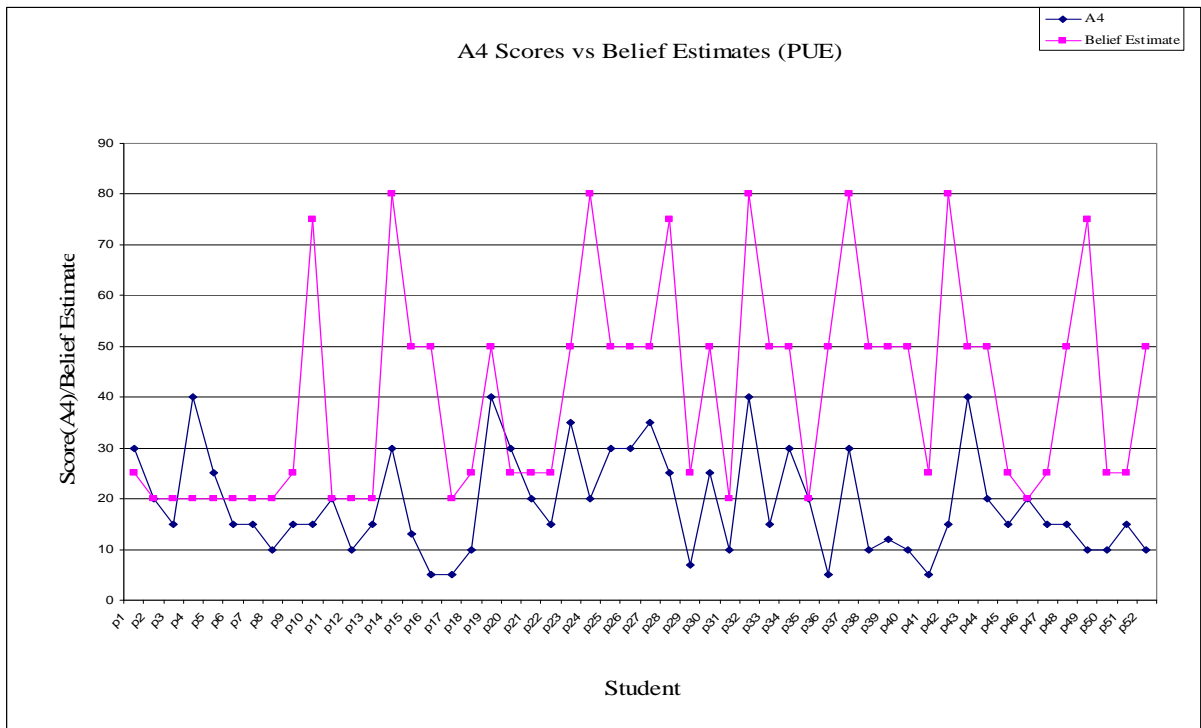


Figure 7.15: Assessor A4 scores vs. model belief estimates for *PUE*.

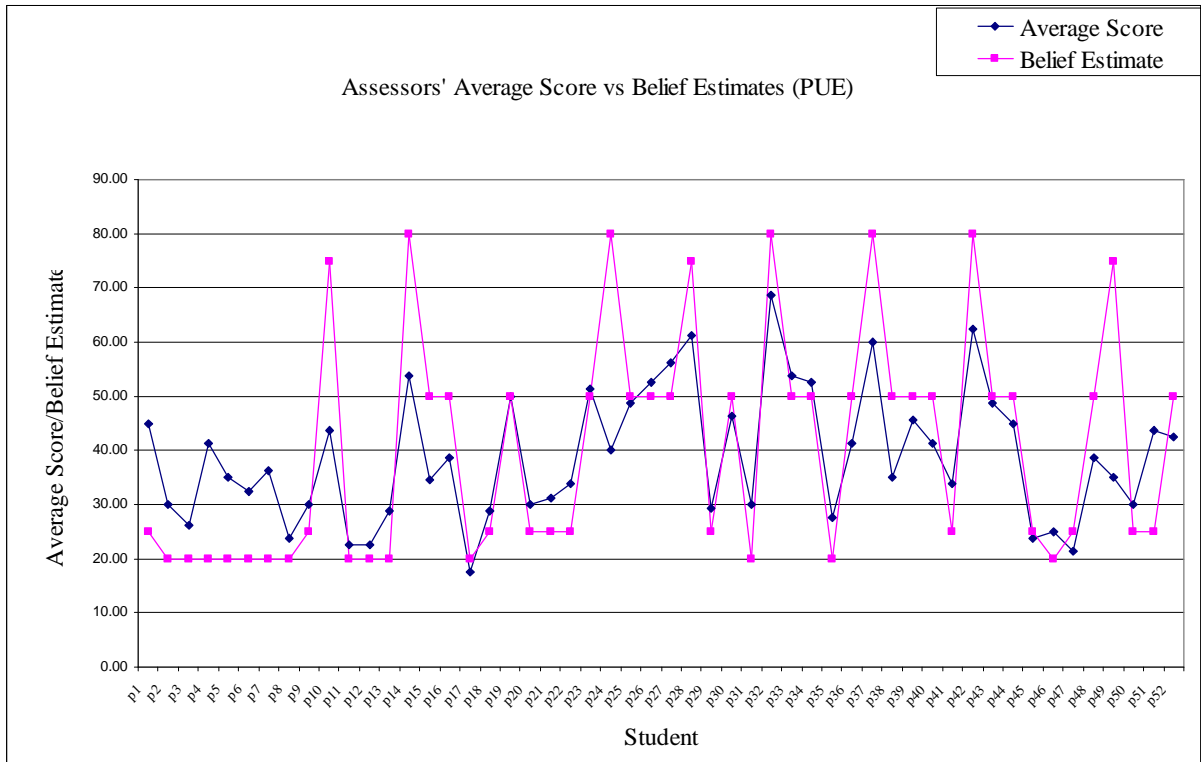


Figure 7.16: Assessors average scores vs. model belief estimates for PUE.

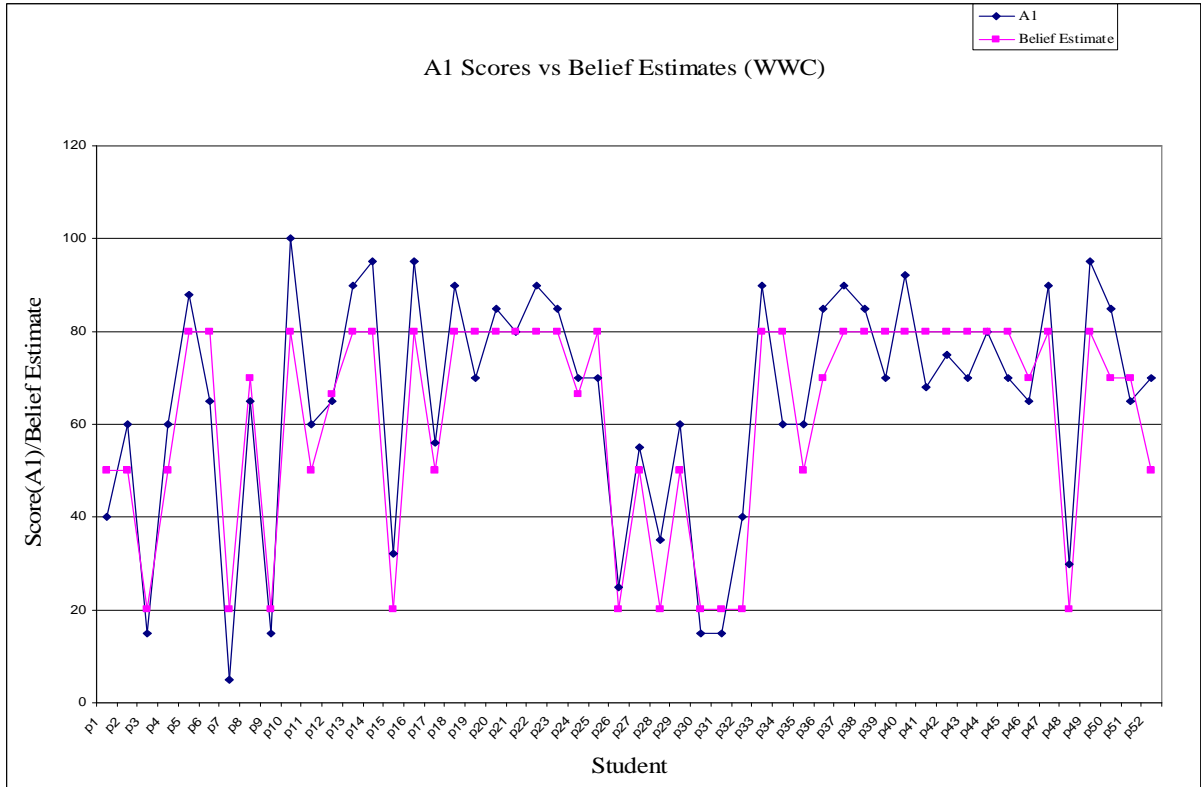


Figure 7.17: Assessor A1 scores vs. model belief estimates for WWC.

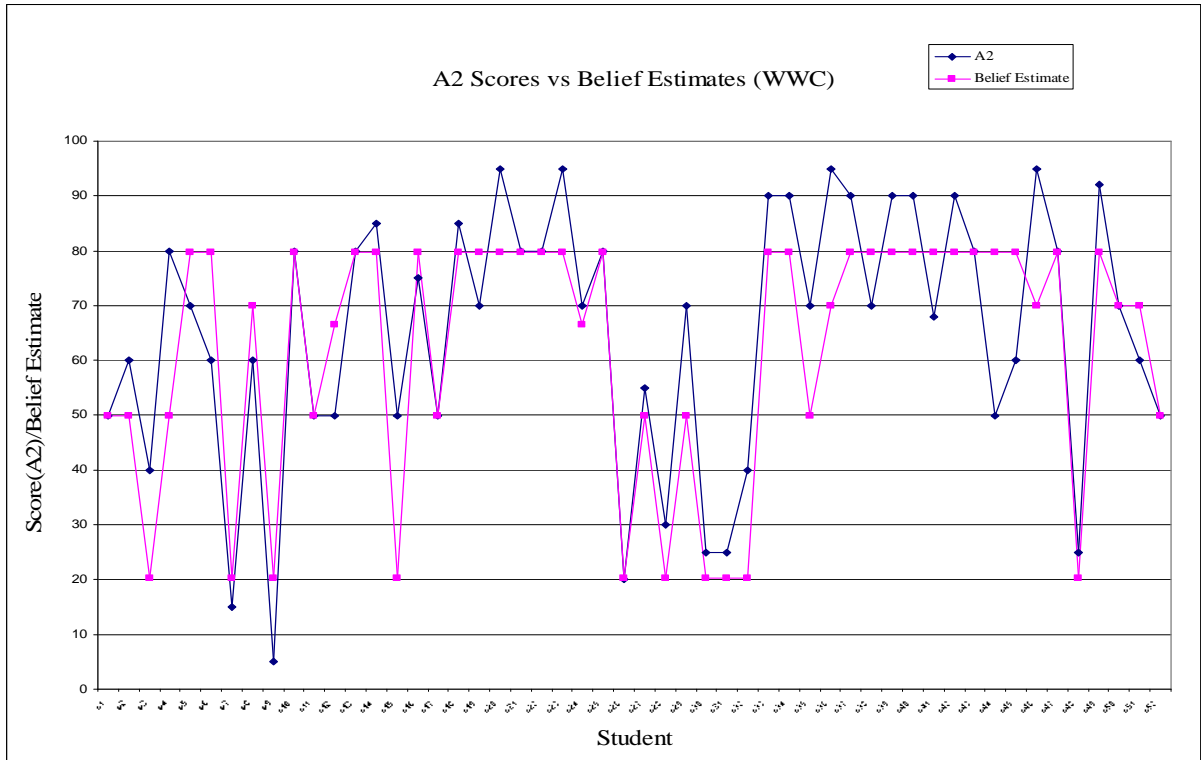


Figure 7.18: Assessor A2 scores vs. model belief estimates for WWC.

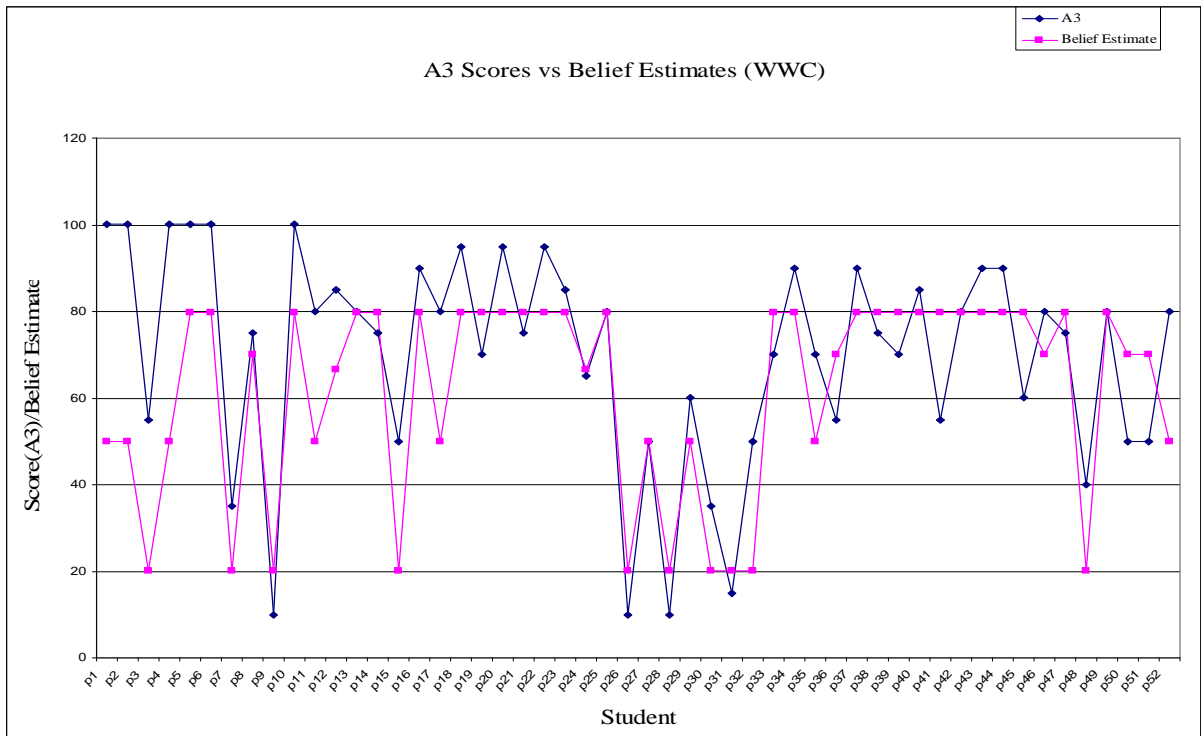


Figure 7.19: Assessor A3 scores vs. model belief estimates for WWC.

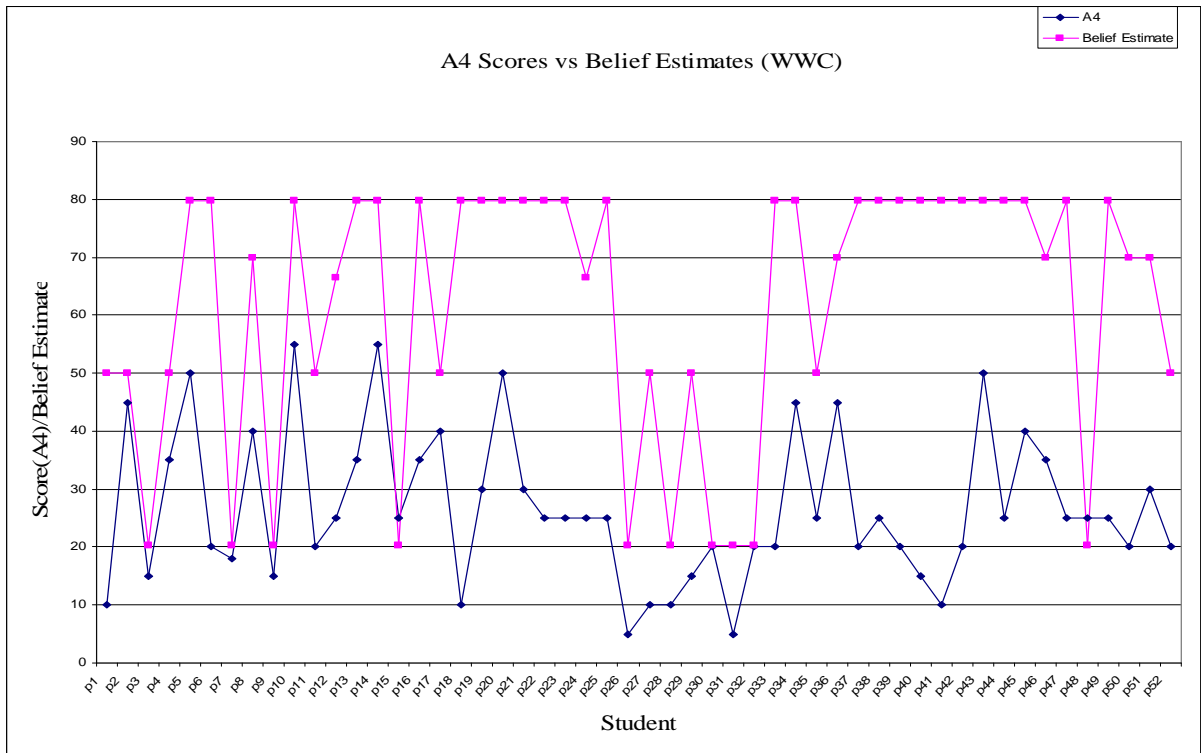


Figure 7.20: Assessor A4 scores vs. model belief estimates for WWC.

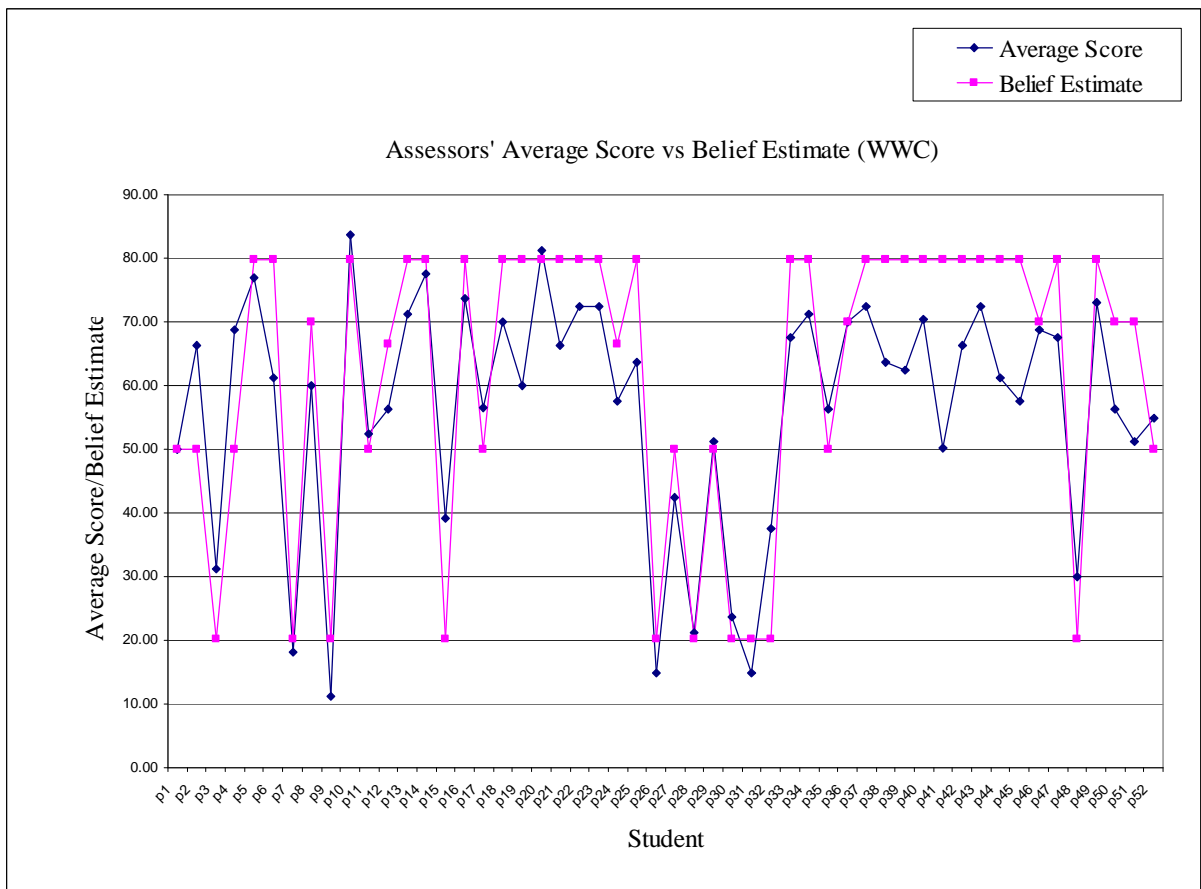


Figure 7.21: Assessors average scores vs. model belief estimates for WWC.

The above figures give a pictorial representation of the assessors score assignments, based on specific criteria, to the individual performance factors, and the comparison of the scores to the model estimated measures of belief in the hypotheses represented by the performance factors. Each assessor's scores for all the assessed performance factors were averaged to obtain the generalized performance score by that assessor, with respect to the assessed performance factors only. Figures 7.22, 7.23, 7.24, and 7.25 show assessors A1, A2, A3, and A4 performance scores, with respect to the seven assessed performance factors, against the belief estimates for performance, given the evidencing of the performance indices for the seven assessed performance factors, while Figure 7.26 shows the average performance scores by all the assessors, against the estimated measure of belief in performance by the LAP model.

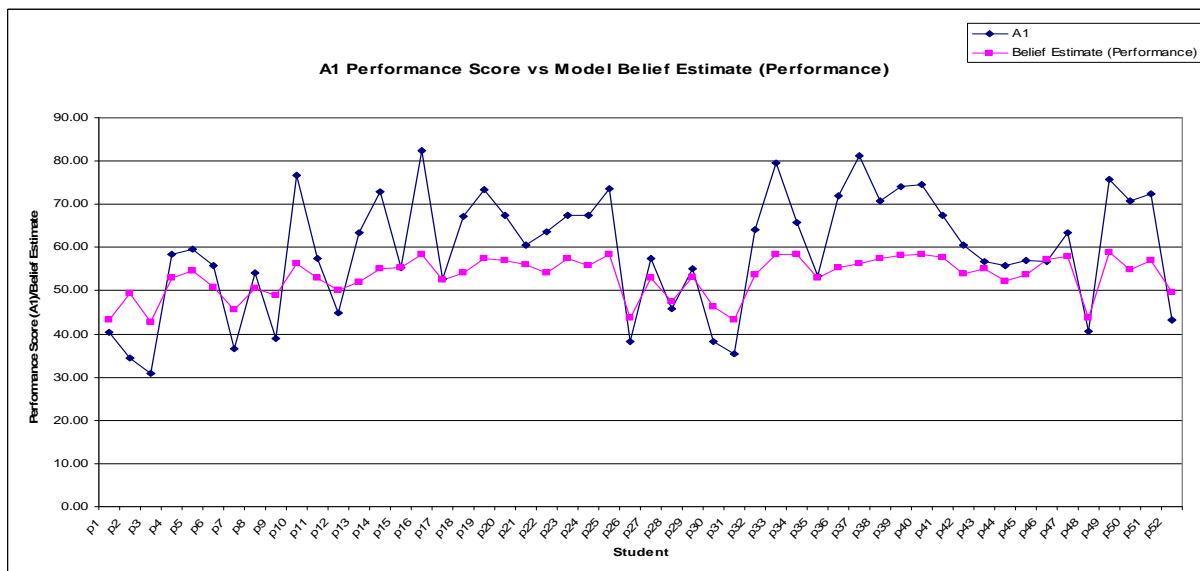


Figure 7.22: Assessor A1 performance scores vs. model belief estimates for *Performance*.

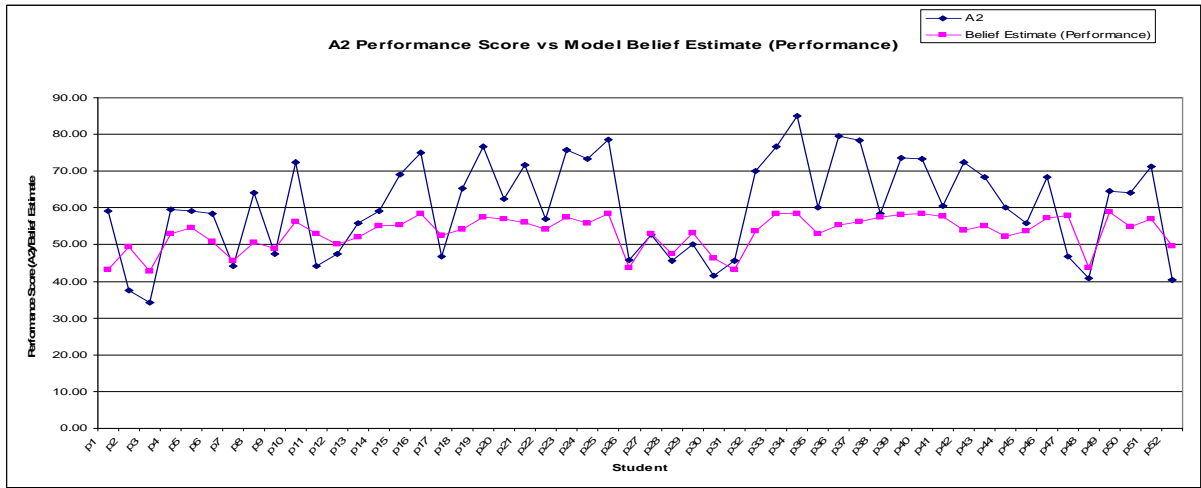


Figure 7.23: Assessor A2 performance scores vs. model belief estimates for *Performance*.

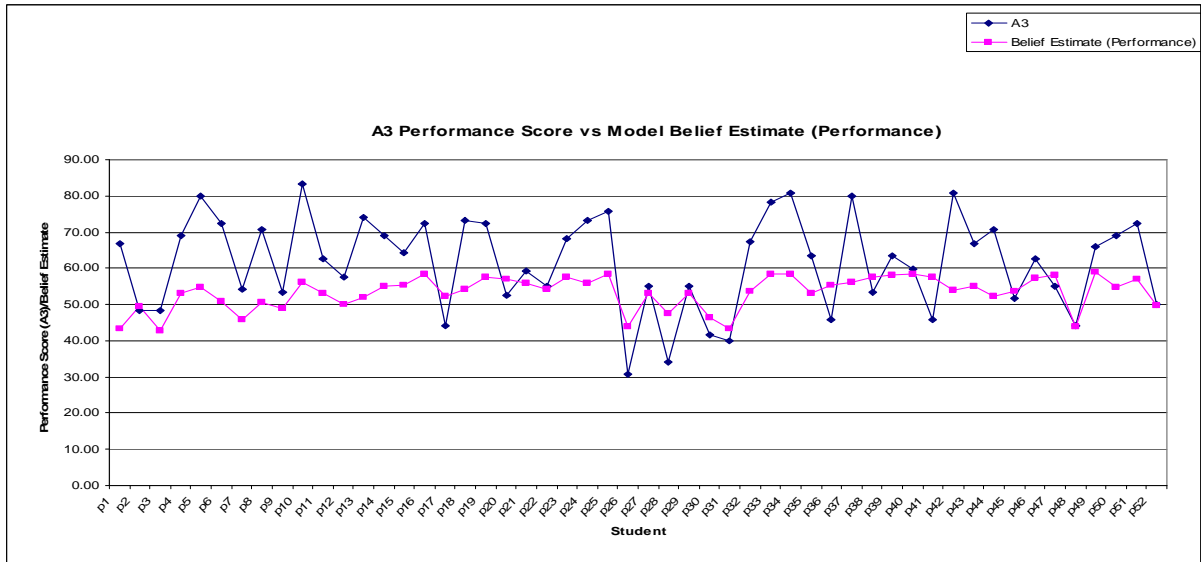


Figure 7.24: Assessor A3 performance scores vs. model belief estimates for *Performance*.

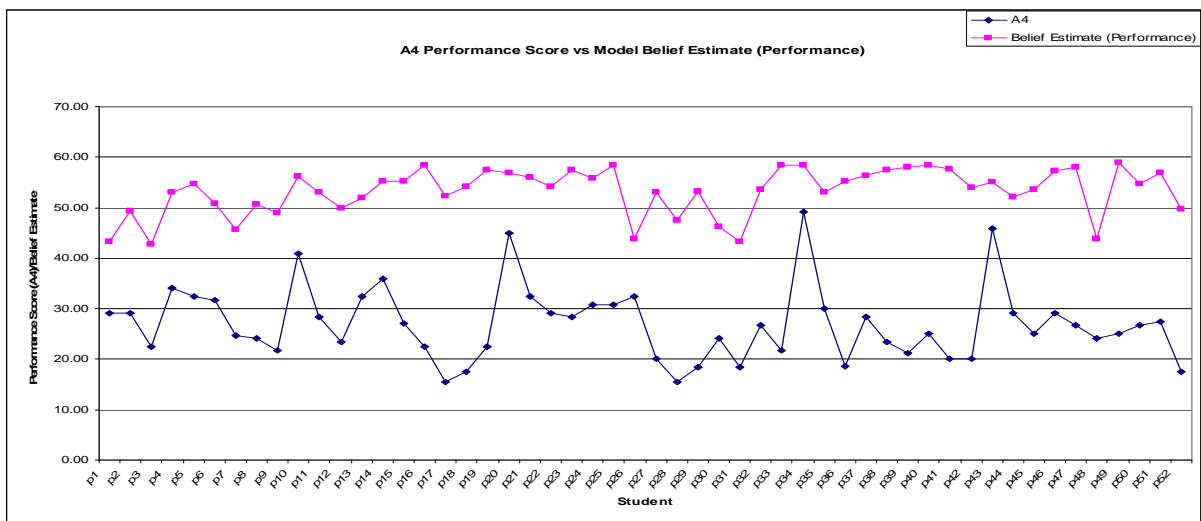


Figure 7.25: Assessor A4 performance scores vs. model belief estimates for *Performance*.

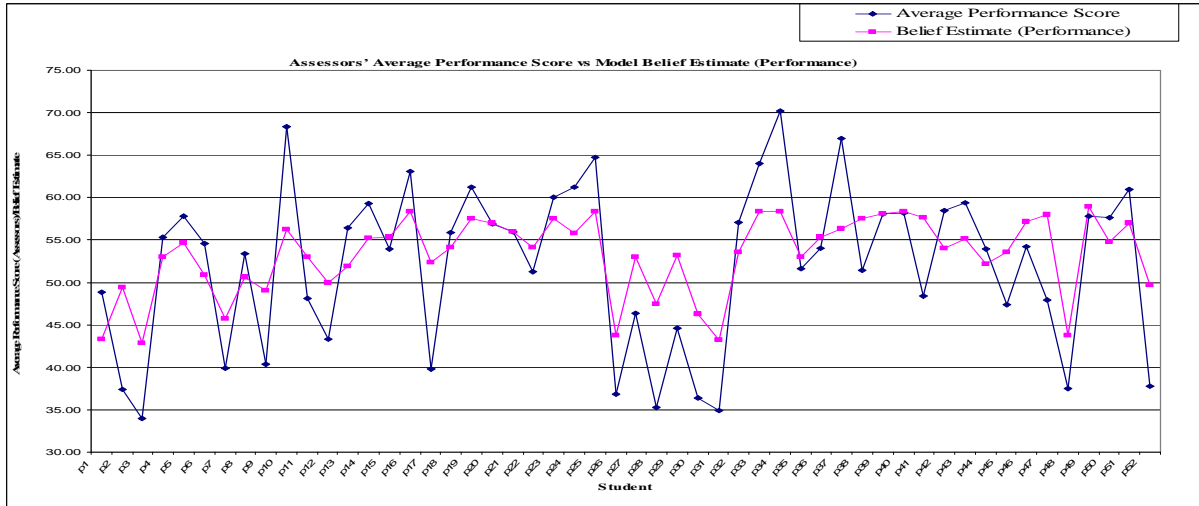


Figure 7.26: Assessors average performance scores vs. model belief estimates for *Performance*.

It was also sought to find the association between the assessors’ scores assignment to the performance factors, *PUE*, *WWC*, *SRC*, and *MAC*, and *PERFORMANCE* and the measures of belief in the hypotheses estimated by the LAP model, using bivariate analysis, in the form of Pearson product-moment correlation measure. There were strong positive correlations between them. The relationships were statistically significant. Table 7.2 highlights the correlation factors and p-values for the associations.

Table 7.2: Correlation Factors and P-values of Assessors Scores for *PUE*, *WWC*, *SRC*, *MAC*, and *Performance*, Compared to the Model Belief Estimates.

Hypothesis	Model Belief Estimate				Assessors Average Score
	A1	A2	A3	A4	
<i>PUE</i>	$r = 0.773485$ $p = 0.000000$	$r = 0.742712$ $p = 0.000000$	$r = 0.65564$ $p = 0.000000$	$r = 0.269926$ $p = 0.052985$	$r = 0.777951$ $p = 0.000000$
<i>WWC</i>	$r = 0.909431$ $p = 0.000000$	$r = 0.847425$ $p = 0.000000$	$r = 0.715468$ $p = 0.000000$	$r = 0.438325$ $p = 0.001154$	$r = 0.889819$ $p = 0.000000$
<i>SRC</i>	$r = 0.845432$ $p = 0.000000$	$r = 0.564213$ $p = 0.000013$	$r = 0.695996$ $p = 0.000000$	$r = 0.249134$ $p = 0.074944$	$r = 0.761581$ $p = 0.000000$
<i>MAC</i>	$r = 0.767547$ $p = 0.000000$	$r = 0.838831$ $p = 0.000000$	$r = 0.740793$ $p = 0.000000$	$r = 0.366913$ $p = 0.007463$	$r = 0.882051$ $p = 0.000000$
<i>Performance</i>	$r = 0.896886$ $p = 0.000000$	$r = 0.754626$ $p = 0.000000$	$r = 0.552468$ $p = 0.000022$	$r = 0.231143$ $p = 0.099276$	$r = 0.801572$ $p = 0.000000$

7.3 MODEL APPLICATION

This section highlights areas of intended application of the LAP assessment model. The model was built for the purposes of estimating measures of belief in a student's abilities/skills, grasp of addressed concept, and performance, with respect to a laboratory activity. An overview of the belief estimation process has been given in Chapter 4 (section 4.4). The application of the model is based on entering the findings at the evidence nodes, updating the network and reading off the estimated beliefs (posterior probability distribution) of the target nodes. Findings are generated based on the performance data extracted from the student's laboratory work behaviour log.

Generally, the LAP model is aimed at being used for the following types of assessment, for a laboratory activity: instantaneous, cumulative and selective.

Instantaneous assessment, in this context, requires that:

- previous belief estimates, based on previously entered evidence, are cleared from the network and the network reset to its original state with fair priors.
- the model is used without the memory/feedback node (the node is deactivated),
- new findings are entered into the model and inference made in order to obtain belief estimates for target variables.

Cumulative assessment, in this context, means that assessment is based on the combined effects of:

- previous belief estimates (previous belief estimates are absorbed into the network),
- past performance outcomes (the model is used with the memory/feedback node),
- new findings, which are entered into the model and inference made in order to obtain belief estimates for target variables.

Selective assessment, in this context, means that the instructor determines what is assessed.

The instructor may:

- choose to test only specific aspects of a student's laboratory work
- turn off/on different nodes of the LAP model
- choose to instantiate only a specific set of evidence variables.

In this case, the model serves as an inquiry tool to possibly diagnose areas of improvement. These assessment options are selectable from the instructor interface GUI (Figure 7.27).

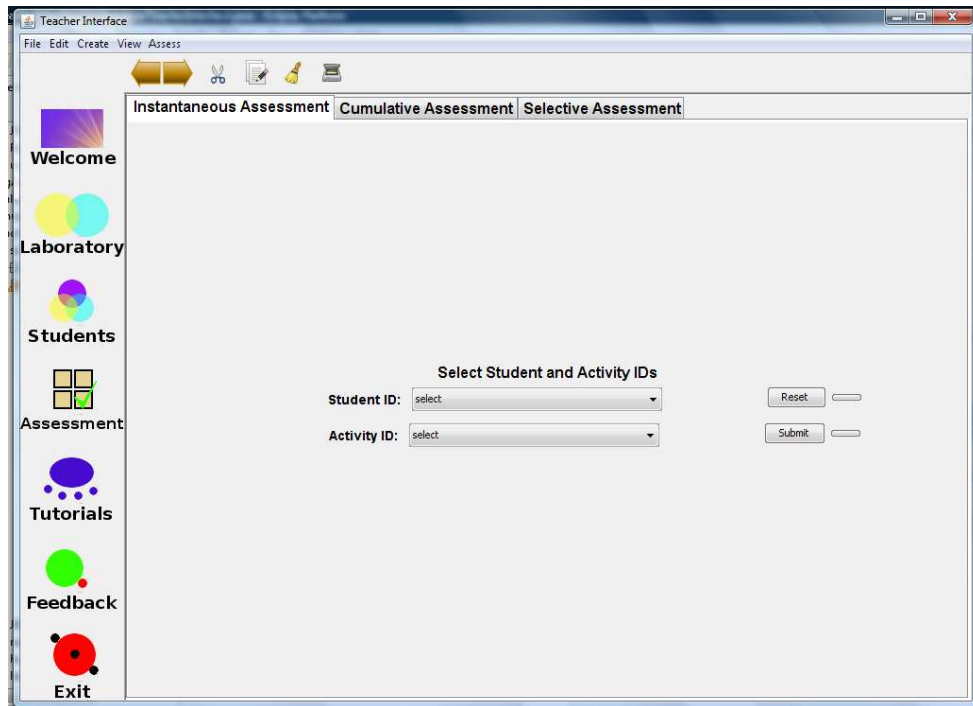


Figure 7.27: A view of the instructor interface GUI.

There are 23 evidence variables in the LAP model. It may not be possible to generate findings for all the evidence variables for every laboratory activity because a laboratory activity may not present the scenarios that will elicit some relevant constructs. Moreover, it may not be necessary to assess everything all of the time. Each laboratory activity does not necessarily have to generate performance data to evidence all the observable nodes and the decision to enter all or some of the findings, at any assessment instance, depends on the instructor. For example, if a laboratory activity does not involve calculations, then nodes 8 (*ACF*) and 9 (*UCV*) (see Figure 6.9) cannot be evidenced. Also, the instructor may seek to evidence a specific set of variables in order to make inference about a specific variable(s). In order to do this, the instructor may choose the nodes to evidence and the target nodes to query, using the instructor interface GUI which has options to enable the instructor turn the evidencing and querying of variables on and off. For example, the instructor may design an activity that will specifically require students to adapt, then turn off the evidencing of other nodes, except those required to infer a student's ability to adapt. If no variables are turned off, findings are generated for all evidence nodes, where possible, and reasoning is generalized.

At runtime, the topology of the model remains the same while the parameters of the memory/feedback node, node 35 "*PreviousPerformance*" (*PP*), changes after each runtime

instance (model update), for cumulative assessment. The change is defined by the relationship between node 35, “*PreviousPerformance*” (*PP*), and node 41, “*Performance*” (*PF*), is expressed as:

$$P(M_t) = P(H_{t-1} | E_{t-1}) \quad (8.2)$$

where $P(M_t)$ represents the probability distribution of the memory/feedback node at time, t , and $P(H_{t-1} | E_{t-1})$ represents the probability distribution of node 40 (*PF*) at time, $t-1$, given the set of evidence, E , at time, $t-1$. In addition to the parameters of the memory/feedback node changing at each run time instance, the memory/feedback node is clamped to one of its states when the state of node 41, “*Performance*” (*PF*), reaches or exceeded a certain specified threshold value.

7.4 SUMMARY

The evaluation of the BN-based assessment tool for the performance assessment of students' laboratory work has been presented. The evaluation results demonstrate the applicability of the model and the practicability of the proposed virtual-observation based assessment scheme. The possibilities offered by the synergy of the assessment tool and methodology are only limited by the instructor's assessment needs. The next chapter presents a discussion of the work presented in this thesis, its limitations, contributions, and further work, before concluding the thesis.



CHAPTER 8

DISCUSSION, LIMITATIONS, CONTRIBUTIONS, FURTHER WORK, AND CONCLUSION

*Now this is not the end. It is not even the beginning of the end. But it is, perhaps, the
end of the beginning.*

Winston Churchill

1874 - 1965

8.0 INTRODUCTION

This thesis has dealt with the implementation and evaluation of a Virtual Electronic Laboratory (VEL), and the application of probabilistic reasoning AI technique to the construction of an assessment tool, the LAP assessment model, for the performance assessment of students' laboratory work, in the VEL environment. The VEL has been shown to be a feasible instruction aid for enhancing students understanding of taught fundamental engineering concepts. Also, the effectiveness of the LAP model, as an assessment tool, within the context of the VEL environment, has been demonstrated.

Essentially, this chapter summarizes the thesis as follows: section 9.1 discusses the work that has been presented in this thesis, while the limitations of the research work are highlighted in section 9.2; section 9.3 outlines the contributions of the research work; while section 9.4 highlights possible further work and future research directions that build on the contributions outlined in section 9.3. Section 9.5 concludes the thesis.

8.1 DISCUSSION

The VEL was evaluated in one of each of the two categories of engineering faculties, at two separate Universities: an engineering faculty in United Kingdom (UK) (i.e. UoP); and an engineering faculty in Nigeria (i.e. UTON). UTON is challenged by resource constraints and large UE class sizes. As part of the evaluation process, pre-/post-tests were used to assess and quantify the knowledge states of the two sets of student cohorts, before and after undertaking laboratory activities in the VEL environment. The tests results showed that there were positive, statistically significant, differences in the students' knowledge states after the educational intervention, for both student cohorts, with respect to the concepts addressed by the laboratory activities. This shows that the VEL has the capacity to enhance learning, in the light of appropriate learning scenarios. Furthermore, the observed performance differences between the UoP and UTON student cohorts and the greater improvement in the learning of the UTON cohort combine to reinforce the highlighted need for complements and/or alternatives to the traditional laboratory in engineering faculties challenged by resources constraints and large class sizes, which may especially be the case for engineering faculties in developing countries.

Feedback from the student cohorts, about the VEL, indicated an overall positive reaction and the instructors of the course units around which the evaluation laboratory activities were

designed responded well to the virtual electronic laboratory. Students unanimously agreed that the evaluation activities addressed concepts related to the taught course units they had undertaken. The majority of students said the use of the VEL enhanced their understanding of the concepts addressed by the laboratory activities. This is reflected in the PTPT results of the student cohorts. Generally, the feedback indicated a positive and enthusiastic attitude to the VEL, especially at UTON. It was well received and acknowledged as a tool with which to realistically undertake curriculum-based laboratory activities.

The evaluation of the VEL afforded the opportunity to generate the data required for the evaluation of the assessment tool. As students were undertaking laboratory activities in the VEL environment, their behaviours (mouse click streams and key presses) were captured in real-time and logged. Performance data, for the evaluation of the LAP model, were extracted from the behaviour logs.

The construction of the LAP model was undertaken with the committed participation of three domain experts, consisting of a Principal Lecturer, a Senior Lecturer, and a Reader, from the Electronic/Computer and Electrical/Electronic Engineering (ECE and EEE) departments of UoP and UTON, respectively, which have different instruction and laboratory practices. Each expert had over 10 years classroom and instructional laboratory experience. The model was evaluated with the performance data extracted from the behaviour logs of the UTON student cohort because of their larger number, to facilitate drawing more robust conclusions. The assessors involved in the evaluation of the assessment model were domain experts, different from the instructors that taught the course units around which the VEL evaluation laboratory activities were designed, and the domain experts that helped formulate the model. The assessors were unfamiliar with the VEL, the assessment tool, and the proposed assessment methodology. Introductory meetings, demonstrations, and briefing sessions facilitated their understanding and appreciation of the research work, which increased their motivation. After the first assessment session, assessors had time to reflect on the assessment methodology and their experiences, the noticeable effect of which was increased speed of assessment, shorter pause durations, and fewer questions to the researcher. At the end of the assessment exercise, the assessors' feedback on the VEL, the assessment tool, and the assessment methodology were encouraging, indicating that all three had been well received. The assessors acknowledged the capability of the entire system to infuse vigour into engineering laboratory education and change the way students' laboratory work is assessed. They opined that the

assessment tool and scheme can avail the instructor of useful information, with respect to students' laboratory work.

Though the *virtual-observation*-based assessment methodology was relatively time demanding when undertaken by human assessors, it is much less time, space, and logistically demanding than the *physical-observation* scheme, and when automated, the time demand pales into insignificance. The *virtual-observation*-based assessment methodology has the ability to offer the same benefits as the *physical-observation*-based methodology. This ability is further strengthened by the strong positive correlations that existed between the assessors' scores and the belief estimates by the LAP model.

8.2 CHALLENGES

The research work encountered a number of challenges or constraining factors including:

- *Student Participation*: It was difficult getting students to participate in the evaluation of the VEL, especially at UoP, despite the offer of participation tokens. This is probably because students are not in the university to be used for research work but to earn a degree. Therefore, they are mainly interested in assessable learning activities that will earn them inclusive marks.
- *Student population*: The resource conditions and student populations at UoP and UTON, differ significantly, which may not give proportionate effectiveness results.
- *Access to domain experts*: Access to domain experts is limited. Also, it is difficult to bring all the domain experts to a central location for knowledge elicitation, because instructors are busy, each having a totally different time schedule.
- *Paradigm change*: the assessment tool and methodology represent a paradigm change in the way students' laboratory work is assessed, which necessitated attitudinal, perception, and orientation change on the part of the domain experts who are accustomed to thinking of laboratory work assessment in terms of logbooks and written reports. Instead, they were required to think of laboratory work assessment in terms of concepts relating to knowledge, understanding, and abilities/skills.
- *Expert Opinion Disagreement*: Domain experts' disagreement is generally acknowledged. Though methods of resolving expert opinion conflicts and how to obtain composite or consensus opinion are detailed in literature, the disagreements introduced significant delays in the work. The disagreement in domain experts'

opinions, in this context, may be attributed to the fact that instructors, in an assessment, tend to be looking for the same things in different ways.

- *Natural Language Processing (NLP)*: NLP is necessary for free text responses to assessment questions but NLP is outside the scope of this work.
- *Non existence of similar models and historical data*: There were no similar existing models to reference and historical data to use in the empirical investigation for the optimal approach for the construction of the assessment tool.
- *Ethical issues*: Ethical considerations apply to every activity in the University, including research. The student participants had to be given sufficient time and information to enable them decide if they want to be part of it. The volunteers had to sign participation consent forms which they may perceive as connoting commitment on their part. This may have scared away potential participants, thereby limiting the number of actual participants.

8.3 CONTRIBUTIONS

The contributions of this research work include:

1. The conception and implementation of a virtual electronic laboratory, a tool with which the instructor can increase the use of laboratory activities in the instruction process, to support learning and enhance students' understanding of fundamental engineering concepts. This will help lay a more solid foundation for students on which to build in later undergraduate, postgraduate, and professional years. The tool, which can be used to undertake curriculum-based laboratory activities in a realistic manner, may be particularly beneficial to the neophyte engineer challenged by the need to grasp the fundamental concepts. This is a major contribution because although some work on the implementation of virtual laboratories has taken place, they are still not available for the whole spectrum of engineering. For various fields of engineering, research is on-going on virtual laboratories that can emulate the traditional laboratory processes. Hence, more efforts are required because of the numerous diverse experiential needs and requirements of the engineering curriculum. Moreover, existing virtual laboratories address specific concepts or a group of concepts and are therefore not designed for undertaking a range of instructor-designed curriculum-based laboratory activities. In addition, they are mostly based on commercial software tools

- which involve licensing costs. The VEL does not involve licensing costs (at least until its back-end freeware application software, such as Spice Opus, cease to be freeware) and is therefore suitable for use in resources constrained engineering faculties.
2. The quantification of the impact of a virtual electronic laboratory, in terms of students' knowledge gain, for a given set of laboratory activities, at two different engineering faculties, provides appropriate evidence and support to justify the use of virtual laboratories in the educational context generally, and engineering laboratory education particularly. Literature fails to provide appropriate support for the use of virtual laboratories in the educational context, highlighting a gap in the virtual laboratory literature despite its large size. Evidence provided by other studies, for the use of virtual laboratories in the educational context, was based on the use of survey instruments to evaluate the usability, acceptability, and effectiveness of the virtual laboratories, and elicit students' perceptions of their learning experiences.
 3. The quantification of students' knowledge gain from undertaking concept-based laboratory activities in the virtual electronic laboratory provides evidence and support of the effectiveness of the use of carefully designed concept-based laboratory activities and an intuitive laboratory learning tool to enhance students' understanding of taught engineering concepts.
 4. The quantification and comparative analysis of the learning gains of a student cohort in an engineering faculty in a University in a developed country (UoP), and a cohort in an engineering faculty in a University in a developing country (UTON) indicated a greater learning improvement in the UTON student cohort. This provides evidence and support of the greater need for complements and/or alternatives to the traditional laboratory in engineering faculties challenged by resource constraints and/or large class sizes, which is especially the case for engineering faculties in developing countries.
 5. The quantification and comparative analysis of the knowledge gain of the student cohorts which indicated a greater improvement in the learning of the UTON student cohort also provides evidence of the need for more implementation efforts on low-cost virtual laboratories for the various other fields of engineering, in order to extend (to needy engineering faculties) and consolidate the gains of the widely advocated

student-centred learning paradigm, which is already adopted by most engineering faculties in Universities in developed countries.

6. A Bayesian Network-based assessment tool (the LAP model) has been proposed for the performance assessment of students' laboratory work in the virtual laboratory environment. Literature search failed to highlight assessment models for virtual laboratory environments generally, and virtual electronic laboratory environment particularly. The LAP model has the capacity to address the problems of time demand, bias, inconsistency, fairness, and increasing student numbers. This is in addition to facilitating that laboratory work assessment conforms to statements of criteria for engineering learning and assessment. The assessment perspective of the model is also in line with the principles of the prevailing educational paradigm which demands that students demonstrate their abilities/skills for laboratory work, and their knowledge and understanding of concepts addressed by laboratory activities. The effectiveness of the model as an assessment tool, within the context of the VEL, has been demonstrated. The model is anchored on psychological learning models and theories, assertions by cognitive psychologists, and other researchers, while domain expert knowledge and physical observational information are used to define the variables and semantics of the model.
7. The practicability of the performance assessment of students' laboratory work from their observed laboratory work behaviour, based on the concept of *virtual-observation*, has been demonstrated and proposed as an alternative or complement to the written evidence technique often extended from the traditional laboratory setting to the virtual laboratory environment, and the *physical observation* assessment technique that was proposed as an alternative to the written evidence technique, by researchers.
8. A new method of validating BN-based assessment models, in virtual learning environments, the *virtual-observation* technique, has been proposed and its practicability demonstrated. The validation method was applied in the validation of the BN-based assessment tool reported in this thesis. The validation method uses the assessment outcome of human assessors, based on the performance data derived from the virtually observed work behaviour of students, as the "gold standard", to validate the assessment outcome of the BN-based model, using the same set of performance data used by the human assessors, for the same assessment scenarios. The technique

has the capacity to provide adequate evidence needed to justify the validity and use of an assessment model in an educational setting. It satisfies the key criteria for construct validity and gives the research design credibility.

9. The work presented in this thesis has resulted in the following publications:

Articles

- i. **E. Chika**, "Building Better Engineers: Extending and Consolidating the Gains", *IEEE GOLDRush*, September, 2008, pp. 12 - 13

Conference Proceedings

- i. **I. E. Achumba**, D. Azzi, and R. Khusainov, "Bayesian Network Model for Students' Laboratory Work Performance Assessment: An empirical investigation of the optimal construction approach", Accepted for oral presentation at the *International Conference on Computational Intelligence (ICCI 2011)*, Paris, France, July 27 – 29, 2011.
- ii. **I. E. Achumba**, "The Virtual Electronic Laboratory (VEL): Description and Demonstration", *European Conference on Library of Labs (LILA 2011)*, Selwyn College, University of Cambridge, United Kingdom (UK), 11 - 12 April, 2011.
- iii. **I. E. Chika**, D. Azzi, A. Hewitt, & J. Stocker, "A Holistic Approach To Assessing Students' Laboratory Performance Using Bayesian Networks", *Proceedings of IEEE Workshop on Computational Intelligence in Virtual Environments (CIVE '09)*, Mar. 29 to April 2, 2009, Nashville, USA, pp. 26 -32.
- iv. **I. E. Chika**, D. Azzi, Barry P. Haynes, and Roy Williams, "Students Laboratory Work Performance Assessment", *Proceedings of the International Conference on Interactive Computer Aided Learning (ICL 2008)*, Villach, Austria, September 24 – 26, 2008.
- v. **I. E. Chika**, D. Azzi, J. Stocker, and B. P. Haynes, "Genuine Lab Experiences for Students in Resource Constrained Environments: The RealLab with Integrated Intelligent Assessment", *Proceedings of the International Conference on Interactive Mobile and Computer Aided Learning (IMCL)*, Amman, Jordan April 16 - 18, 2008.
- vi. D. Azzi, **I. E. Chika**, and B. P. Haynes, "Engineering Learning and Assessment: Current and Emerging Trends", *Proceedings for the National Engineering Conference*, Abuja, Nigeria, 3 – 7 December, 2007.
- vii. **I. E. Chika**, D. Azzi, and B. P. Haynes, "A Paradigm Shift in Engineering Education: Implications for Assessment, Instruction and Curriculum", *Proceedings of the International Conference for Adaptive Science and Technology (ICAST' 07)*, Accra, Ghana, 10 – 13 December, 2007.

Journal Papers

- i. D. Azzi, **I. E. Achumba**, V. L. Dunn, and G. A. Chukwudebe, "Intelligent Performance Assessment of Students' Laboratory Work in a Virtual Electronic Laboratory Environment", Paper submitted to *IEEE Transactions on Learning Technologies*, for review. Submitted on 09/Dec./2010.
- ii. **I. E. Achumba**, and D. Azzi, "A Virtual Electronic Laboratory for Genuine Practical Experiences: Description and Evaluation", Paper submitted to *IEEE Transactions on Learning Technologies*, for review. Revised and resubmitted on 09/Dec./2010.
- iii. **I. E. Achumba**, D. Azzi, and J. Stoker, "Low-Cost Virtual Laboratory Workbench for Electronic Engineering", *International Journal of Virtual and Personal Learning Environments (IJVPLE)*, vol. 1, no. 4, pp. 6 – 22, Oct.-Dec., 2010.

- iv. **I. E. Chika**, Djamel Azzi, James Stocker and Barry P. Haynes, "Genuine Lab Experiences for Students in Resource Constrained Environments: The Reallab with Integrated Intelligent Assessment", *IEEE Multidisciplinary Engineering Education Magazine*, vol. 3, no. 4, December, 2008.
- v. **I. E. Chika**, D. Azzi, J. Stocker, and . P. Haynes, "Genuine Lab Experiences For Students in Resource Constrained Environments: The Reallab with Integrated Intelligent Assessment", *International Journal of Online Engineering (IJOE)*, vol. 4, no 3, August 2008.

8.4 FURTHER WORK

The areas of further work include the following:

- Evaluation of the VEL based on pre-test/post-test control group experimental design such that:
 - the assignment of participants to experimental (treatment) and control groups are randomised so that all participants have equal chances of being in the experimental or control group. The experimental group is the group that will undertake a set of laboratory activities in the VEL environment, while the control is the group that will undertake the same set of laboratory activities in the traditional laboratory environment.
 - it can be assumed that the two groups are equivalent on all important dimensions (such as group size) and that there are no systematic differences between the two groups.
 - participants in the two groups are matched on most relevant characteristics that could possibly affect the outcome of the research.
 - both group experience the same conditions, with the exception of the experimental group, which receives the influence of the independent variable (the treatment) in addition to the shared conditions of the two groups.
 - relevant experimental conditions are controlled so that nothing but the intervention of the independent variable (the treatment) is assumed to produce the observed changes in the values of the dependent variable (knowledge gain).
- Extension of the capabilities and flexibility of the VEL by:
 - ✓ creating the facility to save and retrieve a constructed circuit
 - ✓ the inclusion of an oscilloscope
 - ✓ enabling the use of both active components and Integrated Circuit (IC) chips to build circuits.

- ✓ Create an Adaptive/Intelligent Tutoring System (ATS/ITS) version of the VEL which will incorporate personalized feedback mechanism and adaptive/tutoring component. This demands the incorporation of an instructor/laboratory assistant support system into the virtual laboratory environment by translating human laboratory guidance into an expert system that supervises laboratory undertakings by students.
- ✓ Incorporate graphical user interfaces for circuit design and analysis.
- Creation of a student modelling component of the LAP model such that the system can automatically learn a personalized assessment model for each student, based on their laboratory work behaviour, in order to predict a student's laboratory work performance difficulties. The outcome of the student model can be used to adapt laboratory activities and assessment questions to the student (using decision-theoretic problem selection), and will also impact on the design choices for the feedback mechanism which should not be intrusive or obstructive (e.g., interrupting students' work in order to provide feedback), nor preventive (e.g., preventing a student from asking for help). In this context, feedback will be personalised and more meaningful.
- Investigation of the potential of facial expression characterization and recognition for identifying and classifying engineering students' mood changes while undertaking laboratory activities in virtual laboratory environment, with a view to predict/diagnose students' areas of difficulties. This is in addition to the information that can be gleaned from the analysis of the student's behaviour log.

8.5 CONCLUSION

It is imperative that students achieve a deeper understanding of fundamental engineering concepts if they are to mature into the much coveted skilled real-life problem solvers needed in the society at large and specially in the industry. It may not be possible to achieve this deeper understanding without meaningful laboratory experiences because laboratory work lies at the heart of the relationship between theory, practice, and meaning making, hence, the important role accorded to laboratory work in the undergraduate engineering course. The increased use of the traditional engineering laboratory in the instructional process, to integrate theory and practice, has proven challenging, which catalyzed the adoption of virtual

laboratories as complements, or alternatives in extreme cases. However, the existing virtual laboratories are not designed for undertaking curriculum-based laboratory activities and are based mostly on relatively expensive commercial software tools that require licensing.

The VEL, a low-cost virtual laboratory, which can be used to realistically undertake a range of instructor-designed curriculum-based laboratory activities, has been presented. A description of its architecture and implementation related issues have been given, in addition to a detailed description of its evaluation processes and results. Students can do and re-do laboratory activities several times and the instructor has a veritable and flexible platform for integrating theory and practice in order to enhance learning. The instructor also, has the opportunity to synchronize laboratory activities with lectures as it is possible to achieve a one-to-one student-equipment ratio, without time constraints. Also, there is room for increase in the number of laboratory activities, without the distraction of equipment setup and teardown.

The VEL was evaluated on two separate occasions with different student cohorts, using carefully planned laboratory learning scenarios aimed at enhancing students' understanding of concepts taught in lectures. The use of the same set of laboratory activities to evaluate the VEL in two different engineering faculties reinforces the uniformity of fundamental engineering concepts and indicates that the manner in which the virtual laboratory can be integrated into laboratory learning is not faculty or unit specific and does not require adaptation.

Furthermore, the critical role of laboratory in undergraduate engineering course makes the ongoing challenge of performance assessment of students' laboratory work one of utmost importance, and more so in the virtual laboratory environments, considering its increased adoption in the undergraduate engineering laboratory education. Assessment drives learning and the performance assessment of students' laboratory work is challenging and demanding in the face of increased student numbers. Literature highlighted the conventional laboratory work assessment practices, in the traditional laboratory setting, based on marking students' laboratory logbooks and written reports, among engineering faculties. This assessment method was criticised as failing to address espoused aims, coupled with the inherent potential for bias, and inconsistency. The *physical-observation*-based assessment method that was proposed as an alternative to the written evidence technique is costly (time, space, human resource, logistics), and difficult to apply to large class sizes. Consequently, the *virtual-observation*-based assessment method, using a BN-based assessment tool (the LAP model),

for the performance assessment of students' laboratory work in the VEL environment has been proposed. The model has been described in details and its evaluation results presented.

The results have demonstrated the effectiveness of the LAP model as a performance assessment tool, within the context of the VEL; the practicability of performance assessment of students' laboratory work from their "observed" behaviour; and the practicability of the virtual-observation-based method for validation evaluation of BN-based assessment models. The LAP model and the virtual-observation-based assessment method could reduce the cost and burden of marking students' laboratory work, promote fair, consistent and timely assessment, especially for large class sizes. The model is intended to assess not only what a student knows and understands about the concept(s) addressed by a laboratory activity, but what the student is actually able to do, through the analysis of a log of discrete behaviours. Inference about a student's performance is made on the basis of evidence extracted from the student's behaviour log. The model, though designed within a VEL environment, can be adapted for use in other virtual laboratory environments.

The assessment of undergraduate engineering students' laboratory work is an ongoing challenge for engineering faculties and constitutes a significant problem in the virtual laboratory environment, which adoption in undergraduate engineering laboratory education has increased. The work presented in this thesis demonstrates that the application of Bayesian network artificial intelligence technique to the solution of the problem is feasible. However more work remains to be done. Just as the domains of cognitive psychology, psychology of learning, artificial intelligence, graph theory, probability theory, and human-computer interaction were combined to produce the result of the work presented in this thesis, future work will almost certainly require an even broader interdisciplinary approach.

The work presented in this thesis has built on existing literature on virtual laboratories, performance assessment in virtual learning environments, and students' laboratory work assessment. It is the belief and hope of the author that the work has contributed to: the field Bayesian Network (BN) -based modelling (focused on reasoning under uncertainty), particularly with respect to assessment model validation; and the growing field of intelligent performance assessment in the virtual learning environment.



REFERENCES

- [1] Budhu, M. (2002). Virtual Laboratories For Engineering Education. Paper presented at the International Conference on Engineering Education, August 18. Retrieved March 31, 2007, from the INEER website: <http://www.ineer.org/Events/ICEE2002/Proceedings/Papers/Index/O334-O337/O334.pdf>
- [2] Davies, C. (2008). *Learning and Teaching in Laboratories: An Engineering Subject Centre Guide*, Retrieved December 7, 2009, from The Higher Education Academy website: <http://www.engsc.ac.uk/teaching-guides/laboratories/>.
- [3] Carter, G. (1980). Assessment of Undergraduate Electrical Engineering Laboratory Studies. *IEEE Proceeding*, 127(7), 460-472.
- [4] Harb, S. & Batarseh, I. (2006). Development of Web-based circuit simulator Teaching Electrical Engineering Courses. Proceeding of the Interactive Mobile And Computer Aided Learning (IMCL) Conference, April 19-21, 2006, Amman Jordan.
- [5] Gilbert, H. D. (2003). Electrical Engineering 215 – A case Study in Assessment. ASEE/IEEE Frontiers in Education Conference, Nov. 5 – 8, Boulder, CO, T4B - 9.
- [6] Tobin, P. (2007). The Role PSpice in the Engineering Teaching Environment. International Conference on Engineering Education, Coimbra, Portugal. <http://icee2007.dei.uc.pt/proceedings/papers/614.pdf>
- [7] Wankat, P. C., & Oreovicz, F. S. (1993). *Teaching Engineering*, McGraw-Hill, New York. <https://engineering.purdue.edu/ChE/AboutUs/Publications/TeachingEng/Book.pdf>
- [8] Hartfield, J., Scott, D., & Szmyd, D. (1995). A freshman Electrical Engineering Course and Laboratory for all Engineering Majors. In *Proceeding of Frontiers in Education Conference*, Vol. 2. (4c2.1-4c2). Retrieved from <http://fie.engrng.pitt.edu/fie95/4c2/4c21/4c21.htm>.
- [9] Collacott, R. A., (1971). Laboratory Objectives, *Education and Training*. 13(3), pp. 79 – 80.
- [10] Feisel, D. L., & Rosa, D. (2005). The Role of the Laboratory in Undergraduate Engineering Education, *Journal of Engineering Education*, 94(1), 121 – 130.
- [11] QAA (Quality Assurance Agency), (2006a). *Subject Benchmark Statements: Engineering*. Retrieved February 3, 2008, from the QAA website: <http://www.qaa.ac.uk/academicinfrastructure/benchmark/statements/Engineering06.pdf>
- [12] ABET (Accreditation Board for Engineering and Technology), (2011). *Criteria for Accrediting Engineering Technology Programs*. Retrieved March 6, 2011, from the ABET website: <http://www.abet.org/Linked%20Documents-UPDATE/Program%20Docs/abet-tac-criteria-2011-2012.pdf>
- [13] O’Sullivan, J. J. (2008). Laboratory Teaching in Undergraduate Hydraulic Engineering: Addressing the Negative Sentiment. International Conference on Engineering Education, 27-31 July 2008, Budapest, Hungary. http://icee2008hungary.net/download/fullp/full_papers/full_paper172.pdf
- [14] Mannix, M. (2000). *The Virtues of Virtual Labs*. ASEE PRISM. Retrieved May 8, 2007, from <http://www.highbeam.com/doc/1P3-60667340.html>.
- [15] Ma, J., & Nickerson, J. V. (2006). Hands-on, Simulated and Remote Laboratories: A Comparative Literature Review. *ACM Computing Surveys*, 38(3), pp. 1 – 24.

- [16] Moure, M. J., Valdes, M. D., & Mandado, E., et al. (1999). Educational Application of Virtual Instruments based on Reconfigurable Logic. IEEE International Conference on Microelectronic Systems Education, 19 – 21 July.
- [17] Corter, J. E., Nickerson, J. V., Esche, S. K., Chassapis, C., et al. (2007). Constructing reality: A study of remote, hands-on, and simulated laboratories. *ACM Transactions on Computer-Human Interaction*, 14(2).
- [18] Korzenny, F. (1978). A theory of electronic propinquity: Mediated communications in organizations. *Communication Research*, 5(1), pp. 3-24.
- [19] Reilly, R. (2005). Guest Editorial on Web-based Instruction: Doing things Better and doing better things. *IEEE transaction on Education*, 48(4).
- [20] Gravier, C., Fayolle, J., Bayard, B., Ates, M. & Lardon, J. (2008). State of the Art About Remote Laboratories Paradigms: Foundations of Ongoing Mutations. *International Journal of Online Engineering*, 4(1), 19 – 25.
- [21] Hoyer, H., Jochheim, A., Rohrig, C., & Bischoff, A. (2004). A Multiuser Virtual-Reality Environment for a Tele-Operated Laboratory. *IEEE Transactions on Education*, 47(1), 121 - 126
- [22] Carnevali, G., & Buttazzo, G. (2003). A Virtual Laboratory Environment for Real-time Experiment. Proceedings of the 5th IFAC International Symposium on Intelligent Components and Instrument for Control Applications (SICICA 2003), Aveiro, Portugal, 9 – 11 July.
- [23] Auer, M. E. (2000). Virtual Lab versus Remote Lab. Retrieved on October 1, 2008, from http://www.online-lab.net/rel/documents/auer_icde.pdf.
- [24] Auer, M. E., Dorminger, C., Kleber, H., Oswald, K., & Seper, W. (2007). Online Lab Austria: Results and Experiences. International Conference on Computer Aided Learning (ICL), 26 – 28 September, Villach, Austria.
- [25] Bell, J. T., & Fogler, H. S. (1999). Virtual Laboratory Accidents Designed to Increase Safety Awareness. Annual Meeting and Conference of the American Society for Engineering Education, Charlotte, NC.
- [26] Edward, N. S. (2002). The role of laboratory work in engineering education: student and staff perceptions. *International Journal of Electrical Engineering Education*, 39(1), 11-19.
- [27] Zywano, M. S., & Waalen, J. K. (2001). Student outcomes and attitudes in Technology-enabled and Traditional Education: a case study. *Global Journal of Engineering Education*, 5(1), 49-56.
- [28] Colman, J. N., Kinniment, D. J., & Burns, F. P., et al. (1998). Effectiveness of Computer- Aided Learning as a Direct Replacement for Lecturing in Degree-Level Electronics. *IEEE Transactions on Education*, 41(3), 177 – 184.
- [29] Striegel, A. (2001). Distance education and its impact on computer engineering education. 31st ASEE/IEEE Frontier in Education Conference, 10 - 13 October, Reno, NV, USA.
- [30] Campbell, J. O., Bourne, J., Mosterman, P. & Brodersen, A. J. (2002). Effectiveness of Learning Simulations For Electronic Laboratories. *Journal of Engineering Education*, 91(1), 81-87.
- [31] Hughes, I. (2001). Alternatives to Laboratory Practicals - Do They Meet the Needs?. *Innovations in Education and Teaching International*, 38(1), 3 - 7.
- [32] Madhavan, K., Schroeder, J., & Xian, H. (2009). Patterns of technology adoption and perceptions of virtual laboratories among undergraduate engineering students. Research in Engineering Education Symposium, July 20-23, 2009, Palm Cove, Queensland, Australia. Retrieved on January 29, 2008 from www.ci4ene.org/wp-content/uploads/2010/.../rees2009_submission_77.pdf

- [33] Crosier, J. K., Cobb, S. V., & Wilson, J. R. (2000). Experimental Comparison of Virtual Reality with Traditional Teaching Methods for Teaching Radioactivity. *Education and Information Technologies*, 5(4), 329-343.
- [34] Oakley, B. (1996). A virtual classroom approach to teaching circuit analysis. *IEEE Transactions on Education*, 39(3), 287 – 296.
- [35] Benesova, A., Boland, S., & Galloway, W. (2002). Virtual learning environments. Retrieved July 21, 2008 from http://www.dcs.napier.ac.uk/~mm/socbytes/feb2002_i/3.html
- [36] Balamuralithara, B., & Woods, P. C. (2008). Virtual Laboratories in Engineering Education: The Simulation Lab and Remote Lab. *Computer Applications in Engineering Education*, 17(1), 108 – 118.
- [37] Ertugrul, N. (2000). Towards Virtual Laboratories: a survey of LabVIEW-based Teaching/Learning Tools and Future Trends. *International Journal of Engineering Education*, 16(3), 171-180.
- [38] Medin, M. I., Ross, B. H., & Markman, A. B. (2002). *Cognitive Psychology* (3rd Ed.). New York: John Wiley & sons Inc.
- [39] Felder, R., & Brent, R. (2003). Designing and Teaching Courses to Satisfy the ABET Engineering Criteria. *Journal of Engineering Education*, 92(1), 7 - 25.
- [40] QAA (Quality Assurance Agency), (2006b). *Code of Practice Section 6: Assessment of Students*. Retrieved February 3, 2008, from the QAA website: <http://www.qaa.ac.uk/academicinfrastructure/codeOfPractice/section6/default.asp>
- [41] Kizlik, B., (2009, September). *Measurement, Assessment, and Evaluation in Education*. Retrieved October 5, 2008, from <http://www.adprima.com/measurement.htm>.
- [42] Rowntree, D. (1999). *Designing an assessment system*. Retrieved February 23, 2010, from http://www.sfubusiness.ca/files/PDF/teaching_facilitation/designing_an_assessment_system.pdf
- [43] Booth, R., B. Clayton, B., Hartcher, R., Hungar, S., Hyde, P. & Wilson, P. (2003). *The development of quality online assessment in vocational education and training*. In R. Byrnes, A. and Ellis, (2006). The prevalence and Characteristic of online assessment in Australian universities. *Australian Journal of Education Technology*, 22(1), 104 – 125.
- [44] Byrnes, R. & and Ellis, A. (2006). The prevalence and Characteristic of online assessment in Australian universities. *Australian Journal of Education Technology*, 22(1), 104 – 125.
- [45] Cohen, L., Manion, L., & Manion, K. (2007). *Research Methods in Education*, New York: Routledge.
- [46] Denscombe, M. (2002). *Ground Rules for Good Research*. Maidenhead: Open University Press.
- [47] Palop, J.M.G., & Teruel, J.M.A. (2000). Virtual work bench for electronic instrumentation teaching. *IEEE Transactions on Education*, 43(1), 15–18.
- [48] de Magistris, M. (2005). A MATLAB-based virtual laboratory for teaching introductory quasi-stationary electromagnetics. *IEEE Transactions on Education*, 48(1), 81-88.
- [49] Vanfretti, L. (2006). A comprehensive Virtual Laboratory for Teaching Power Systems Dynamics and Control at Undergraduate Level. *IEEE Power Systems Conference and Exposition*, Atlanta, Georgia, http://www.vanfretti.com/Dr_Luigi_Vanfrettis_Website/Publications.html
- [50] Sevgi, L. & Uluisik, C. (2006). A LabView-Based Virtual Instrument for Engineering Education: A Numerical Fourier Transform Tool. *Turk Journal of Electrical Engineering*, 14(1).
- [51] Chen, X., & Han, Y. (2008). Design and Implementation of Digital Signal Processing Soft Laboratory. *Proceedings of the International Conference on Computer Science and Software Engineering*. 5, 260–262.

- [52] Vasko, C., Novotny, M., & Szendiuch, I. (2007). Virtual laboratory of Microelectronic Technologies. *Proceedings of the 30th International Spring Seminar on Electronics Technology, Cluj-Napoca, Romania*, 564–566.
- [53] Hassan, R., Omari, N., Arshad, H., Sahran, S., & Yusof, N. H. (2010). DigiLab: A Virtual Lab for IT Students. *WSEAS Transactions on Advances in Engineering Education*, 7(5), 151 – 160.
- [54] Zhou, G.T., & Lo, H. (1999). Developing Java-based virtual laboratory Tools for an Undergraduate Random Signals and Noise Course. *Proceedings of the American Society for Engineering Education Annual Conference*, Charlotte, NC, 1 - 8. Retrieved on March 22, 2007 from <http://www.succeed.ufl.edu/papers/99/00838.pdf>.
- [55] Duarte, B.P. Butz, S.M. Miller, & Mahalingam, A. (2008). An Intelligent Universal Virtual Laboratory (UVL). *IEEE Transactions on Education*, 51(1), 2-9.
- [56] Iz, O. S. B. (2009, March). *iMedia Centre: Bringing Learning and Technology Together* (The iMedia Centre (iMC) is a media production outfit within Ngee Ann Polytechnic, Singapore). Retrieved May 27, 2010 from <http://www.learnerstogether.net/imedia-centre-bringing-learning-and-technology-together/288>
- [57] Fedák V. & Bauer, P. (2005). E-Learning in Education of Electrical Drives and Power Electronics: Opportunities and Challenges. *Proceedings of the 15th International Conference on Electrical Drives and Power Electronics*, 26-28 September, Dubrovnik, Croatia, 1 – 9 (paper E05-L4.pdf). Retrieved November 20, 2010 from <http://www.esa.fer.hr/edpe2005/technical/lectures/E05-IL4.pdf>
- [58] Van Poppel, B. P., Benson, M. J., Boettner, D. D. & Ozer, A. (2004). Virtual Laboratory Development for Undergraduate Engineering Course. *Proceedings of the Fifth International Conference on Information Technology Based Higher Education and Training*, 644 - 649.
- [59] Wilkins, B. R. (1973). The design of a first-year laboratory course. *Proceedings of the Conference on The teaching of electronic engineering in degree courses*, University of Hull. In Carter, G., (1980). Assessment of Undergraduate Electrical Engineering Laboratory Studies. *IEEE Proceedings*, 127(7), 460-472.
- [60] McPhun, M. K. (1973). A Programmed laboratory course in electronics for first year students. *Proceedings of the Conference on The teaching of electronic engineering in degree courses*, University of Hull, 1973. In Carter, G., (1980). Assessment of Undergraduate Electrical Engineering Laboratory Studies. *IEEE Proceedings*, 127(7), 460-472.
- [61] SEEDCU (School of Electronic Engineering, Dublin City University). Undergraduate Laboratory Handbook: The Laboratory Logbook. Retrieved October 10, 2009 from <http://www.eeng.dcu.ie/EE/253-EE.html> .
- [62] Pramanik, A. & Ring, D. (1977). An evolving teaching laboratory for first year students of electrical and electronic engineering. In Carter, G., (1980). Assessment of Undergraduate Electrical Engineering Laboratory Studies. *IEEE Proceedings*, 127(7), 460-472.
- [63] Dearden, G. J. (1976). Laboratory assessment: some generalised findings. *Proceeding of the Conference on the teaching of electronic engineering in degree courses*, University of Hull, 12/1-12/9. In Carter, G., (1980). Assessment of Undergraduate Electrical Engineering Laboratory Studies. *IEEE Proceedings*, 127(7), 460-472.
- [64] Jordana, J., Sánchez, F. J. (2010). Cooperative work and continuous assessment in an Electronic Systems laboratory course in a Telecommunication Engineering degree. *Proceedings of the 2010 IEEE Conference on Engineering Education (EDUCON)*, 14-16 April 2010, Madrid, 395 – 400.
- [65] Denton, J. P. (2005). Work in progress - application and assessment of student formal laboratory reports using a team approach with an author and editor format. *Proceedings*

- of the 35th Annual Conference on Frontiers in Education, 19-22 Oct. 2005, Indianapolis, IN, T4F – 17.
- [66] Watai, L. L., Francis, S. A., & Brodersen, A. J. (2007). A Qualitative and Systematic Assessment Methodology for Course Outcomes from Formal Laboratory Work Products in Electrical Engineering. 37th ASEE/IEEE Frontiers in Education Conference, October 10 – 13, 2007, Milwaukee’.
- [67] Wankat, P. C., & Oreovicz, F. S. (1993). *Teaching Engineering*, New York: McGraw-Hill Retrieved April 12, 2009 from <https://engineering.purdue.edu/ChE/AboutUs/Publications/TeachingEng/Book.pdf>
- [68] Hughes, I. (2004, May). Coping Strategies for Staff Involved in Assessment of Laboratory Write-Ups, BEE-1, 3. Retrieved on April 12, 2009 from <http://bio.ltsn.ac.uk/journal/vol3/beej-3-4.htm>
- [69] NCREL (North Central Regional Educational Laboratory), (2011), Performance Assessment. Retrieved on January 20, 2011 from <http://www.ncrel.org/sdrs/areas/issues/methods/assment/as8lk5.htm>.
- [70] Chamot, A., & O'Malley, M. (1994). *The CALLA Handbook*. MA: Addison-Wesley Publishing Company. In Anonymous, (2000). World Languages Assessment tasks. Central Jersey World Languages Professional Development Institute. Retrieved on August 14, 2009 from <http://www.globalteachinglearning.com/cjwlpdi/assessment/assess.shtml>.
- [71] Linn, R. L., Baker, E., & Dunbar, S. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher*, 20(8), 15-21. In Olfos, R. & Zulantay, H., (2007). Reliability and Validity of Authentic Assessment in a Web Based Course. *Educational Technology & Society*, 10(4), 156-173.
- [72] Tedick, D. J. (ed) (1998). *Proficiency-Oriented Language Instruction and Assessment: A Curriculum Handbook for Teachers*. In anon. World Languages Assessment tasks. Central Jersey World Languages Professional Development Institute, 2000. <http://www.globalteachinglearning.com/cjwlpdi/assessment/assess.shtml>
- [73] Chen, H., Chiu, M., & Tsai, Y., et al, (1995). The Use of Performance-Based Assessment in an Integrated Chemistry Laboratory Program. Retrieved August 28, 2009, from <http://www.ntnu.edu.tw/acad/docmeet/95/a3/a302.doc>.
- [74] Ganiel, U., & Hofstein, A. (1981). Objective and Continuous Assessment of Student Performance in the Physics Laboratory. Presented at the Annual Meeting of the National Association for Research in Science Teaching (54th, Grossinger’s in the Catskills), Ellenville, NY.
- [75] Bahri, R. Z., & Trevelyan, J. P. (2009). An effective way to measure practical intelligence from a laboratory experience. Research in Engineering Education Symposium. Retrieved August 28, 2009, from http://rees2009.pbworks.com/f/rees2009_submission_74.pdf
- [76] Alinier, N., & Alinier, G. (2005, October). Design of an objective assessment tool to evaluate students’ basic electrical engineering skills (The Higher Education Academy-Engineering Subject Centre (THEA-ESC) Mini-project report). Retrieved September 21 2009, from <http://www.engsc.ac.uk/mini-projects/design-of-an-objective-assessment-tool-to-evaluate-students-basic-electrical-engineering>
- [77] Harden, R. M., & Gleeson, F. A. (1979). Assessment of clinical competence using an objective structured clinical examination (OSCE)”, *Medical Education*, 13(1), 39–54.
- [78] Tesfazgi, S. H. (2003, April). Survey on Behavioural Observation Methods in Virtual Environments. <http://graphics.tudelft.nl/~vrphobia/research-samson.pdf>
- [79] Arrington, R. (2010, March). Study Finds Virtual Observations Better Than Naked Eye in Examining Moon Phases. *UVAToday*, 9 March 2010. <http://www.virginia.edu/uvatoday/newsRelease.php?id=11215>

- [80] Cheung, Y. K., & Elkind, M. V. (2001). Stochastic approximation with virtual observations for dose-finding on discrete levels. *Biometrika*, 97(1), 109–121.
- [81] Gallimore, R., & Stigler, J. (2003, January - March). LESSONLAB: Evolving Teaching into a Profession. *TechKnowLogia*, 32 – 34. http://www.techknowlogia.org/TKL_Articles/PDF/462.pdf
- [82] Bennett, R. E. (1990). Toward Intelligent Assessment: An Integration of Constructed Response Testing, Artificial Intelligence, and Model-Based Measurement 9Report No. ETS-RR-90-5). Princeton N.J.: Educational Services Testing
- [83] Tan, C. S., & Tay, Y. H. (2010). A Preliminary Study of the Progress on Using Bayesian Networks for Educational Assessment. Proceedings of the 2010 IEEE Conference on Sustainable Utilization and Development in Engineering and Technology, 20 – 21 November 2010, Kuala Lumpur, Malaysia.
- [84] Moussavi, M., & McGinn, N., (2009). A Bayesian Network for School Performance. <http://baw-uai09.intel-research.net/papers/Moussavi-UAIpaper--final.pdf>
- [85] Garcia, P., Amandi, A., Schiaffino, S., & Campo, M. (2007). Evaluating Bayesian networks' precision for detecting students' learning styles. *Computers and Education*, 49, 794-808.
- [86] Felder, R., & Silverman, L. (1988). Learning and teaching styles. *Journal of Engineering Education*, 78(7), 674-681.
- [87] Xenos, M., (2004). Prediction and assessment of student behaviour in open and distance education in computers using Bayesian networks. *Computers and Education*, 43(4), 345-359.
- [88] Corbett, A., Trask, H., & Scarpinato, K., et al. (1998). A formative evaluation of the PACT Algebra II Tutor: Support for Simple Hierarchical Reasoning. In Goettle, B., Half, H., and Redfield, C., et al (Eds.) *Proceedings of the 4th International Conference on ITS*, Springer-Verlag, 374-383.
- [89] Mislevy, R. J., Almond, R. G., & Yan, D., et al. (2000). Bayes Nets in Educational Assessment: Where the numbers come from. *Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence*, 437–446.
- [90] Collins, V., Greer, J. E., & Huang, S. X. (1996). Adaptive assessment using granularity hierarchies and bayesian nets. *Lecture Notes in Computer Science*, 1086, 569-577.
- [91] Zhang, L., Zhuang, Y., Yuan, Z., & Zhan, G. (2007). Auto Diagnosing: An Intelligent Assessment System Based on Bayesian Networks. *37th ASEE/IEEE Conference on Frontiers in Education*, T1G-7 – T1G-10.
- [92] Liu, C. C., Chen, G. D., & Wang, C. Y., et al. (2002). Student Performance Assessment Using Bayesian Network and Web Portfolios. *Journal of Educational Computing Research*, 27(4), 437-469.
- [93] VanLehn , K. (2001). Olae: A Bayesian Performance Assessment for Complex Problem Solving. Conference on Measurement in Education, April 3, Seattle, WA. <http://www.pitt.edu/~vanlehn/distrib/NCMEOlae3.html> .
- [94] VanLehn , K., & Martin, J. (1995). Student Assessment Using Bayesian Nets. *International Journal of Human-Computer Studies*. 42, 575 – 591.
- [95] VanLehn , K., & Martin, J. (1998). Evaluation of an assessment system based on Bayesian student modelling. *International Journal of Artificial Intelligence and Education*, 8(2), 179–221.
- [96] Noguez, J., Sucar, E., & Ramos, F. (2007). A Probabilistic relational Student Model for Virtual Laboratories. 11th International Conference on User Modeling, 25 - 29 June, Corfu, Greece. http://www.cs.usyd.edu.au/~aied/vol3/vol3_noguez.pdf.
- [97] Jensen, F. V. (2001). *Bayesian Networks and Decision Graphs*, USA: Springer.
- [98] Collins, V., Greer, J. E., & Huang, S. X. (1996). Adaptive assessment using granularity hierarchies and Bayesian nets”, *Lecture Notes in Computer Science*, 1086, 569-577.

- [99] Neil, M., Fenton, N., Nielson, L. (2000). Building Large-Scale Bayesian Networks. *The Knowledge Engineering Review*, 15(3), 257 – 284.
- [100] Kjaerulff, U. B., & Madsen, A. L. (2008). *Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis*, New York, USA: Springer Science+Business Media, LLC.(9780387741000).
- [101] Nilsson, N. J. (1998). *Artificial Intelligence: A New Synthesis*, San Francisco, California: Morgan Kaufmann Inc.
- [102] Mayo, M. J. (2001). Bayesian Student Modelling and Decision-Theoretic Selection of Tutorial Actions in Intelligent Tutoring Systems. Unpublished Ph.D Thesis, University Of Canterbury, Canterbury.
- [103] Korb, K. B., & Nicholson, A. E. (2004). *Bayesian Artificial Intelligence*, London: Chapman and Hall/CRC.
- [104] Lauritzen, S. L., & Spiegelhalter, D. J. (1988). Local Computations with Probabilities on Graphical Structure and Their Application to Expert Systems. *Journal of Royal Statistic Society*, 50(2), 157-224.
- [105] Cowell, R. (1999a). Introduction to Inference in Bayesian Networks. In Jordan M. (Ed.) *Learning in Graphical Models*, MIT Press, 9-26.
- [106] Myllymaki, P. (2010). Advantages of Bayesian Networks in Data Mining and Knowledge Discovery. <http://www.bayesit.com/docs/advantages.html>
- [107] Laskey, K. B., & Mahoney, S. M. (1997). Network fragments: Representing knowledge for constructing probabilistic models. *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*. In Geiger, D., & Shenoy, P., (eds), San Francisco, CA: Morgan Kaufmann, 334 – 341.
- [108] Hilbert, D. M., & Redmiles, D. F. (2000). Extracting Usability Information from User Interface Events. *ACM Computing Surveys*, 32(4), 384–421.
- [109] Mostow, J. (2004). Some useful design tactics for mining ITS data. *Proceedings of the ITS2004 Workshop on Analyzing Student-Tutor Interaction Logs to Improve Educational Outcomes* (pp. 20 – 28).
- [110] Guzdial, M., Walton, C., Konemann, M., & Soloway, E. (1993). Characterizing process change using log file data, (Technical Report GIT-GVU-93-44). <http://smartech.gatech.edu/bitstream/1853/3653/1/93-44.pdf>
- [111] Muehlenbrock, M. (2005). Automatic Action Analysis in an Interactive Learning Environment. Proceedings of the Conference on Artificial Intelligence in Engineering Education (AIED). <http://www.muehlenbrock.de/publications/Muehlenbrock-WS%20AIED%202005.pdf>
- [112] Scheuer, O., Muehlenbrock, M., & Melis, E., (2007). Results from action analysis in an interactive learning environment. *Journal of Interactive Learning Research*, 18(2), 185–205.
- [113] Scandura, J. M. (2001). Structural (Cognitive Task) Analysis: An Integrated Approach to Software Design and Programming. *Journal of Structural Learning and Intelligent Systems*, 14(4), 433-458.
- [114] Clark, R. E., & Estes, F. (1996). Cognitive task analysis. *International Journal of Educational Research*, 25(5), 403-417.
- [115] Militello, L. G., & Klein, G. (1997). Cognitive Task Analysis: Bringing a Powerful Tool into Wide Use. *CSERIAC GATEWAY*, 8(4). In Baber, C., Borrás, C., Hone, G., MacLeod, I., et al., (2005). Cognitive Task Analysis: Current use and Practice in the UK Armed Forces and Elsewhere, HFIDTC/WP2.3.1/1, 1. Retrieved March 3, 2011, from <http://www.hfidtc.com/research/methods/methods-reports/phase-1/cta-review-armed-forces.pdf>.
- [116] Schraagen, J. M., Chipman, S. F., & Shalin, V. L. (Eds) (2000). *Cognitive Task Analysis*. London: Psychology Press.

- [117] Chipman, S. F., Schraagen, J. M., & V. L. Shalin (2000). Introduction to cognitive task analysis. In Schraagen, J. M., Chipman, S. F., & Shute, V. J. (Eds.) *Cognitive Task Analysis*. Mahwah, NJ: Lawrence Erlbaum Associates, 3-23.
- [118] Clark, R. E., Feldon, D. F., van Merriënboer, J. J. G., Yates, K., & Early, S., (2006, October 14). *Cognitive Task Analysis*, 577 – 593.
- [119] Ryder, J. M., Weiland, M. Z., Szczepkowski, M. A., & Zachary, W. W. (1998). Cognitive Engineering of a New Telephone Operator Workstation Using COGNET. *International Journal of Industrial Ergonomics*, 22.
- [120] Seamster, T. L., Redding, R. E., & Kaempf, G. L. (2000). A Skill-Based Cognitive Task Analysis Framework. In Chipman, S. F., Schraagen, J. M., & V. L. Shalin, (2000). Introduction to cognitive task analysis. In Schraagen, J. M., Chipman, S. F., & Shute, V. J. (Eds.), *Cognitive Task Analysis*. Mahwah, NJ: Lawrence Erlbaum Associates, 135 – 141.
- [121] Zachary, W., Ryder, J., and Zubritzky, M. (1989). *Information processing models of human computer interaction in naval air anti-submarine warfare* (Technical Report No. 891215.8704), CHI Systems, Inc., Spring House, PA.
- [122] Zachary, W.W., Zaklad, A.L., Hicinbothom, J. H., Ryder, J.M., & Purcell, J.A. (1993), COGNET representation of tactical decision-making in Anti-Air Warfare. In *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting*, Human Factors and Ergonomics Society, Santa Monica, CA, 1112-1116.
- [123] Gagne, R.M., & Fleishman, E.A. (1959). *Psychology and Human Performance: An Introduction to Psychology*, New York: Holt-Dryden and Company.
- [124] Dember, W. N. and Jenkins, J. J. (1970). *General Psychology: Modelling Behaviour and Experience*, Englewood Cliffs, New Jersey: Prentice-Hall Inc.
- [125] Elander, J. (2003). Student assessment from a psychological perspective. *Psychology Learning and Teaching*, 3(2), 114-121.
- [126] Hamza, K., Alhalabi, B., & Marcovitz, D.M. (2000). Remote Labs, *Proceedings of the International Conference on Society for Information Technology and Teacher Education, Association for the Advancement of Computing in Education (AACE/SITE 2000)*, (ERIC Document Reproduction Service No. ED444475). Retrieved May 25, 2007 from <http://www.cse.fau.edu/~bassem/CADET/AACE2000-SanDiego.pdf>
- [127] del Alamo, J. A., Chang, V., Hardison, J., Zych, D., & Hui, L. (2003). An Online Microelectronics Device Characterization Laboratory with a Circuit-like User Interface. *Proceedings of the International Conference on Engineering Education*, July 21. Retrieved May 29, 2007 from <http://www-mtl.mit.edu/~alamo/pdf/2003/RC-95%20delAlamo%20ICEE%202003%20paper.pdf>
- [128] Mavrikis, M., & Maciocia, A. (2003). Wallis: a Web-based ILE for Science and Engineering Students Studying Mathematics. *International Conference on Artificial Intelligence in Education*, Sydney, Australia. Retrieved May 28, 2007 from http://www.cs.usyd.au/~aied/Supp_procs.html.
- [129] Siegel, M. (2002). What to do when the teaching lab instruments are too sophisticated for the students. *Proceeding of the International Symposium on Virtual and Intelligent Measurement Systems (VIMS2002)*, Alaska, USA, 123-128.
- [130] Murthy, R. S., (2004, September 17). A Simplified Introduction to Circuit Simulation using SPICE OPUS. Retrieved October 12, 2007, from <http://www.spiceopus.si/download/tutorial-rsam.pdf>.
- [131] SOHP (Spice Opus Home Page) (2006, December 2). Spice Opus: Frees Analogue Circuit Simulation. <http://fides.fe.uni-lj.si/spice/download/download1.html>
- [132] GHP (Gnuplot Home Page). (1986). <http://www.gnuplot.info/>
- [133] XHP (XAMPP Home Page). (2007). XAMPP for Windows. <http://www.apachefriends.org/en/xampp-windows.html>

- [134] Dugard, P., & Todman, J. (1995). Analysis of Pre-test-Post-test Control Group Designs in Educational Research. *Journal of Educational Psychology*, 15(2), 181-198.
- [135] Harkness, S. J. (2005). Student Assessment: Pre-Test/Post-Test and the Accumulation of Knowledge Across Sequential Prerequisites. *Proceedings of the Annual Meeting of the American Political Science Association*, Marriott Wardman Park, Omni Shoreham, Washington Hilton, Washington, DC. Retrieved October 14, 2009, from http://www.allacademic.com/meta/p40270_index.html.
- [136] Nicol, D. (2007). E-assessment by Design: using Multiple-choice Tests to Good Effect. *Journal of Further and Higher Education*, 31(1), 53-64.
- [137] DiGironimo, N. (2007). A Concentration Analysis of Student Responses on the 1995 Version of the Force Concept Inventory. Retrieved February 6, 2009, from <http://www.docstoc.com/docs/2396428/A-Concentration-Analysis-of-Student-Responses>.
- [138] Wattanakasiwich, P. (2006). Investigation of student understanding of force and motion by using concentration analysis. *Proceedings of the 32nd Congress on Science and Technology of Thailand*, 10 - 12 October, Bangkok. Retrieved February 6, 2009, from
- [139] Obaidat, I., & Malkawi, E. (2009). The Grasp of Physics Concepts of Motion: Identifying Particular Patterns in Students' Thinking. *International Journal for the Scholarship of Teaching and Learning*, 3(1), <http://www.georgiasouthern.edu/ijstol>
- [140] Bao, L. & Redish, E. F. (2001). Concentration Analysis: A Quantitative Assessment of Student States. *American Journal of Physics*, 65(7), 45-53.
- [141] Jarvis, M. (2000). *Theoretical Approaches in Psychology*. London: Routledge, Taylor and Francis Group.
- [142] Ioannides, C. & Vosniadou, S. (2002). The changing meanings of force. *Cognitive Science Quarterly*, 2(1), 5-62.
- [143] Lindsay, E. Koretsky, M., Richardson, J. J., & Mahalinga, M. (2007). Work in Progress – How Real is Student Engagement in using Virtual Laboratories. *Proceedings of the 37th ASEE/IEEE Conference on Frontiers in Education*, F3G-13 – F3G-14.
- [144] Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika*, 29, 350 – 362.
- [145] Ng, P., Pinto, J., & Williams, S. K. (2008, January). The Effect of Learning Styles on Course Performance: A Quantile Regression Analysis. Working Paper Series—08-02. Retrieved on February 25, 2010 from <http://www.cba.nau.edu/pin-ng/working/LearningStyles.pdf>.
- [146] Lindsay, E., & Good, M. (2007). Learning Styles a Potential Predictor of Student Achievement in Remote and Virtual Laboratory Classes. *Proceedings of the AaeE Conference*, Melbourne, Australia.
- [147] Viola, S.R., Graf, S., Kinshuk, & Leo, T. (2006). Analysis of Felder-Silverman Index of Learning Styles by a Data-Driven Statistical Approach. *Proceedings of the 8th IEEE International Symposium on Multimedia (ISM'06)*, 959-964.
- [148] Hamada, M., (2008). An Integrated Virtual Environment for Active and Collaborative e-Learning in Theory of Computation. *IEEE Transactions on Learning Technologies*, 1(2), 2008.
- [149] Masters, J., & Madhyastha, T. (2005). Educational Applets for Active Learning in Properties of Electronic Materials. *IEEE Transactions on Education*, 48(1).
- [150] Rosati, P. (1998). The Learning Preferences of Engineering Students from Two Perspectives. *Proceedings of the 28th Annual Conference on Frontiers in Education (FIE '98)*, 29-32.
- [151] Dumas, J. (1999). *Usability Testing Methods: Subjective Measures, Part II - Measuring Attitudes and Opinions*”, American Institutes for Research. Retrieved January 26, 2011 from http://www.upassoc.org/html/1999_archive/usability_testing_methods.html
- [152] Tittle, C., & Hill, R. (1967). Attitude Measurement and Prediction of Behaviour: An Evaluation of Conditions and Measurement Techniques. *Sociometry*, 30. In Page-Bucci, H. (2003, February). *The value of Likert scales in measuring attitudes of online*

- learners. Retrieved January 26, 2011 from <http://www.hkadesigns.co.uk/websites/msc/reme/likert.htm>
- [153] Neuman, W.L. (2000). *Social Research Methods: Qualitative and Quantitative Approaches*. USA: Allyn & Bacon. In Page-Bucci, H. (2003, February). *The value of Likert scales in measuring attitudes of online learners*. Retrieved January 26, 2011 from <http://www.hkadesigns.co.uk/websites/msc/reme/likert.htm>
- [154] Smith, M. K., (2003). *Learning theory*, The Encyclopaedia of Informal Education, Retrieved May 16, 2008 from www.infed.org/biblio/b-learn.htm.
- [155] Lave, J., & Wenger, E. (199). *Communities of Practice. Learning, meaning and identity*, Cambridge: Cambridge University Press.
- [156] Race, P. (1998). Five Factors for Successful Learning. Learning Technology Dissemination Initiative (LTDI) Value Conference, May 13, 1998, Retrieved July 17, 2008 from <http://www.icbl.hw.ac.uk/ltidi/workshop/valuetalks/race/sld022.htm>
- [157] Marzano, R. J., Pickering, D., & McTighe, J. (1993). *Assessing Student Outcomes: Performance Assessment Using the Dimensions of Learning Model*. Association for Supervision and Curriculum Development Conference, Virginia, USA. Retrieved May 20, 2008 from http://eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/19/c8/33.pdf,
- [158] Honderich, T. (1995). *The Oxford Companion to Philosophy*, Oxford University Press, Oxford,
- [159] Wikipedia. Assessment: Educational Assessment. 2008, http://en.wikipedia.org/wiki/Educational_assessment
- [160] Saljo, R. (2004). Learning about learning. *Higher Education*, 8, 443-451.
- [161] Harvey, L. (2004). Analytic Quality Glossary, Quality Research International. <http://www.qualityresearchinternational.com/glossary/>.
- [162] EC (Engineering Council) UK-SPEC (2004). The accreditation of Higher Education Programmes (UK Standard for Professional Engineering Competence). Retrieved on September, from the EC^{UK} website: <http://www.engc.org.uk/professional-qualifications/standards/uk-spec>
- [163] Norman, G. R., Newble, D. (2002). *International Handbook of Research in Medical Education*, Springer.
- [164] Schwartz, B., & Robbins, S. J. (1995). *Psychology of Learning and behaviour* (4th Ed.), New York, USA: W. W. Norton and Company Inc.
- [165] Fenton, N., Neil, M., & Caballero, J. G. (2006). Using Ranked Nodes to Model Qualitative Judgements in Bayesian Networks. *IEEE Transactions on Knowledge and Data Engineering*, 19(10), 1420–1432.
- [166] Mohler, J. L. (2001). Using Interactive Multimedia Technologies to Improve Student Understanding of Spatially-Dependent Engineering Concepts. *GraphiCon'2001*, 2001, Nizhny Novgorod, Russia. <http://www.tech.purdue.edu/cg/>
- [167] Pashler, H., McDaniell, M., Rohrer, D., & Bjork, R. (2008). Learning Styles: Concepts and Evidence. *Psychological Science in the Public Interest*, 9(3), 105 – 119.
- [168] Marrison, D. L., & Frick, M. J. (1994). The Effect of Agricultural Students' Learning Styles on Academic Achievement and their Perceptions of Two Methods of Instruction. *Journal of Agricultural Education*, 35(1). Retrieved on February 25, from <http://202.198.141.77/upload/soft/001/35-01-26.pdf>.
- [169] Zwyno, M. S. (2003). A contribution to validation of score meaning for Felder-Soloman's index of learning styles. Proceedings of the American Society for Engineering Education Conference and Exposition, 2003.
- [170] Litzinger, T. A., Lee, S. A., Wise, J. C., & Felder, R. M. (2005). A study of the reliability and validity of the felder-soloman index of learning styles. *Proceedings of the American Society for Engineering Education Annual Conference and Exposition*.

- [171]Felder, R. M., & Spurlin, J. (2005). Applications, reliability, and validity of the index of learning styles. *International Journal of Engineering Education*, 21(1), 103-112.
- [172]Stathacopoulou, R., Magoulas, G.D., & Grigoriadou, M. (1999). Neural Network-based Fuzzy Modeling of the Student in Intelligent Tutoring Systems. *Proceedings of the International Joint Conference on Neural Networks*, Washington DC, 5, 3517–3521.
- [173]Lansford, C. E. (1999). Using Pre-test/Post-test Data to Evaluate the Effectiveness of Computer Aided Instruction: A Study of CAI and Its Use with Developmental Reading Students. *Proceedings of the 10th International Conference on Society for Information Technology and Teacher Education*, San Antonio, TX, February 28 –March 4, 1-7.
- [174]Cowell, R. G. (1999b). Parameter Learning from Incomplete Data for Bayesian Networks. Retrieved from <http://www.staff.city.ac.uk/~rgc/webpages/aistats99.pdf>
- [175]Horvitz, E., Breeze, J., & Heckerman, D. (1998). The Lumiere Project: Bayesian User Modeling for Inferring the Goals and Needs of Software Users. *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence* July, 256-265. Retrieved from <http://research.microsoft.com/en-us/um/people/horvitz/lumiere.HTM>
- [176]Pollino, C. A., Woodberry, O., Nicholson, A., Korb, K., & Hart, B. T. (2007). Parameterisation and evaluation of a Bayesian network for use in an ecological risk assessment. *Elsevier: Journal of Environmental Modelling & Software*, 22, 1140-1152.
- [177]Brosnan, A.J. (2006). Use of Bayesian Belief Networks for Enemy Course of Action Assessment At the Tactical Level. *Journal of Battlefield Technology*, 9(1), 31–39.
- [178]Rajabally, R., Sen, P., Whittle, S., & Dalton, J. (2004). Aids to Bayesian Belief Network Construction. *Proceedings of the 2nd International IEEE Conference on Intelligent Systems*, 2, 457 – 461.
- [179]Das, B. (2008, August 4). *Generating Conditional Probabilities for Bayesian Networks: Easing the Knowledge Acquisition Problem*. Retrieved on February 24, 2010, from <http://arxiv.org/ftp/cs/papers/0411/0411034.pdf>.
- [180]Druzdzel, M. J., & van der Gaag, L. C. (2000). Building Probabilistic Networks: Where Do the Numbers Come from? (Guest Editors' Introduction). *IEEE Transactions on Knowledge and Data Engineering*, 12(4), 1-485.
- [181]Almond, R. G. (2007). Modelling Diagnostic Assessments with Bayesian Networks. *Journal of Educational Measurement*, 44(4), 341 – 359.
- [182]Cooper, G. F., & Herskovits, E. (1992). A Bayesian Method for the Induction of Probabilistic Networks from Data. *Machine Learning*, 9(4), 309–347.
- [183]Friedman, N. (1998). The Bayesian Structural EM Algorithm. *Proceedings of the 14th Annual Conference on Uncertainty in Artificial Intelligence (UAI-98)* (pp. 129 – 138), Morgan Kaufmann.
- [184]Pan, Y., & Burnside, E. S. (2004). The Effects of Training Parameters on Learning a Probabilistic Expert System for Mammography. *International Congress Series*, 1268, 1027–1032.
- [185]de Jongh, M., & Dru'zd'zel, M. J. (2009). A Comparison of Structural Distance Measures for Causal Bayesian Network Models. *Recent Advances in Intelligent Information Systems*, pp. 443–456. Retrieved September 07, 2010, from <http://iis.ipipan.waw.pl/2009/proceedings/iis09-43.pdf>.
- [186]ZADEH, L. A., (1958). What is Optimal?. *IRE Transactions on Information Theory*, pp. 3. Retrieved September 08, 2010, from <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=01057441>.
- [187]Druzdzel, M. J., & Simon, H. A.,(2006). Causality in Bayesian Belief Networks. *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, 3-11.
- [188]Fenton, N., & Neil, M.,(2006). Comment: Expert Elicitation for Reliable System Design. *Statistical Science*, 21(4), 451–453.
- [189]Ng, K., & Abramson, B.,(1992). Consensus Diagnosis: A Simulation Study. *IEEE Transactions on Systems, Man, and Cybernetics*, 22(5), 916–928.

- [190] Druzdzel, M. J., & Onisko, A., (2008). Are Bayesian Networks Sensitive to Precision of Their Parameters?. *Intelligent Information Systems*, 35–44. Retrieved on March 10, 2011, from <ftp://ftp.pitt.edu/users/d/r/druzdzel/iis08.pdf>
- [191] Heinrich, G., (2008, August). Parameter estimation for text analysis. Technical Note Version 2.4, vsonix GmbH and University of Leipzig. Retrieved on July 19, 2010 from <http://www.arbylon.net/publications/text-est.pdf>.
- [192] Oteniya, L. (2008). Bayesian belief networks for dementia diagnosis and other applications: a comparison of hand-crafting and construction using a novel data driven technique, Unpublished PhD Thesis, Department of Computing Science, University of Stirling, Stirling, Scotland.
- [193] Spirtes, P., Glymour, C., & Scheines, R. (1991). An algorithm for fast recovery of sparse causal graphs. *Social Science Computer Review*, 9, 62-72. Retrieved on July 22, 2010, from http://www.hss.cmu.edu/philosophy/techreports/15_Spirtes.pdf.
- [194] Sahin, F., Yavuz, M. Ç., Arnavut, Z., & Uluyol, Ö. (2007). Fault diagnosis for airplane engines using Bayesian networks and distributed particle swarm optimization, *Parallel Computing*, 33(2), 124-143.
- [195] Chickering, D. M. Geiger, D., & Heckerman, D., (1995). Learning Bayesian Networks: Search Methods and Experimental Results. Retrieved from <http://research.microsoft.com/en-us/um/people/dmax/publications/aistats95.pdf>
- [196] Heckerman, D. (1996). A Tutorial on Learning With Bayesian Networks (Technical Report MSR-TR-95-06). USA: Microsoft Research. Retrieved from <http://research.microsoft.com/pubs/69588/tr-95-06.pdf>
- [197] Lam, W., & Bacchus, F. (1994). Learning Bayesian belief networks: an approach based on the MDL principle. *Computational Intelligence*, 10, 269–293.
- [198] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723.
- [199] Schwarz, G. (1978). Estimation the dimension of a model. *Annals of Statistics* vol. 6, 1978, pp. 462–464.
- [200] Heckerman, D., Geiger, D. and Chickering, D. M. (1995). Learning Bayesian Networks: The Combination of Knowledge and Statistical Data. *Machine Learning*. 20(3), 197–243. Retrieved from <http://research.microsoft.com/en-us/um/people/heckerman/HGC94uai.pdf>
- [201] Buntine, W. (1991). Theory refinement on Bayesian networks. *Proceedings of the 7th Annual Conference on Uncertainty in Artificial Intelligence*, Los Angeles, CA, 52–60.
- [202] Yang, S. and Chang, K. (2002). Comparison of Score Metrics for Bayesian Network learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, 32(3), 419-428. Retrieved on July 22, 2010 from http://volgenau.gmu.edu/~kchang/publications/journal_pdf/com_score_matrices.pdf.
- [203] Cooper, G. F. (1993). A Bayesian Method for Learning Belief Networks that Contain Hidden Variables. AAAI (Technical Report WS-93-02). Retrieved from <http://www.aaai.org/Papers/Workshops/1993/WS-93-02/WS93-02-011.pdf>
- [204] Forster, M., & Sober, E. (2010). AIC Scores as Evidence – a Bayesian Interpretation, Retrieved from <http://philosophy.wisc.edu/sober/forster%20and%20sober%20AIC%20Scores%20as%20Evidence%20jan%2028%202010.pdf>
- [205] Chickering, D. M., Geiger, D., and Heckerman, D. (1994). Learning Bayesian networks is np-hard (Technical Report MSR-TR-94-17). Microsoft Research, Retrieved on July 22, 2010 from <http://research.microsoft.com/apps/pubs/default.aspx?id=69598>.
- [206] Robinson, R. W. (1973). Counting labelled acyclic digraphs. In Harary, F. (Ed.), *New Directions in the Theory of Graphs*, New York: Academic Press, 239-273.
- [207] Anderson, S., Madigan, D., and Perlman, M. (1997). A characterization of markov equivalence classes for acyclic digraphs. *Annals of Statistics*, 25, 505 -541.

- [208] Druzdzal, M. J. (1999, November). *GeNIe: A development environment for graphical decision-analytic models*. In *Proceedings of the 1999 Annual Symposium of the American Medical Informatics Association (AMIA-1999)*, page 1206, Washington, D.C., Retrieved from <http://www.pitt.edu/~druzdzal/psfiles/amia99.pdf>
- [209] NSC (Norsys Software Corp), (2008). Netica-J Reference Manual (Version 3.25). http://www.norsys.com/netica-j/docs/NeticaJ_Man.pdf
- [210] ARG (Automated Reasoning Group) (2004). SamIam. <http://reasoning.cs.ucla.edu/samiam/>
- [211] Chickering, D. M. (2002, October). The WinMine Toolkit” (Technical Report: MSR-TR-2002-103). Retrieved from <http://research.microsoft.com/en-us/um/people/dmax/WinMine/WinMine.pdf>
- [212] Ramoni, M., & Sebastiani, P. (1997). Learning Bayesian networks from incomplete databases (Technical Report KMi-TR-43). *Intelligent Data Analysis Journal*, 2(1).
- [213] Pearl, J. and Russell, S. (2000, November). *Bayesian Networks*. Retrieved from <http://www.cs.berkeley.edu/~russell/papers/hbttbn-bn.ps> Accessed 14/02/2010
- [214] MR (Microsoft Research), WinMine Toolkit Tutorial, Retrieved on February 02, 2011 from <http://research.microsoft.com/en-us/um/people/dmax/WinMine/Tutorial/Tutorial.html>.
- [215] BNTT (Bayesian Network Toolbox Tutorial). (2007, October). *How to use the Bayes Net Toolbox*. Retrieved on February 02, 2011 from <http://bnt.googlecode.com/svn/trunk/docs/usage.html>
- [216] Pennock, D. (2006, December). *Evaluating probabilistic predictions*. Retrieved from <http://blog.oddhead.com/2006/12/26/evaluating-probabilistic-predictions/>
- [217] Blattenberger, G., & Lad, F. (1985). Separating the Brier Score into Calibration and Refinement Components: A Graphical Exposition. *The American Statistician*, 39(1), 26-32. Retrieved from <http://www.jstor.org/stable/pdfplus/2683902.pdf>
- [218] Morgan, M., & Henrion, M. (1990). *Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis*. London: Cambridge University Press.
- [219] Doshi, P., Greenwald, L., & Clarke, J. (2002). Towards Effective Structure Learning for Large Bayesian Networks (AAAI Technical Report WS-02-14). Retrieved from <http://www.aaai.org/Papers/Workshops/2002/WS-02-14/WS02-14-003.pdf>
- [220] Good, I. J. (1952). Rational decisions. *Journal of the Royal Statistical Society*, 14, 107-114. In Roulston, M. The Logarithmic Scoring Rule a.k.a. “ignorance”. Retrieved from <http://www.cawcr.gov.au/bmrc/wefor/staff/eee/verif/Ignorance.html>
- [221] Jensen F. V. (2001). *Bayesian Networks and Decision Graphs*. USA: Springer.
- [222] Mahalingam, A., Butz, B. P., & Duarte, M. (2005). An intelligent circuit analysis module to analyze student queries in the universal virtual laboratory. *Proceedings of the 35th ASEE/IEEE Frontiers in Education Conference*, F4E-1–6.
- [223] Vago, I. (1985). *Graph Theory: Application to the Calculation of Electrical Networks*. Neth.: Elsevier Science Publishers and Akademiai Kiado.
- [224] Bouckaert, R. R. (2002). Low level information extraction: a Bayesian network based approach. *Proceedings of the 19th International Conference on Machine Learning: Workshop on Text Learning*,.
- [225] Chua L., & Lin, P. (1975) *Computer-Aided Circuit Analysis of Electric Circuits: Algorithms and Computational Techniques*. Prentice-Hall Professional Technical Reference.
- [226] Pedrycz, W., & Gomide, F. (1998). *An introduction to fuzzy sets: analysis and design*. USA: MIT Press.
- [227] Coupe, V. H. M., van der Gaag, L.C., & Habbema, J.D.F. (2000). Sensitivity analysis: an aid for belief-network quantification. *The Knowledge Engineering Review*, 15(3), 215-232.

- [228]Laskey K. B., & Mahoney, S. M., (2000). Network engineering for agile belief network Models. *IEEE Transactions on Knowledge and Data Engineering*, 12, 487–498.
- [229]Chan, H., & Darwiche, A. (2004). Sensitivity Analysis in Bayesian Networks: From Single to Multiple Parameters. *UAI*, 67–75.
- [230]Lanini, S. (2006). Water Management Impact Assessment Using a Bayesian Network. *7th International Conference on Hydro-informatics*, (HIC'06), Nice, France.
- [231]Pearl, J. (1998). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. California: Morgan Kaufmann.
- [232]Das, B. (1999). Representing Uncertainties Using Bayesian Networks (Report DSTO-TR-0918). Australia: Department of Defence.
- [233]Druzdzal, M. J., & Onisko, A. (2008). Are Bayesian Networks Sensitive to Precision of Their Parameters?. *Intelligent Information Systems*, 35–44.
- [234]Cozman, F. G. (2002, December). Bayesian Networks in Java User Manual. Received on March 22, 2011 from <http://www.pmr.poli.usp.br/ltd/Software/javabayes/Home/node8.html>.
- [235]Harlen, W. (2007). Criteria for Evaluating Systems for Student Assessment. *Studies in Educational Evaluation*, 33, 15–28.
- [236]Russel, S., & Norvig, P. (2003). *Artificial Intelligence: A Modern Approach*. USA: Pearson Educ. Inc.
- [237]Stathacopoulou, R., Grigoriadou, M., and Samarakou, M., & Mitropoulos, D. (2007, May). Monitoring Students' Actions and Using Teachers' Expertise in Implementing and Evaluating the Neural Network-Based Fuzzy Diagnostic Model. *Expert Systems with Applications*, 32(4), 955–975.
- [238]Daum, S. G. (1997). Using Simulated Students to Improve Residents' Teaching. *Academic Medicine*, 72(5).
- [239]Vanlehn, K., Ohlsson, S., and Nason, R. (1994). Applications of Simulated Students: An Exploration. *Journal of Artificial Intelligence in Education*, 5(2), 135–175.
- [240]Brown, J., & Eskenazi, M. (2006). Using Simulated Students for the Assessment of Authentic Document Retrieval. In Ikeda, M., Ashley, K., & Chan, T. W. (Eds) (2006). *Intelligent Tutoring Systems*, 685–688.
- [241]Runyon, R. P., Coleman, K. A., & Pittenger, D. J. (2000). *Fundamentals of Behavioural Statistic*. Boston USA: McGraw-Hill.
- [242]Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performance as scientific inquiry into scoring meaning, *American Psychologist*, 9, 741-749.
- [243]Cronbach, L. J. (1971). Test validation. In Thorndike, R. L. (Ed.), *Educational Measurement* (2nd Ed.). Washington, D. C.: American Council on Education.
- [244]Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newsbury Park: Sage Publications.



APPENDIX A: PRE-TEST AND POST-TEST SCORES

Table A1: UoP participants' pre-/post-test Scores and Percentage Gain

UoP Cohort Pre-Test and Post-Test Scores and Gain				
S/No	Participant ID	Pre-Test Score (%)	Post-Test Score (%)	% Gain
1	P05	55	82	33
2	P06	64	73	12
3	P07	91	91	0
4	P09	64	82	22
5	P10	64	82	22
6	P11	36	73	51
7	P12	55	45	-22
8	P13	64	82	22
9	P14	64	91	30
Mean (μ)		$\mu_{p1} = 61.89$	$\mu_{p2} = 77.89$	$M_{pg} = 18.89$
Median (Md)		$Md_{p1} = 63$	$Md_{p2} = 82$	$Md_{pg} = 22$
Spread		55	46	

Table A2: UTON participants' pre-/post-test Scores and Percentage Gain

S/No	Participant ID	Pre-Test Score (%)	Post-Test Score (%)	% Gain
1	p1	45	55	18
2	p2	36	18	-100
3	p3	36	55	35
4	p4	73	82	11
5	p5	45	55	18
6	p6	45	73	38
7	p7	64	73	12
8	p8	27	36	25
9	p9	55	45	-22
10	p10	45	64	30
11	p11	27	55	51
12	p12	18	45	60
13	p13	55	73	25
14	p14	55	82	33
15	p15	45	45	0
16	p16	27	55	51
17	p17	36	45	20
18	p18	18	36	50
19	p19	45	55	18
20	p20	55	55	0
21	p21	45	55	18
22	p22	45	73	38
23	p23	09	36	75
24	p24	45	64	30
25	p25	64	82	22
26	p26	27	36	25
27	p27	45	64	30
28	p28	45	36	-25
29	p29	36	36	0
30	p30	36	27	-33
31	p31	36	36	0
32	p32	45	45	0
33	p33	45	73	38
34	p34	55	73	25
35	p35	09	36	75
36	p36	36	64	44
37	p37	27	64	58
38	p38	36	73	51
39	p39	55	73	25
40	p40	18	45	60
41	p41	36	64	44
42	p42	45	45	0
43	p43	45	73	38
44	p44	27	64	58
45	p45	64	82	22
46	p46	27	55	51
47	p47	36	55	35
48	p48	64	45	-42
49	p49	64	82	22
50	p50	55	64	14
51	p51	18	45	60
52	p52	36	55	35
Mean Score		$\mu_{P1} = 40.75$	$\mu_{P2} = 56.02$	$M_{Pg} = 24$
Median (Md)		$Md_{P1} = 45$	$Md_{P2} = 55$	$Md_{Pg} = 25$
Spread		64	64	



APPENDIX B: CONCENTRATION ANALYSIS FORMULA, DETAILS, AND STUDENTS' RESPONSE DISTRIBUTIONS

APPENDIX B1: THE CONCENTRATION ANALYSIS FORMULA AND DETAILS

The CAM uses a synergy of the *Score (S)*, and the *Concentration Factor (C)*, both of which are functions of the students' responses, to derive the Students' Response Patterns (SRPs) for each item in a test. The *C* value for each test item is derived using equation B1.

$$c_j = \frac{\sqrt{m}}{\sqrt{m} - 1} \times \left(\frac{\sqrt{\sum_1^m n_i^2}}{N} - \frac{1}{\sqrt{m}} \right) \quad (\text{B1.1})$$

where $j = 1$ to k (k = no of test items), m = no of choices per item, n_i = no of selections for choice i , and N = total no of responses to item j . *C* maps the response of a class on a test item to the interval [0,1] with zero corresponding to students selecting a random distribution of answers and one corresponding to all students selecting the same answer. The *S* value for an item is the ratio of the number of correct responses to the total number of students that responded to the item. *S* also takes values in the range [0, 1]. Combining the codes for the score, *S*, and the concentration factor, *C*, gives the *SC* coding, the SRPs, for each of the items [Bao & Redish, 2001]. The SRP is based on the 3-level coding scheme of Table B1.1, where “L” stands for low, “M” for medium and “H” for high. A situation of low score ($S < 0.4$) but high concentration value ($C > 0.5$) is represented with an LH type of response pattern and often indicates that students are likely to have a very popular incorrect knowledge state fragment. In a situation of medium score ($0.4 \sim 0.7$) and medium concentration ($0.2 \sim 0.5$), referred as an MM type of response pattern, students are often in a mixed state between the correct and incorrect knowledge state fragment (two different dominant knowledge state fragments: one correct, one incorrect).

Table B1: Three-level coding scheme for Score and Concentration Factor
(Source: items [Bao & Redish, 2001])

SCORE RANGE AND LEVEL		CONCENTRATION FACTOR RANGE AND LEVEL	
Score (<i>S</i>)	Level	Concentration Factor (<i>C</i>)	Level
0 ~ 0.4	L	0 ~ 0.2	L
0.4 ~ 0.7	M	0.2 ~ 0.5	M
0.7 ~ 1.0	H	0.5 ~ 1.0	H

APPENDIX B2: STUDENTS' RESPONSE DISTRIBUTION FOR THE PRE-TEST POST-TEST ITEMS

Table B2.1: UoP Response Distribution for each Choice of the Pre-Test Items

UoP Response Distribution for each Choice of the Pre-test Items					
Item Number	A	B	C	D	Total Number of Answers
1	2	6	0	1	9
2	5	2	2	0	9
3	0	6	3	0	9
4	8	1	0	0	9
5	3	1	0	5	9
6	3	6	0	0	9
7	2	7	0	0	9
8	7	2	0	0	9
9	0	1	0	8	9
10	0	1	6	1	8
11	0	0	6	3	9

Table B2.2: UoP Response Distribution for each Choice of the Post-Test Items

UoP Response Distribution for each Choice of the Post-test Items					
Item Number	A	B	C	D	Total Number of Answers
1	2	6	0	1	9
2	8	0	1	0	9
3	0	8	1	0	9
4	8	1	0	0	9
5	2	1	0	6	9
6	3	6	0	0	9
7	0	9	0	0	9
8	8	1	0	0	9
9	0	0	0	9	9
10	0	4	4	0	8
11	0	0	9	0	9

Table B2.3: UTON Response Distribution for each Choice of the Pre-Test Items

UTON Response Distribution for each Choice of the Pre-test Items					
Item Number	A	B	C	D	Total Number of Answers
1	7	36	6	3	52
2	17	7	17	10	51
3	7	11	28	6	52
4	35	10	5	2	52
5	19	14	4	14	51
6	15	21	5	10	51
7	14	28	8	1	51
8	20	20	9	1	50
9	22	12	9	8	51
10	8	11	26	5	50
11	0	11	26	15	52

Table B2.4: UTON Response Distribution for each Choice of the Post-Test Items

UTON Response Distribution for each Choice of the Post-test Items					
Item Number	A	B	C	D	Total Number of Answers
1	4	45	1	2	52
2	16	25	7	4	52
3	5	26	15	6	52
4	42	10	0	0	52
5	15	7	6	24	52
6	9	39	3	1	52
7	3	43	5	1	52
8	39	4	8	1	52
9	13	6	9	23	51
10	2	25	22	3	52
11	3	3	37	9	52

APPENDIX B3: THE CONCENTRATION ANALYSIS RESULTS AND DISCUSSION

The results of the CAM analysis, in terms of the students' *S*, *C*, and *SRP* distribution for the pre-test items and their corresponding post-test items, are shown in Tables 3.4 and 3.5. The Questions (*Q*) are numbered 1 to 11 (*Q1*, *Q2*, ... ,*Q11*) for the purposes of the analysis result presentation.

Table B3.1: Score, Concentration factor, and SRPs for the UoP Cohort

Pre-Test (UoP)											
<i>Q</i>	<i>Q1</i>	<i>Q2</i>	<i>Q3</i>	<i>Q4</i>	<i>Q5</i>	<i>Q6</i>	<i>Q7</i>	<i>Q8</i>	<i>Q9</i>	<i>Q10</i>	<i>Q11</i>
<i>S</i>	0.67	0.56	0.67	0.89	0.56	0.33	0.78	0.78	0.89	0.11	0.67
<i>C</i>	0.42	0.28	0.49	0.79	0.31	0.49	0.62	0.62	0.79	0.54	0.49
<i>SRP</i>	MM	MM	MM	HH	MM	LM	HH	HH	HH	LH	MM
Post-Test (UoP)											
<i>Q</i>	<i>Q1</i>	<i>Q2</i>	<i>Q3</i>	<i>Q4</i>	<i>Q5</i>	<i>Q6</i>	<i>Q7</i>	<i>Q8</i>	<i>Q9</i>	<i>Q10</i>	<i>Q11</i>
<i>S</i>	0.67	0.89	0.89	0.89	0.67	0.33	1.0	0.89	1.0	0.44	1.0
<i>C</i>	0.42	0.79	0.79	0.79	0.42	0.49	1.0	0.79	1.0	0.41	1.0
<i>SRP</i>	MM	HH	HH	HH	MM	LM	HH	HH	HH	MM	HH

Table B3.1 shows the UoP student cohort's response analysis results, which highlights some interesting themes as itemized below:

1. are four HH in pre-test responses and 7 HH in post-test responses, an indication of knowledge convergence and improved learning in the UoP cohort after, undertaking the laboratory activities in the VEL.
2. is no LL in both the pre-test and post-test responses, implying the absence of guessing in the UoP student cohort's responses for both tests.
3. is one LH in the pre-test responses (for Q10) and no LH in the post-test responses, implying that the one incorrect knowledge state fragment some of the students had, in relation to item Q10, before undertaking the laboratory activities in the VEL, was corrected such that almost half of the affected students now have the correct knowledge state fragment, while the remaining half retained the incorrect knowledge state fragment, after undertaking the laboratory activities in the VEL environment.
4. is one LM in both the pre-test and post-tests for the same test item, Q6, which indicates that undertaking the laboratory activities in the VEL environment failed to correct the students incorrect knowledge state fragment for this item.
5. are five MM in pre-test responses (for Q1, Q2, Q3, Q5, Q11) and 2, out of these five MMs in pre-test responses still exist in the post-test responses (for Q1, Q5), while the other three (for Q2, Q3, Q11) had been replaced by HH. This means that the three items for which the cohort had two different dominant knowledge state fragment (one correct, one incorrect), after undertaking the laboratory activities in the VEL environment, their knowledge was corrected to one dominant correct knowledge state fragment (HH), which indicates improved learning for the student cohort.

Table B3.2: Score, Concentration factor, and SRPs for UTON Cohort

Pre-Test (UTON)											
<i>Q</i>	<i>Q1</i>	<i>Q2</i>	<i>Q3</i>	<i>Q4</i>	<i>Q5</i>	<i>Q6</i>	<i>Q7</i>	<i>Q8</i>	<i>Q9</i>	<i>Q10</i>	<i>Q11</i>
<i>S</i>	0.69	0.33	0.21	0.67	0.27	0.29	0.55	0.40	0.16	0.22	0.50
<i>C</i>	0.43	0.06	0.21	0.42	0.09	0.10	0.27	0.19	0.09	0.19	0.23
<i>SRP</i>	MM	LL	LM	MM	LL	LL	MM	LL	LL	LL	MM
Post-Test (UTON)											
<i>Q</i>	<i>Q1</i>	<i>Q2</i>	<i>Q3</i>	<i>Q4</i>	<i>Q5</i>	<i>Q6</i>	<i>Q7</i>	<i>Q8</i>	<i>Q9</i>	<i>Q10</i>	<i>Q11</i>
<i>S</i>	0.86	0.31	0.50	0.84	0.46	0.17	0.83	0.75	0.45	0.48	0.71
<i>C</i>	0.74	0.18	0.19	0.66	0.14	0.54	0.67	0.54	0.12	0.29	0.47
<i>SRP</i>	HH	LL	ML	HH	ML	LH	HH	HH	ML	MM	HM

Table B3.2 shows the UTON student cohort's response analysis results, which also highlights some interesting themes as itemized below:

1. no HH in the pre-test responses and 4 HH in the post-test responses for (Q1, Q4, Q7, Q8), where, before undertaking the laboratory activities in the VEL environment, the pre-test SRP indicated two different dominant CKS fragments (one correct, one incorrect) for the items, Q1, Q4, and Q7, and a guessing (LL) for Q8.
2. 6 LL in the pre-test responses and only 1 LL in the post-test responses, implying that the six items, Q2, Q5, Q6, Q8, Q9, and Q10, for which most students had no clue about the correct response, thereby resorting to guessing, has been reduced to only 1 LL (for item Q2), after perusing the VEL.
3. 4 MM in the pre-test responses (for Q1, Q4, Q7, Q11) and 3 of which were corrected (for Q1, Q4, Q7) to HH in the post-test responses. This means that the three of the four items for which the cohort had two different dominant CKS fragments (one correct, one incorrect), after undertaking the laboratory activities in the VEL environment, their knowledge was corrected to one dominant correct CKS fragment (HH), which indicates improved learning for the student cohort.



APPENDIX C: STUDENTS' SCORES ON THE LEARNING STYLE PREFERENCES SCALES AND USABILITY SURVEY ITEMS

Table C1: UoP students' scores on the LS preferences scales

S/No	Participant ID	Learning Style (LS) Points								LS
		Type of Information Readily Perceived		Input Source (Communication Channel)		Information Processing Method		Method of Progression Towards Understanding		
		<i>Sensing/Intuitive</i>		<i>Visual/Verbal</i>		<i>Active/Reflective</i>		<i>Sequential/Global</i>		
		<i>Sensing (Sen)</i>	<i>intuitive (I)</i>	<i>visual (V)</i>	<i>verbal (Ver)</i>	<i>active (A)</i>	<i>reflective (R)</i>	<i>sequential (Seq)</i>	<i>global (G)</i>	
1	P05	1		7		1			1	
2	P06		1	9			7	5		
3	P07	1		11			1	3		
4	P09		1	9		1		3		
5	P10		3	9			5		3	
6	P11	11		11		7		1		
7	P12	3		11		9		1		
8	P13	1		7		5			7	
9	P14	1		9		7			7	

Table C2: UTON students' scores on the LS preferences scales

S/No	Participant ID	Learning Style Points								LS
		Type of Information Readily Perceived		Input Source (Communication Channel)		Information Processing Method		Method of Progression Towards Understanding		
		Sensing/Intuitive		Visual/Verbal		Active/Reflective		Sequential/Global		
		Sensing (Sen)	intuitive (I)	visual (V)	verbal (Ver)	active (A)	reflective (R)	sequential (Seq)	global (G)	
1	p1	3		7		7		1		
2	p2		5	7			3		1	
3	p3		3	5			5		1	
4	p4		1		5	5		1		
5	p5		5	7			7		3	
6	p6		7		5	1		7		
7	p7	3		1		1			11	
8	p8	1		5			1	5		
9	p9		3	7			3		1	
10	p10	5		1		3			3	
11	p11	5		5			1	7		
12	p12	5			5		1		1	
13	p13	5		3			1	3		
14	p14	1		5		3		5		
15	p15		5	3			1	3		
16	p16	5		3			1	1		
17	p17	5		7			1		5	
18	p18	9		5		7		7		
19	p19	3		7			3		3	
20	p20		9	1			5		1	
21	p21	5			7		3	5		
22	p22	7			1		3	7		
23	p23		7	5			3		5	
24	p24	5			7		7	1		
25	p25	1		5			1	5		
26	p26		5	11		3		3		
27	p27	7		1			5	5		
28	p28		1	3		1		1		
29	p29		5	3			1	3		
30	p30	7		7		7		7		
31	p31	5		11		3			3	
32	p32	1		5		1		3		
33	p33		5		11	1		1		
34	p34	5		1			5	1		
35	p35		3	11		3			3	
36	p36	9		7			5	3		
37	p37	3		3		1		5		
38	p38	3		7		5			5	
39	p39	9		7		1			1	
40	p40		3	1		5			3	
41	p41	3		1			1		3	
42	p42		5		5	7		1		
43	p43		5	5			3		7	
44	p44	7			7	5		1		
45	p45		1	5		1		1		
46	p46	5		1		1			3	
47	p47	3		9			1	1		
48	p48		1	1		5		5		
49	p49		3	5			7	5		
50	p50		7	3		1			1	
51	p51		1		3	3			3	
52	p52	5		7		1			7	

Table C3: UoP Scores for some of the Usability Survey Items

UoP Scores of the Usability Survey Items							
S/No	Participant ID	Q1 Easy	Q2 Enhanced	Q3 Not Difficult	Q4 Enjoyed	Q5 Recommended	Q6 Excellent
1	P05	3	4	4	4	4	4
2	P06	5	3	4	5	5	4
3	P07	4	4	4	4	4	4
4	P09	4	4	3	4	4	4
5	P10	4	4	3	4	4	4
6	P11	4	4	4	4	4	4
7	P12	4	4	3	4	5	5
8	P13	3	4	2	4	4	4
9	P14	4	4	4	4	5	4
Mean		3.89	3.89	3.44	4.11	4.33	4.11
SD		0.60	0.33	0.73	0.33	0.50	0.33
Min		3	3	2	4	4	4
Max		5	4	4	5	5	5

Table C4: UTON Scores for some of the Usability Survey Items

UTON Student Cohorts' Scores of the Survey Items							
S/No	Participant ID	Q1 Easy	Q2 Enhanced	Q3 Not Difficult	Q4 Enjoyed	Q5 Recommend	Q6 Excellent
1	p1	5	5	5	5	5	5
2	p2	2	4	2	5	5	3
3	p3	4	5	4	5	4	3
4	p4	4	5	3	5	4	5
5	p5	3	5	4	5	5	4
6	p6	3	4	4	4	5	4
7	p7	4	4	3	4	4	4
8	p8	3	3	2	3	3	3
9	p9	3	5	5	5	5	5
10	p10	5	5	5	5	5	5
11	p11	4	4	4	3	5	5
12	p12	4	4	2	4	4	4
13	p13	4	5	3	4	4	4
14	p14	5	4	4	4	5	4
15	p15	4	3	3	4	5	5
17	p17	3	2	3	4	4	4
18	p18	3	5	5	5	5	5
19	p19	4	4	4	4	5	4
20	p20	4	5	1	5	5	5
21	p21	4	5	3	5	5	5
22	p22	4	5	3	5	5	5
23	p23	4	3	3	4	4	5
24	p24	4	4	4	3	5	4
25	p25	3	5	3	5	5	5
26	p26	3	4	3	4	4	4
27	p27	5	4	4	5	5	4
28	p28	3	5	4	5	5	5
29	p29	5	4	4	4	4	4
30	p30	5	5	2	4	4	4
31	p31	4	3	3	4	5	5
32	p32	5	2	3	5	5	5
34	p34	4	5	3	5	5	5
35	p35	5	4	2	4	5	4
36	p36	5	5	3	5	5	5
37	p37	4	5	2	5	5	4
38	p38	5	4	3	4	4	4
39	p39	4	5	5	5	5	5
40	p40	5	5	5	5	5	5
41	p41	4	4	4	4	5	4
42	p42	4	5	4	5	5	5
43	p43	5	4	3	5	5	5
44	p44	4	5	5	5	5	5
45	p45	3	5	3	5	5	5
46	p46	4	4	3	3	4	4
47	p47	5	5	3	5	5	5
48	p48	5	5	5	5	5	4
49	p49	4	4	3	5	5	5
50	p50	3	4	3	4	4	4
51	p51	4	5	4	4	5	5
52	p52	4	4	4	5	5	4
	Mean	4.02	4.34	3.44	4.48	4.70	4.46
	SD	0.77	0.80	0.97	0.65	0.51	0.61
	Min	2	2	1	3	3	3
	Max	5	5	5	5	5	5



APPENDIX D: DOMAIN EXPERTS' LINK RATINGS

Table D1: Domain Expert Link Weight Assignments

Expert Link Weight Assignments							
S/No	LINK		Expert 1 Weighting	Expert 2 Weighting	Expert 3 Weighting	Average Link Weight	Link Weight Used
	Parent node	Child node					
1	1 (UCC)	4 (WWC)	40	40	40	40.000	40
2	2 (CPC)	4 (WWC)	20	40	40	33.333	33
3	3 (UAC)	4 (WWC)	40	20	20	26.667	27
4	5 (OSP)	7 (PUE)	50	50	50	50.000	50
5	6 (OEC)	7 (PUE)	50	50	50	50.000	50
6	3 (UAC)	11 (ADA)	90	50	50	63.333	63
7	10 (MAC)	11 (ADA)	10	50	50	36.667	37
8	9 (UCV)	10 (MAC)	50	35	50	45.000	45
9	8 (ACF)	10 (MAC)	50	65	50	55.000	55
10	12 (SAC)	14 (ADC)	50	50	50	50.000	50
11	13 (GLC)	14 (ADC)	50	50	50	50.000	50
12	7 (PUE)	15 (CC)	20	20	30	23.333	23
13	14 (ADC)	15 (CC)	30	20	20	23.333	23
14	4 (WWC)	15 (CC)	30	40	30	33.333	34
15	11 (ADA)	15 (CC)	20	20	20	20.000	20
16	7 (PUE)	16 (CM)	10	12.5	10	10.833	11
17	14 (ADC)	16 (CM)	10	25	10	15.000	15
18	4 (WWC)	16 (CM)	20	12.5	40	24.167	24
19	11 (ADA)	16 (CM)	30	25	10	21.667	22
20	10 (MAC)	16 (CM)	30	25	30	28.333	28
21	20 (RDG)	21 (SRC)	50	50	40	46.667	47
22	19 (FOI)	21 (SRC)	50	50	60	53.333	53
23	10 (MAC)	39 (UNC)	40	25	60	41.667	42
24	21 (SRC)	39 (UNC)	20	25	20	21.667	22
25	38 (CUN)	39 (UNC)	40	50	20	36.667	37
26	21 (SRC)	37 (KNC)	20	20	20	20.000	20
27	8 (ACF)	37 (KNC)	40	40	40	40.000	40
28	36 (FKN)	37 (KNC)	40	40	40	40.000	40
29	15 (CC)	24 (LAS)	20	30	40	30.000	30
30	16 (CM)	24 (LAS)	30	20	10	20.000	20
31	17 (MAM)	24 (LAS)	30	20	30	26.667	27
32	23 (GAS)	24 (LAS)	20	30	20	23.333	23
33	21 (SRC)	23 (GAS)	40	35	20	31.667	32
34	18 (APG)	23 (GAS)	20	15	30	21.667	21
35	22 (IER)	23 (GAS)	40	50	50	46.667	47
36	24 (LAS)	40 (PF)	25	20	25	22.500	24
37	39 (UNC)	40 (PF)	25	25	25	25.000	25
38	37 (KNC)	40 (PF)	25	25	25	25.000	25
39	34 (OFA)	40 (PF)	10	20	5	15.000	8
40	35* (PP)	40 (PF)	15	20	20	17.500	18
41	33 (EUV)	34 (OFA)	20	25	20	21.667	22
42	32 (BGK)	34 (OFA)	60	50	60	56.667	57
43	29 (LSE)	34 (OFA)	20	25	20	21.667	21
44	30 (PTS)	32 (BGK)	50	50	40	46.667	47
45	31 (PUT)	32 (BGK)	50	50	60	53.333	53
46	25 (PIT)	29 (LSE)					25
47	26 (ICC)	29 (LSE)					25
48	27 (MIP)	29 (LSE)					25
49	28 (PTU)	29 (LSE)					25

Table D2: Pearson Correlation factor of the expert link weightings

Pearson Correlation Factor of the Domain Expert Link Weight Assignments			
Domain Expert	DE1	DE2	DE3
DE1	1		
DE2	$r = 0.62032$ (p =)	1	
DE3	$r = 0.5469$ (p =)	$r = 0.6419$ (p =)	1



APPENDIX E: EMPIRICAL INVESTIGATION RESULTS

Table E1: Average performance metric values for the parameteric groups of models and their maximum standard deviation
(models constructed with the Dp set of training training datasets)

Model Group Type	Model Group Name	Names of Member models of the Group	Training Dataset used for the Construction of the models	Software Tool Used	Structure Learning Approach	% of CP Links Learnt	Parameter Learning Algorithm	Group Average Error Rate and (max STD)	Group Average Logloss and (max STD)	Group Average Brier Score and (max STD)	Group Average Sensitivity and (max STD)
sedacen	PNDp	PND1p PND2p, PND3p, PND4p,	D1p, D2p, D3p, D4p, respectively	Netica	-	-	EM	28.1500 (0.2364)	0.5451 (0.0016)	0.3694 (0.0016)	71.6800 (0.2956)
	PGDp	PGD1p PGD2p, PGD3p, PGD4p	D1p, D2p, D3p, D4p, respectively	Genie	-	-	EM	28.3150 (0.3142)	0.5408 (0.0136)	0.3670 (0.0096)	71.2525 (0.6868)
	PSDp	PSD1p PSD2p, PSD3p, PSD4p	D1p, D2p, D3p, D4p, respectively	Samiam	-	-	EM	48.1556 (0.6107)	0.7007 (0.0245)	0.5018 (0.0178)	14.3169 (0.7753)
todacen	SPGaDp	SPGaD1p, SPGaD2p, SPGaD3p, SPGaD4p	D1p, D2p, D3p, D4p, respectively	Genie	Cl-Test (PC)	100	EM	28.3150 (0.3142)	0.5413 (0.0136)	0.3674 (0.0190)	71.2525 (0.6868)
	SPGbDp	SPGaD1p, SPGaD2p, SPGaD3p, SPGaD4p	D1p, D2p, D3p, D4p, respectively	Genie	SaS (GTT/K2)	100	EM	28.3150 (0.3142)	0.5414 (0.0133)	0.3673 (0.0094)	71.2525 (0.6868)
	SPGcDp	SPGaD1p, SPGaD2p, SPGaD3p, SPGaD4p	D1p, D2p, D3p, D4p, respectively	Genie	SaS (GTT/BDeu)	100	EM	28.3150 (0.3142)	0.5423 (0.0138)	0.3677 (0.0101)	71.2525 (0.6868)
	SPWDp	SPWD1p, SPWD2p, SPWD3p, SPWD4p	D1p, D2p, D3p, D4p, respectively	WinMine	SaS (GSA/BDu)	100	EM	*	*	*	*

* The performance of the models constructed using the WinMine Toolkit could not be evaluated in a similar manner as the other models because of its network file format (.xmod). The file extension .xmod is a representation of the abbreviated words Export Module, and it stores the various graphics and images that are required to generate high resolution outputs. The .xmod network files constructed with WinMine could not be opened by other software tools that support facilities for the evaluations undertaken in this context.

Table E2: Average performance metric values for the non-parameteric groups of models and their maximum standard deviation

(models constructed with the Dnp set of training training datasets)

Model Group Type	Model Group Name	Names of Member models of the Group	Training Dataset used for the Construction of the models	Software Tool Used	Structure Learning Approach	Min/Max No of Links Learnt	Parameter Learning Algorithm	Group Average Error Rate (max STD)	Group Average Logloss (max STD)	Group Average Brier Score (Max STD)	Group Average Sensitivity
sedacen	PNDnp	PND1np, PND2np, PND3np, PND4np,	D1np, D2np, D3np, D4np, respectively	Netica	-	-	EM	48.9056 (0.0006)	0.6929 (0.0006)	0.4997 (0.0006)	52.5475 (0.8283)
	PGDnp	PGD1np, PGD2np, PGD3np, PGD4np	D1np, D2np, D3np, D4np, respectively	Genie	-	-	EM	48.9206 (0.0006)	0.6932 (0.0006)	0.4997 (0.0006)	52.5038 (2.3618)
	PSDnp	PSD1np, PSD2np, PSD3np, PSD4np	D1np, D2np, D3np, D4np, respectively	SamIam	-	-	EM	50.0925 (0.0000)	0.6931 (0.0000)	0.5000 (0.0000)	0.0000 (0.0000)
todacen	SPGaDnp	SPGaD1np, SPGaD2np, SPGaD3np, SPGaD4np	D1np, D2np, D3np, D4np, respectively	Genie	CI-Test (PC)	6/16	EM	50.0825 (0.0002)	0.6931 (0.0001)	0.5 (0.0001)	50.2125 (1.8934)
	SPGbDnp	SPGaD1np, SPGaD2np, SPGaD3np, SPGaD4np	D1np, D2np, D3np, D4np, respectively	Genie	SaS (GTT/K2)	1/2	EM	-	-	-	-
	SPGcDnp	SPGaD1np, SPGaD2np, SPGaD3np, SPGaD4np	D1np, D2np, D3np, D4np, respectively	Genie	SaS (GTT/BDeu)	0/0	EM	-	-	-	-
	SPWDnp	SPWD1np, SPWD2np, SPWD3np, SPWD4np	D1np, D2np, D3np, D4np, respectively	WinMine	SaS (GSA/BDu)	0/0	EM	-	-	-	-



APPENDIX F: QUANTIZATION OF EVIDENCE VARIABLES

APPENDIX F1: Quantization of Node 2: “CorrectPlacementOfComponents”

The *quantization of CPC* necessitates the given circuit or network is completely described by a graph, G , without self-loops, so that the KVL and KCL equations of the network can be derived from the graph by using the relevant components of the graph (path, subgraph, loop, cutset, tree, and cotree) to generate the necessary matrices (Incident Matrix (IM), A_a , and Reduced Incident Matrix (RIM), A ; Loop Matrix (LM), B_a , and Fundamental Loop Matrix (FLM), B ; Cutset Matrix (CM), D_a , and Fundamental Cutset Matrix (FCM), D). The IM, A_a , is generated directly from G . For example, the network of Figure F1.1, is described by the graph, G , of Figure F1.2.

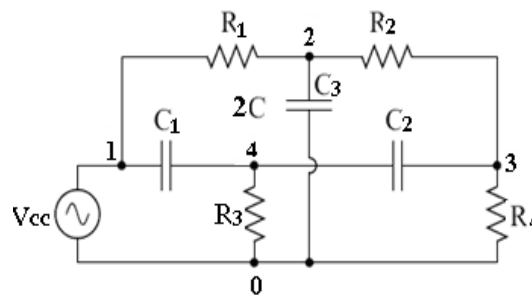


Figure F1.1: The schematic of one of the circuits constituting part of the laboratory activities designed for the evaluation of the VEL: the *Twin-T Notch filter*.

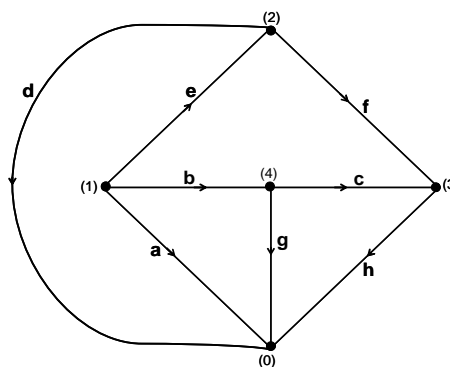


Figure F1.2: G : The graph description of the network of Figure F1.1.

In G , the edges represent the branches of the network, so that a, b, c, d, e, f, g , and h represent $v_{cc}, c1, c2, c3, r1, r2, r3$, and $r4$, respectively. G has n vertices and b edges corresponding to

the n nodes and b branches of the network. The IM, A_a (equation C1.1), is the $n \times b$ matrix, $A_a = [a_{ij}]$, where: $a_{ij} = 1$ if the branch j is incident at node i and the arrow is pointing away from node i ; $a_{ij} = -1$ if branch j is incident at node i , and the arrow is pointing towards node i ; and $a_{ij} = 0$ if branch j is not incident at node i . $A_a i = 0$, where i is the $b \times 1$ column vector, representing the branch currents.

$$A_a = \begin{matrix} & \begin{matrix} \text{nodeNo} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{bmatrix} a & b & c & d & e & f & g & h \end{bmatrix} \\ \begin{matrix} \\ \\ \\ \\ \\ \end{matrix} & & \begin{bmatrix} -1 & 0 & 0 & -1 & 0 & 0 & -1 & -1 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & -1 & 0 & 1 \\ 0 & -1 & 1 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \end{matrix} \begin{matrix} \\ \\ \\ \\ \\ \end{matrix} \begin{matrix} adgh \\ abe \\ def \\ cfh \\ bcg \end{matrix} \quad (\text{F1.1})$$

and $A_a i = 0$ (F1.2)

where i is the $b \times 1$ column vector,

$$i(t) = \begin{bmatrix} i_a \\ i_b \\ i_c \\ i_d \\ i_e \\ i_f \\ i_g \\ i_h \end{bmatrix} \quad (\text{F1.3})$$

representing the branch currents. Each row of A_a corresponds to a vertex, $i, i=1$ to n , of G , and each column corresponds to an edge, $j, j = 1$ to b . $A_a i = 0$ is the set of n KCL equations characterizing the network. The equations are not linearly independent. The required $n-1$ linearly independent KCL equations are obtained by transforming the IM, A_a , to the RIM, A (equation C1.4) in order to derive the equation, $Ai = 0$, which is the required set of $n-1$ independent KCL equations.

$$A = \begin{array}{c} \text{nodeNo} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \left[\begin{array}{cccccccc} a & b & c & d & e & f & g & h \end{array} \right] \\ \left. \begin{array}{l} -1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & -1 & 0 & 1 \\ 0 & -1 & 1 & 0 & 0 & 0 & -1 & 0 \end{array} \right\} \begin{array}{l} \text{abe} \\ \text{def} \\ \text{cfh} \\ \text{agh} \end{array} \end{array} \quad (\text{F1.4})$$

$$\text{and} \quad A_i = 0 \quad (\text{F1.5})$$

G can be partitioned into a chosen tree, T , and its cotree, T_C , so that all its edges are split into two groups: those that form the tree and those that form the cotree. Hence, A can be partitioned as $A=[A_T|A_C]$ where the columns of A_T correspond to the tree branches of the chosen tree, T , and the columns of A_C correspond to the branches of the cotree, T_C (referred to as *chords*), and $\det A_T \neq 0$. If, for example, the chosen tree is as shown in Figure F1.3, with branches $(b\ c\ d\ h)$, then A can be partitioned as in equation F1.6.

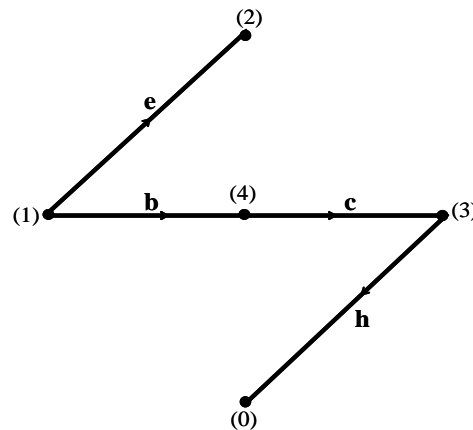
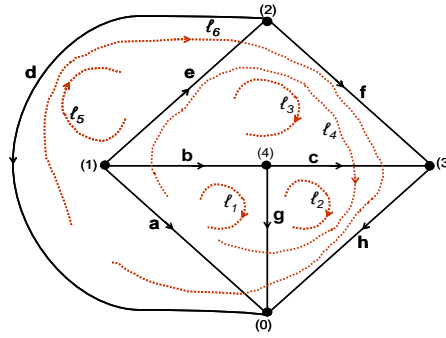


Figure F1.3: A chosen tree, T , of the graph, G .

$$A = \begin{array}{c} \text{nodeNo} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \left[\begin{array}{cccc|cccc} b & c & d & h & a & e & f & g \\ 1 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 1 & 0 \\ 0 & -1 & 0 & 1 & 0 & 0 & -1 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 & 0 & -1 \end{array} \right] \end{array} \quad (\text{F1.6})$$

The LM, B_a (equation 7.8), is formed from the q oriented loops of G , highlighted in Figure F1.4.

Figure F1.4: Oriented Loops of the graph, G .

$$B_a = \begin{matrix} \text{loopNo} \\ l_1 \\ l_2 \\ l_3 \\ l_4 \\ l_5 \\ l_6 \end{matrix} \begin{bmatrix} a & b & c & d & e & f & g & h \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 & 1 \\ 0 & -1 & -1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ -1 & 0 & 0 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & - & 1 & 0 & 1 \end{bmatrix} \quad (\text{F1.7})$$

$$\text{and} \quad B_a V = 0 \quad (\text{F1.8})$$

Equation F1.7 is the set of q KVL equations describing the network, where v is the $b \times 1$ column vector (equation F1.9),

$$v(t) = \begin{bmatrix} v_a \\ v_b \\ v_c \\ v_d \\ v_e \\ v_f \\ v_g \\ v_h \end{bmatrix} \quad (\text{F1.9})$$

representing the branch voltages. Hence, the LM, $B_a = [b_{ij}]$, is a $q \times b$ matrix, where $b_{ij} = 1$ if the branch j is in loop i , and their arrow directions agree; $b_{ij} = -1$ if branch j is in loop i , and their arrow directions oppose; $b_{ij} = 0$ if branch j is not in loop i . $B_a V = 0$ is the set of q KVL equations describing the network, where V is the $b \times 1$ column vector representing the branch

voltages. The q KVL equations are not linearly independent. The required set of linearly independent KVL equations is $BV = 0$, where B is the FLM and is systematically generated from G , using the chosen tree, T and its cotree, T_C . B can also be generated from the systematic transformation of B_a . The columns of B can be partitioned as $B = [B_T | B_C]$, where the columns of B_T and B_C correspond to the branches of the tree, T , and the chords of its cotree, T_C , respectively.

Furthermore, the oriented cutsets of G , highlighted in Figure F1.5, based on chosen tree, T , are used to form the CM, D_a . G has $m = n-1$ oriented cutsets so that D_a is the $m \times b$ matrix, $D_a = [d_{ij}]$, where $d_{ij} = 1$ if the branch j is in cutset i , and their arrow directions agree; $d_{ij} = -1$ if branch j is in cutset i , and their arrow directions oppose; $d_{ij} = 0$ if branch j is not in cutset i . $D_a i = 0$ is a set of linearly dependent KCL equations.

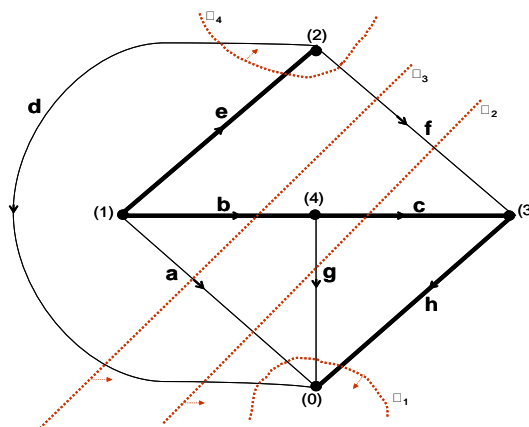


Figure F1.5: FSC associated with the tree, T , of the graph G .

$$D = \begin{matrix} & \text{cutsetNo} & \begin{bmatrix} a & b & c & d & e & f & g & h \end{bmatrix} & \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix} & & \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ -1 & 0 & 1 & -1 & 0 & 1 & -1 & 0 \\ -1 & 1 & 0 & -1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & -1 & 0 & 0 \end{bmatrix} & \begin{matrix} adgh \\ acdfg \\ abdf \\ def \end{matrix} \end{matrix} \quad (\text{F1.10})$$

The FCM, D , an $m \times b$ matrix ($m = n - 1$), required to form the set of linearly independent KCL equations, $Di = 0$, is derived from the FSC. The FSC is systematically generated from G using the tree, T and cotree, T_C . Or, D can be systematically generated directly from D_a

such that D consists of the maximum number of independent rows of D_a . D can also be partitioned as $D=[D_T|D_C]$. The rows of D are linearly independent and

$$D_i = 0 \quad (\text{F1.11})$$

Equation F1.11 represents a more generalized set of KCL equations for the network [Chua and Lin, 1975].

So far, the generation of three sets of KCL and KVL equations have been described. From the way B is generated, using T and T_c , each chord of T_c in B_C constitutes a unit column matrix of B_C so that $B = [B_T|1_U]$, where $u = b - n + 1$ and $T = \text{tree}$. Similarly, each branch of T in D_T constitutes a unit column matrix of D_C so that $D = [1_p | D_C]$, where $p = n - 1$ and $c = \text{chords}$. The check for topological equivalence of two networks is done by comparing their sets of KCL equations ($B_i = 0$) and/or KVL equations ($A_i = 0$, or $D_i = 0$), which, in this context, reduces to matrix comparison. Hence, for the purpose of the work presented in this thesis, a student's built circuit is compared to the reference circuit, for topological equivalence, by comparing their D matrices, since $D_i = 0$, according to [Chua and P. Lin, 1975], represents the set of more generalized KCL equations for a network. The result of the comparison is used to generate the quantized value for the node, CPC , using equation F1.12.

$$CPC = \frac{100}{n} \times q \left(\frac{\sum_{i=1}^n x_i}{m} \right) \quad (\text{F1.12})$$

where n = number of different types of reference circuits, m = number of circuits built and simulated by a student, x_i = number of circuits built and simulated by a student that are topologically equivalent to reference circuit type i , and q = number of x_i 's not equal to zero (q is an indication that a circuit of a particular type is built topologically correctly at least once).

APPENDIX F2: Quantization of the node 3: "UseCorrAlternativeComponent"

This node is quantified if and only if (*iff*) a laboratory activity requires students to adapt in the face of non availability of a specified component. For example, if a given laboratory circuit diagram specifies the use of a 220k Ω resistor which is not among the provided components in the Resistor component container, students are expected to adapt by using other resistor values, combing them serially or in parallel to obtain the same circuit effect as a 220k Ω resistor. In this case, the quantization of the node takes a *transfer function* approach by comparing the student's built circuit to the reference circuit, not for component/value and

topological equivalence, but for equivalence in their output characteristics. If the student and reference circuits have the same output characteristics, then the student is assumed to have used a feasible alternative component. Hence, UAC is quantized as follows:

$$UAC = \frac{100}{n} \times q \left(\sum_{i=1}^n x_i / m \right) \quad (F1.12)$$

where n = number of different types of reference circuits in which adaptation is required, m = number of circuits built and simulated by student, x_i = number of circuits built and simulated by student in which adaptation has been made and their output characteristics are equivalent to that of reference circuit type i ($i=1 \dots n$) that required adaptation, and q = number of x_i 's not equal to zero (q is an indication that a circuit of a particular type is built with the correct alternative components, at least once).

APPENDIX F3: Quantization of the node 5: “ObserveSafetyProcedure”

In the traditional laboratory, there usually are safety related issues that students must necessary observe. It may not be possible to completely observe all laboratory safety related procedures in the VEL. However, there is a number of breadboarding related and equipment use safety procedures that students can and should be expected to observe, even in the VEL environment, because it is possible to observe them and they need to imbibe the culture of being safety conscious. Safety requires that the Power Supply (PS) or Function Generator (FG) be:

- connected to the breadboard after the circuit has been built, and not before.
- connected to the breadboard first before it is switched on and not the other way round.
- switched off before disconnection from the breadboard.
- be switched off and possibly disconnected before making changes to the breadboarded circuit such as:
 - repositioning the pins of a component
 - removing a component
 - adding additional components to the circuit
 - rewiring.
- the two pins of a component should not be connected to the same hole on the breadboard.

The quantization of OSP requires that the actions log is checked for the violation or non-violation of any of these safety related issues. Let n = total number of violations observed and

m = total number of good safety practices observed, then $k = m + n$ (k = total number of safety related actions taken by student).

$$OSP = 100 - \left(\frac{n}{k} \times 100 \right) \quad (F3.1)$$

APPENDIX F4: Quantization of the node 6: “Operate Equipment Correctly”

Laboratory equipment should not be mishandled and should be used in the best possible way that promotes safety and longevity. The precautions that can possibly be taken in the virtual laboratory environment with respect to correct equipment use include:

- the +V and ground (GND) terminals of the PS/FG should not be connected to the same hole on the breadboard
- the PS should not be used where the FG is supposed to be used, and vice versa
- PS/FG switch flickering should be avoided. Flickering occurs when equipment is switched on, off, on, off, on, etc in succession.
- *housekeeping*: switch off and disconnect the PS/FG on finishing a laboratory activity, before quitting the laboratory environment.
- the two pins of a component should not be connected to the same hole on the breadboard.

The quantization of *OEC* requires that the actions log is checked for the violation or non-violation of any of these good equipment use issues. Let n = total number of misuses observed and m = total number of correct-uses observed, then $k = m + n$ (k = total number of correct equipment use related actions taken).

$$OEC = 100 - \left(\frac{n}{k} \times 100 \right) \quad (F4.1)$$

APPENDIX F5: Quantization of the nodes 8: “Apply Correct Formula”

These nodes can be quantified *iff* a laboratory activity requires students to make specific calculations for which they are expected to apply specific formulae. MCQs are used to elicit, from a student, the formulae used for the particular calculation. The selected formula applied by the student is checked against the formula specified by the instructor. Let y = no of different types of formulae to be applied for the specified calculations to be made; x = no of formulae used by a student that are correct and right for the specified calculations, then

$$ACF = \frac{x}{y} \times 100 \quad (F5.1)$$

APPENDIX F6: Quantization of the node 9: “UseCorrectValues”

These nodes can be quantified *iff* a laboratory activity requires students to make specific calculations for which they are expected to apply specific formulae, with certain values for the variables in the formulae, to arrive at expected results. The quantize the variable:

- check if the correct formula was applied
- check if the calculation result is correct by comparing the student calculation result with the expected result specified by the teacher.

Correct values are assumed to have been used if the correct formula was applied and the calculation result is correct. The assumption is that each calculation requires the use of only one specific formula. Let y = no of different calculations to be made; w = no of time correct values used in the calculations, then

$$UCV = \frac{w}{y} \times 100 \quad (\text{F6.1})$$

APPENDIX F7: Quantization of the node 12: “SpecificActivityConstraints”

This node can be quantified *iff* for a laboratory activity, the instructor highlights a set of specific constraints that must be adhered to. For example, the instructor may specify a constraint such as: $c1 = \frac{1}{2}c2$. The form of constraint specified by the instructor determines the source of data extract for the quantification of this node. Let y = no of specific activity constraints specified and x = no of specific activity constraints adhered to, then

$$SAC = \frac{x}{y} \times 100 \quad (\text{F7.1})$$

APPENDIX F8: Quantization of the node 17: “AbilityToMakeAppropriateMeasurements”

The variable, MAM , is quantized as:

$$MAM = \frac{100}{n} \times q \left(\frac{\sum_{i=1}^n x_i}{m} \right) \quad (\text{F8.1})$$

where n = number of different types of measurements instructed, m = total number of measurement attempts made by a student, x_i = number of measurements by student that are the same as the instructed measurement type i , and q = number of x_i 's not equal to zero (q is an indication that an instructed measurement type is made correctly, at least once).

APPENDIX F9: Quantization of the node 18: “AbilityToPlotGraph”

This node is instantiated *iff* a laboratory activity requires students to plot a graph(s). It is quantified as follows:

$$APG = \frac{100}{n} \times q \left(\sum_{i=1}^n x_i / m \right) \quad (\text{F9.1})$$

where n = number of different types of graph plots to be made; m = number of graph plot attempts made by a student, x_i = number of graph plot attempts that are correct, corresponding to a specified graph plot (with the correct signal frequency range specified, where necessary), and for which the student constructed circuit that yielded the data used for the graph plot is equivalent to its reference circuit, q = number of x_i 's not equal to zero (q is an indication that an instructed graph plot type is done correctly, at least once).

APPENDIX F10: Quantization of the node 19: “FollowInstructions”

The variable is quantized with respect to the instructional tasks for a laboratory activity such as: *construct a circuit, plot graph (with or without given specifications, e.g. frequency range), modify a constructed circuit, calculate, make a measurement, give all answers to two decimal places, etc.* The number of decimal places for a student answer, for example, is verified by checking if all the student's numerical results/answers are rounded to the number of specified decimal places. Hence, the *FOI* node is quantized as:

$$FOI = \frac{x}{y} \times 100 \quad (\text{F10.1})$$

where y = total number of different instructions given, and x = the number of instructions actually carried out by the student. The outcome of carrying out the instruction may or may not be correct but it is only checked that attempted has been made to do as instructed. The order in which the instructions are carried out is not taken into consideration in this context.

APPENDIX F11: Quantization of the node 20: “ReadDiagramsAndGraphs”

This node can be quantified either partially or fully, or not at all, for every laboratory activity, because the quantization involves two different aspects: reading of diagrams and reading of

graphs. The reading of diagrams can possibly always be quantized for every laboratory activity that includes a circuit diagram, as part of its instructions or if some laboratory activity questions involve reading a diagram(s) in order to answer the questions. For example, if a student's built circuit is equivalent to the reference laboratory schematic then the student has read the given diagram well. The reading of graphs aspect can be quantified *iff* a laboratory activity requires students to read off values from plotted graph(s) or to express their understanding of a plotted graph. Each laboratory activity question has an indicator, highlighting what it is designed to elicit. All the questions that elicit the *RDG* construct are profiled and used for the quantization. Let y = number of laboratory questions and post-laboratory items that elicit reading diagrams and/or graphs, and x = number of laboratory questions and post-laboratory items that elicit reading diagrams and/or graphs for which student's answers/responses are correct. Then, *RDG* is quantized as:

$$RDG = avg \left(UCC + CPC + \frac{x}{y} \times 100 \right) \quad (F11.1)$$

APPENDIX F12: Quantization of the node 22: “AbilityToInterpretExpResults”

This node is quantized and instantiated *iff* a laboratory activity requires students to read off data, explain and/or extract information from a laboratory activity result(s). For example, students' may be asked to state the reason for the difference between two plotted graphs. Each laboratory activity question has an indicator, highlighting what it is designed to elicit. The questions that elicit the *IER* construct are profiled and used for the quantization. Let q = total number of laboratory questions and post-laboratory items that elicit students' ability to interpret experimental results, and p = total number of laboratory questions and post-laboratory items that elicit students' ability to interpret experimental results for which student's answers/response are correct. Then

$$IER = \frac{p}{q} \times 100 \quad (F12.1)$$

APPENDIX F13: Quantization of the node 38: “ConceptualUnderstanding”

This node is quantified from marking a student's answers/responses to laboratory questions and post-laboratory items designed to elicit students' understanding of the concept addressed by a laboratory activity. Each laboratory question and post-laboratory item has an indicator

highlighting what it is designed to elicit. The questions that elicit students' understanding of the concept addressed by a laboratory activity are profiled and used for the quantization. Let y = total number of laboratory activity questions and post-laboratory items designed to elicit students' understanding of the addressed, and x = total number of laboratory activity questions and post-laboratory items designed to elicit students' understanding of the addressed for which students answers/responses are correct. Then

$$CUN = \frac{x}{y} \times 100 \quad (\text{F13.1})$$

APPENDIX F14: Quantization of the node 36: “*FactualKnowledge*”

This node is quantified from marking a student's answers/responses to laboratory questions and post-laboratory items designed to elicit students' knowledge of the concept addressed by a laboratory activity. Each laboratory question and post-laboratory item has an indicator highlighting what it is designed to elicit. The questions that elicit students' knowledge of the concept addressed by a laboratory activity are profiled and used for the quantization. Let y = total number of laboratory activity questions and post-laboratory items designed to elicit students' knowledge of the addressed, and x = total number of laboratory activity questions and post-laboratory items designed to elicit students' knowledge of the addressed for which students answers/responses are correct. Then

$$FKN = \frac{x}{y} \times 100 \quad (\text{F14.1})$$

APPENDIX F15: Quantization of the node 30: “*PreLaboratoryTestScore*”

This node is directly quantified from the database as students' pre-laboratory test scores are stored in the database.

APPENDIX F16: Quantization of the node 31: “*PreRequisiteUnitTestScore*”

This node is directly quantified from the database as students' test scores for pre-requisite units are stored in the database.

APPENDIX F17: Quantization of the node : “*PerceivedInfoType*”; “InfoCommChannel”; “InfoProcessing”; and “ProgressionToUnderstanding”

These nodes are not quantized; rather, they are instantiated directly from the database. Students preferred Learning Styles (LSs) were evaluated with the [Felder and Silverman, 1988] Index of Learning Styles (ILS) instrument. Students LS preferences on the four dimensions of the ILS were generated and stored in the database. Since the nodes are not quantized, they do not require fuzzification.



APPENDIX G: ASSESSOR CONSISTENCY

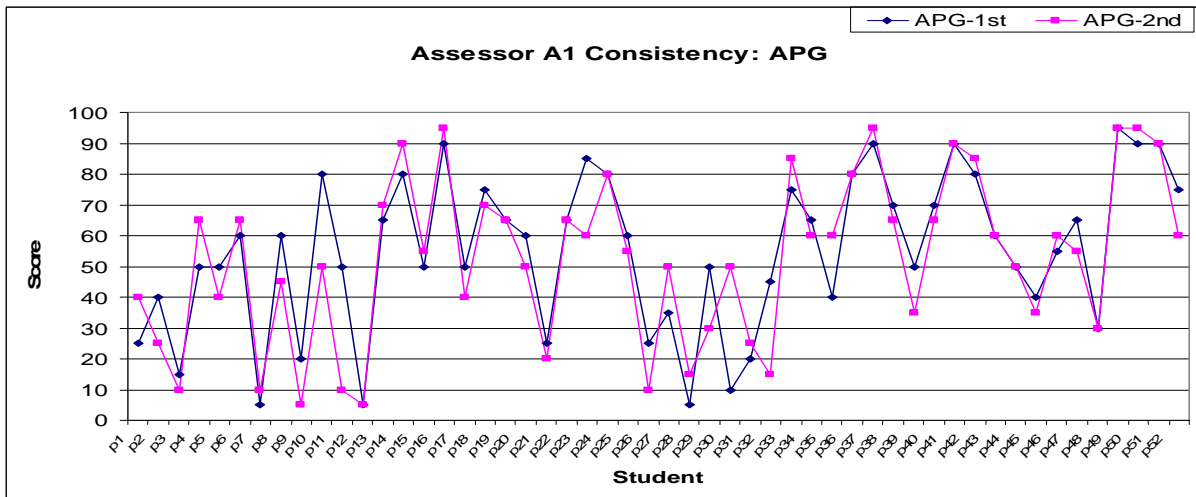


Figure G1: Assessor A1 1st and 2nd assessment scores for the performance factor, APG

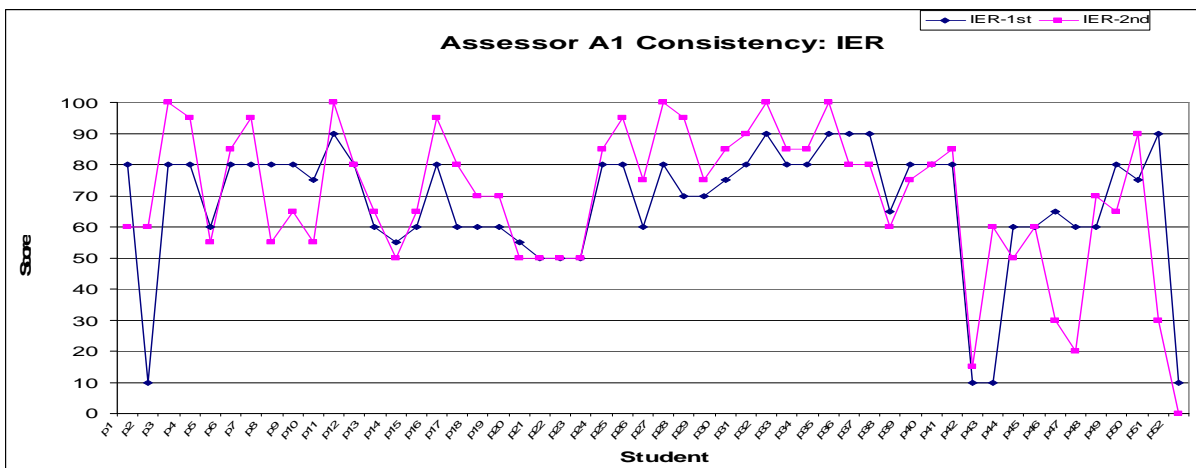


Figure G2: Assessor A1 1st and 2nd assessment scores for the performance factor, IER

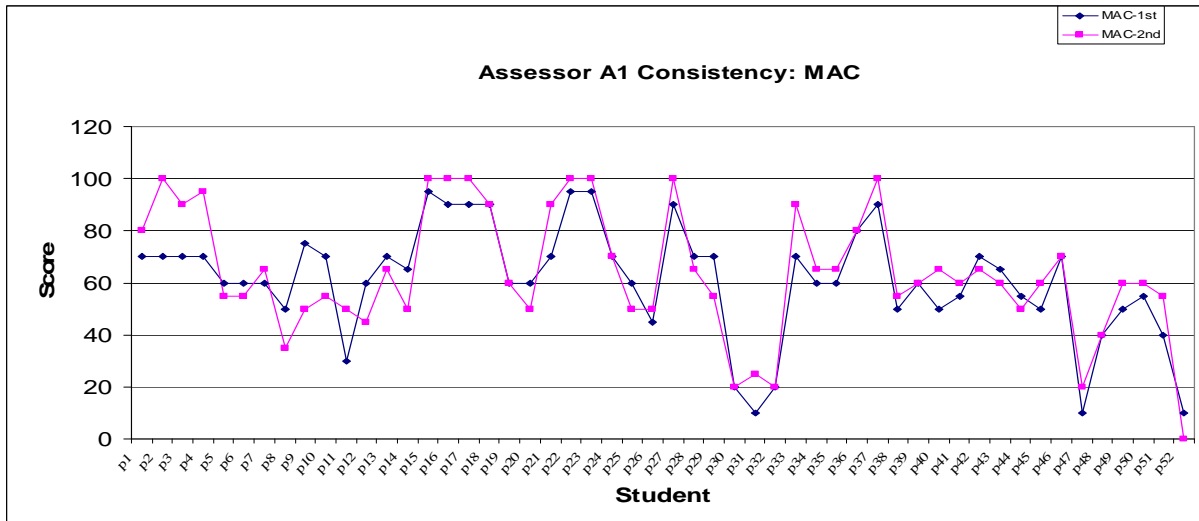


Figure G3: Assessor A1 1st and 2nd assessment scores for the performance factor, MAC

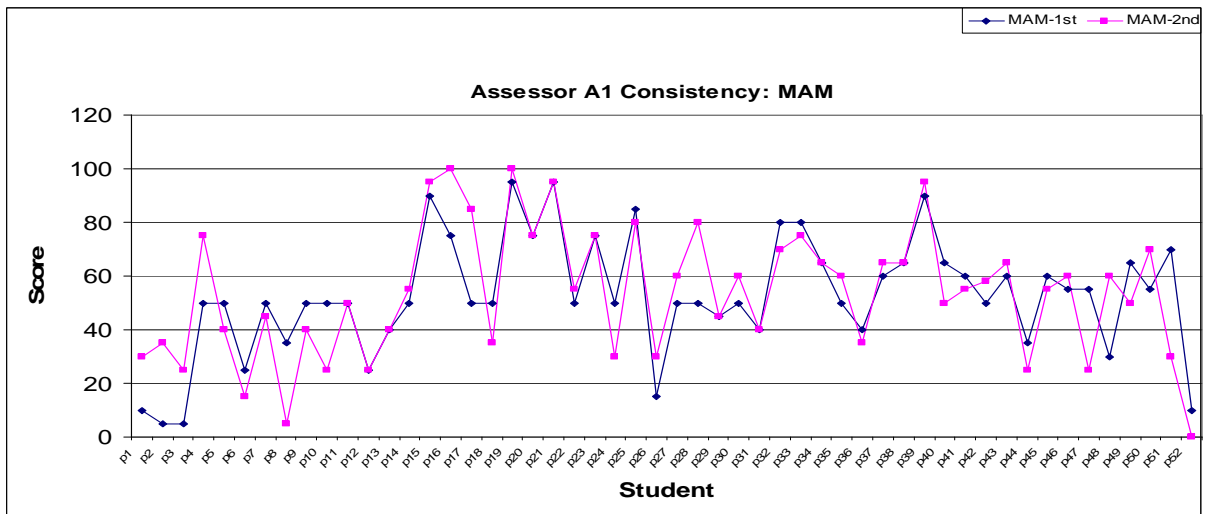


Figure G4: Assessor A1 1st and 2nd assessment scores for the performance factor, MAM

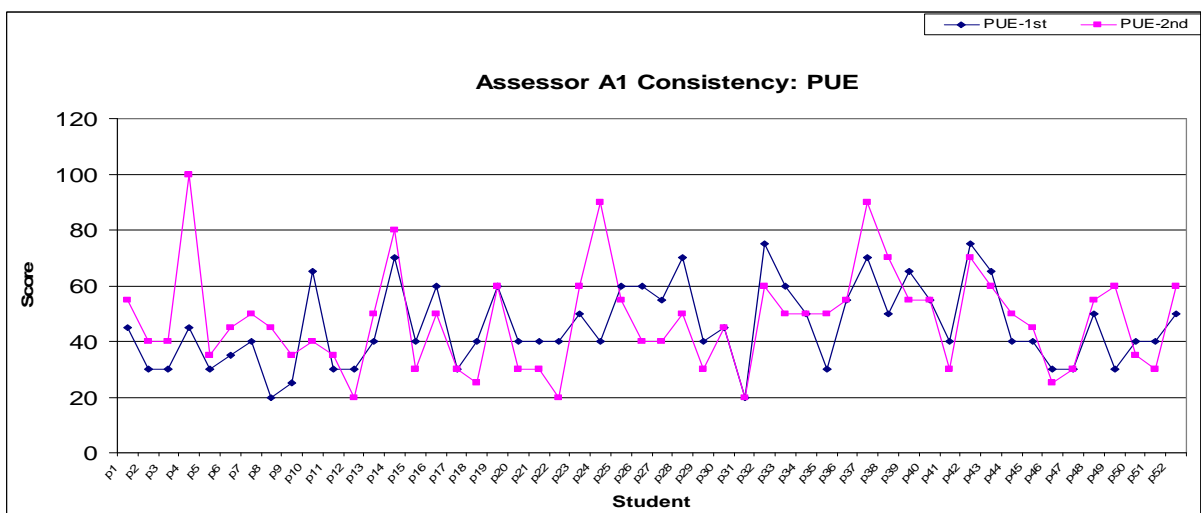


Figure G5: Assessor A1 1st and 2nd assessment scores for the performance factor, PUE

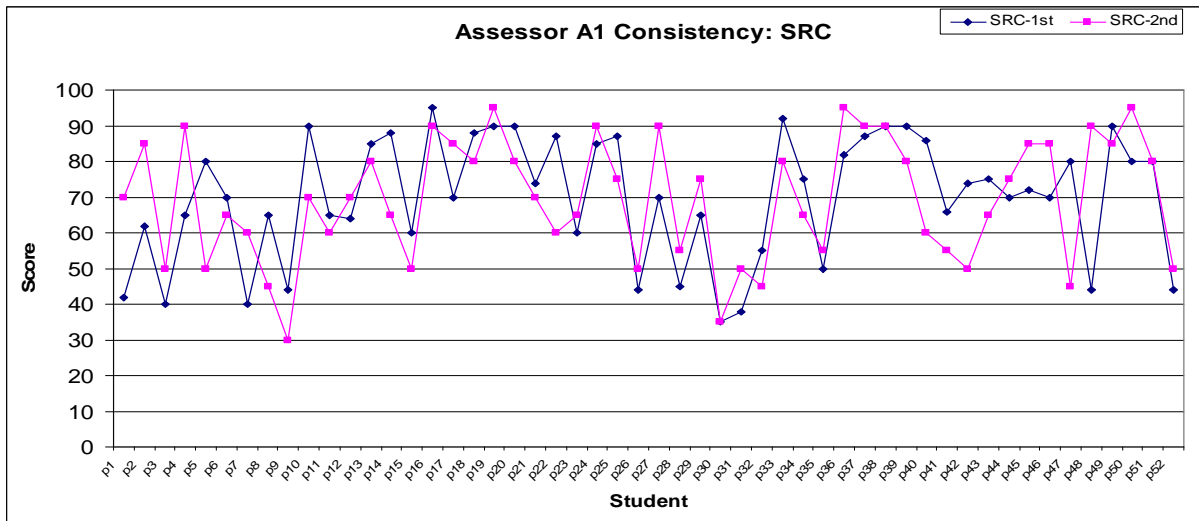


Figure G6: Assessor A1 1st and 2nd assessment scores for the performance factor, SRC

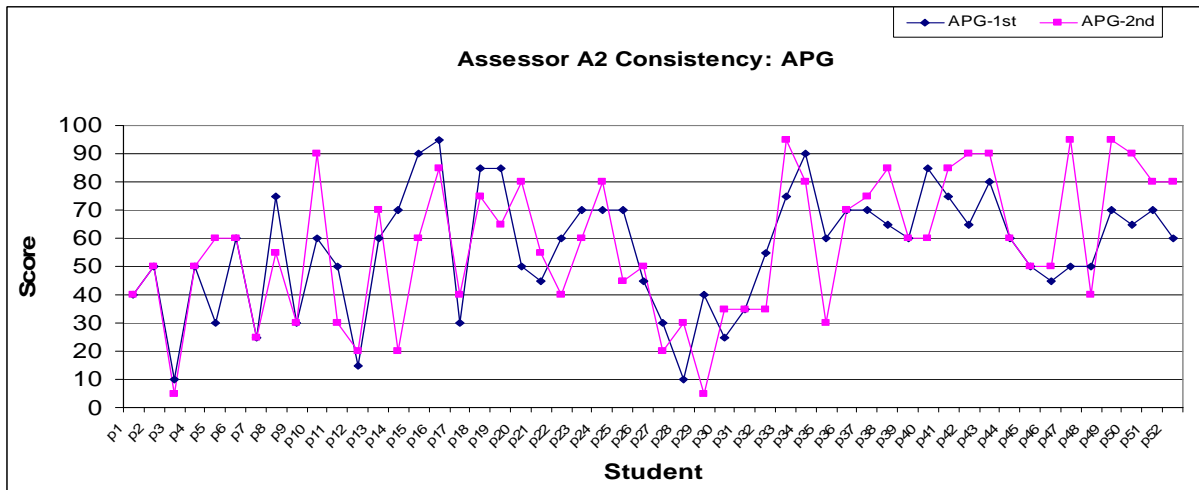


Figure G7: Assessor A2 1st and 2nd assessment scores for the performance factor, APG

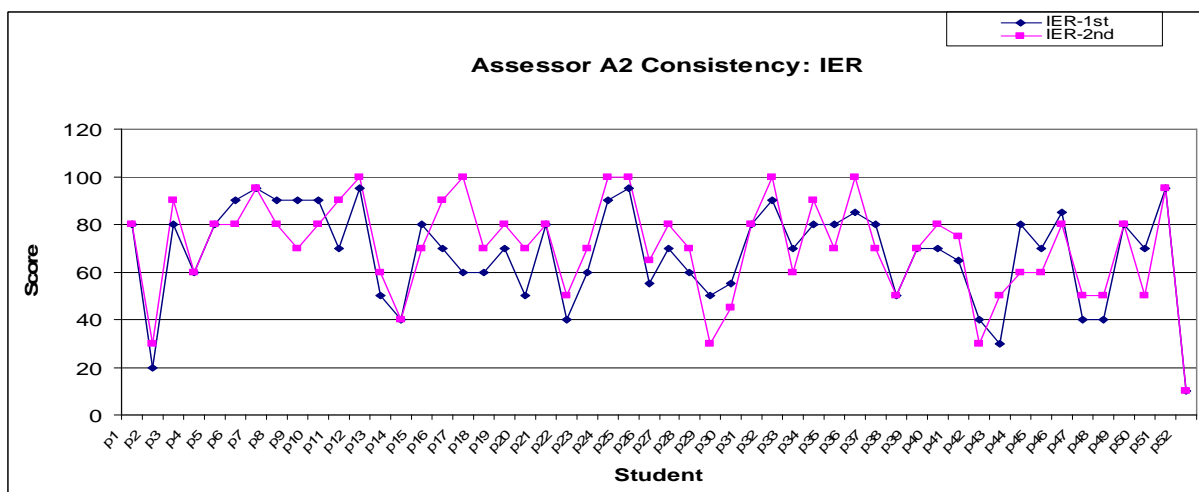


Figure G8: Assessor A2 1st and 2nd assessment scores for the performance factor, IER

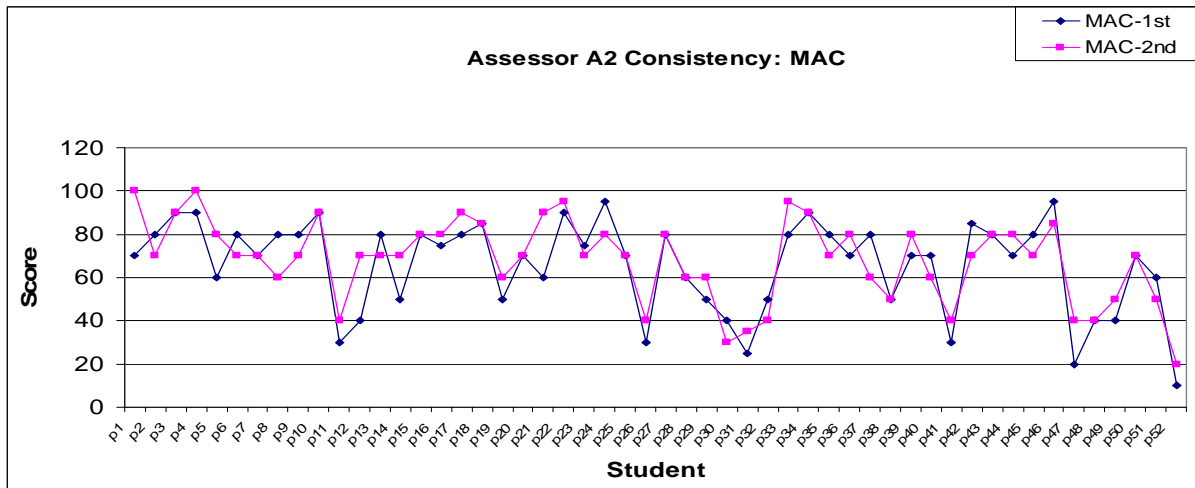


Figure G9: Assessor A2 1st and 2nd assessment scores for the performance factor, MAC

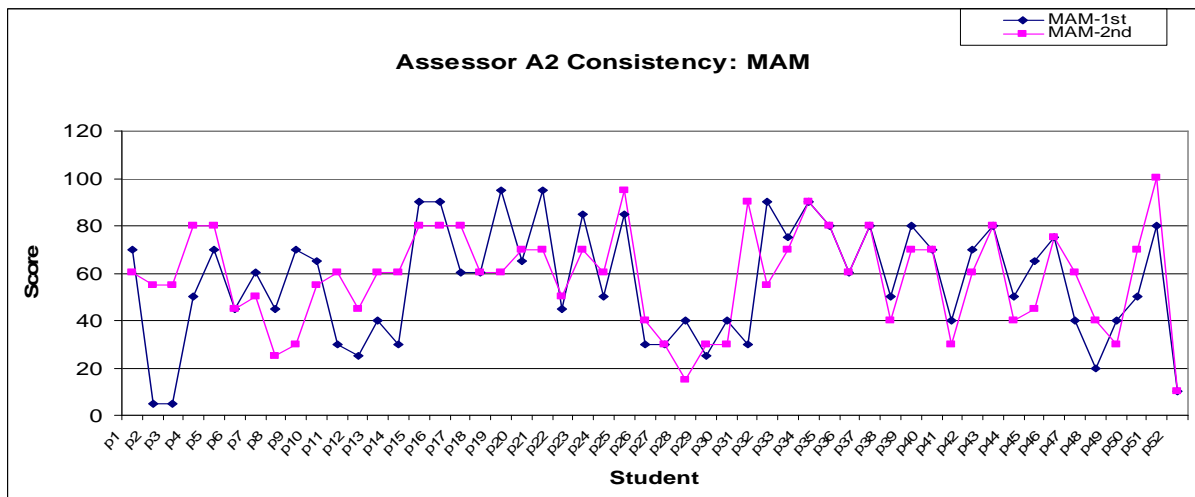


Figure G10: Assessor A2 1st and 2nd assessment scores for the performance factor, MAM

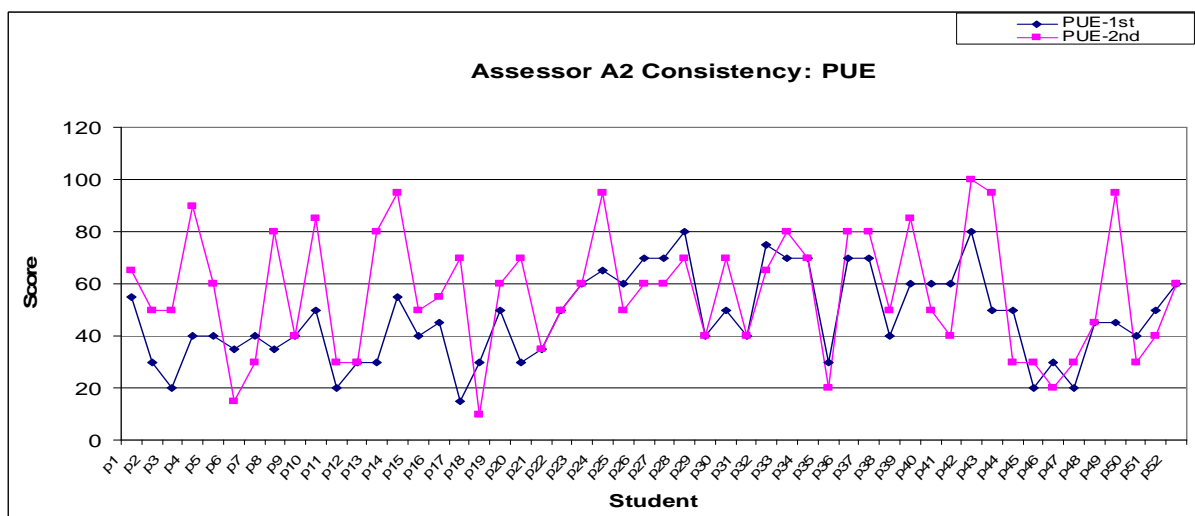


Figure G11: Assessor A2 1st and 2nd assessment scores for the performance factor, PUE

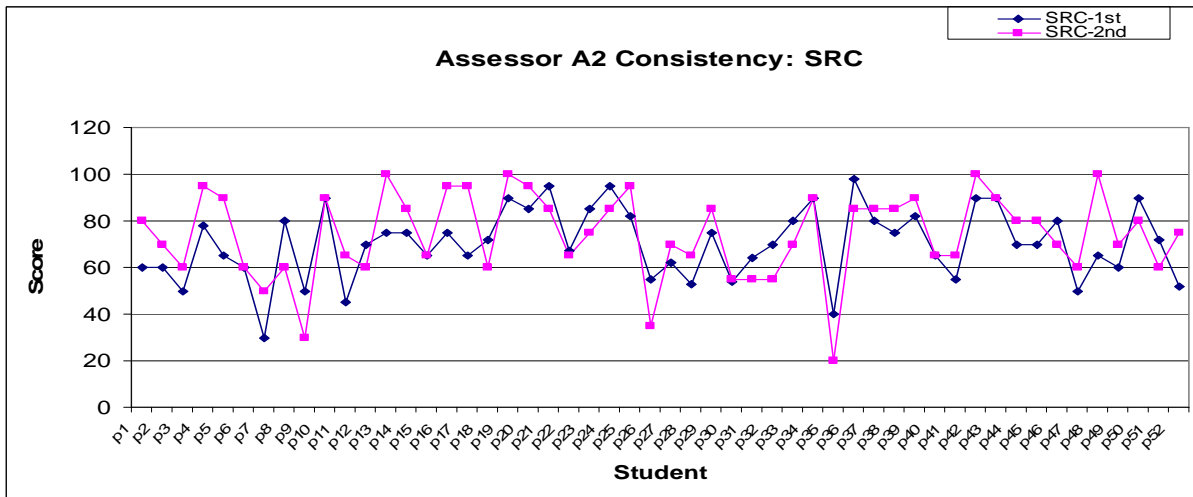


Figure G12: Assessor A2 1st and 2nd assessment scores for the performance factor, SRC

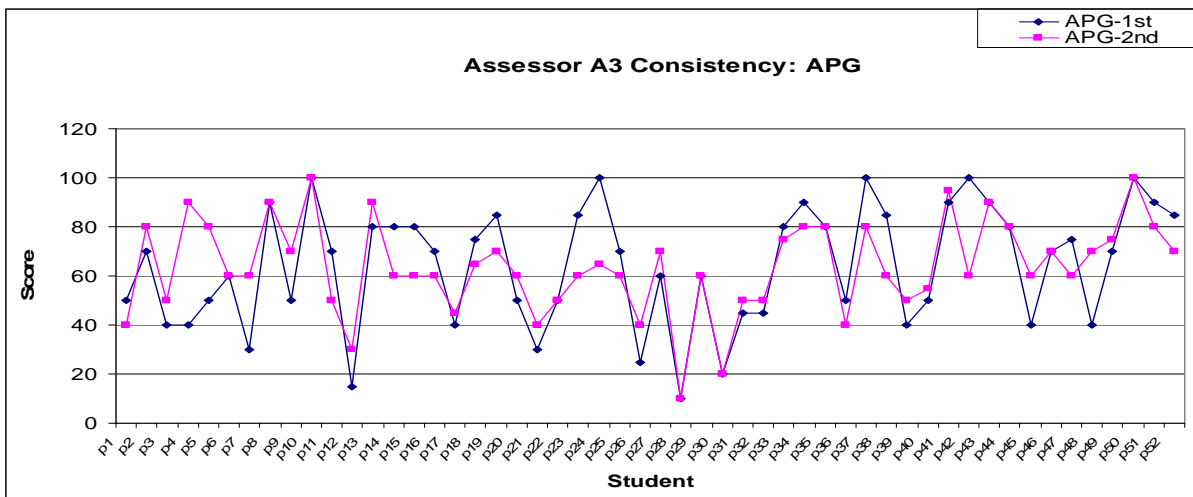


Figure G13: Assessor A3 1st and 2nd assessment scores for the performance factor, APG

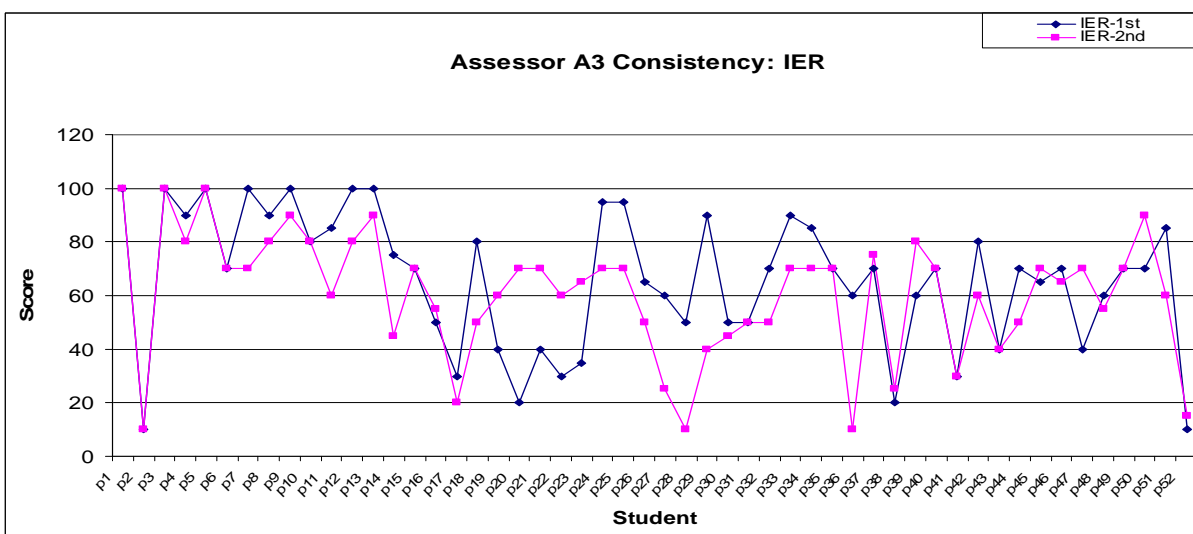


Figure G14: Assessor A3 1st and 2nd assessment scores for the performance factor, IER

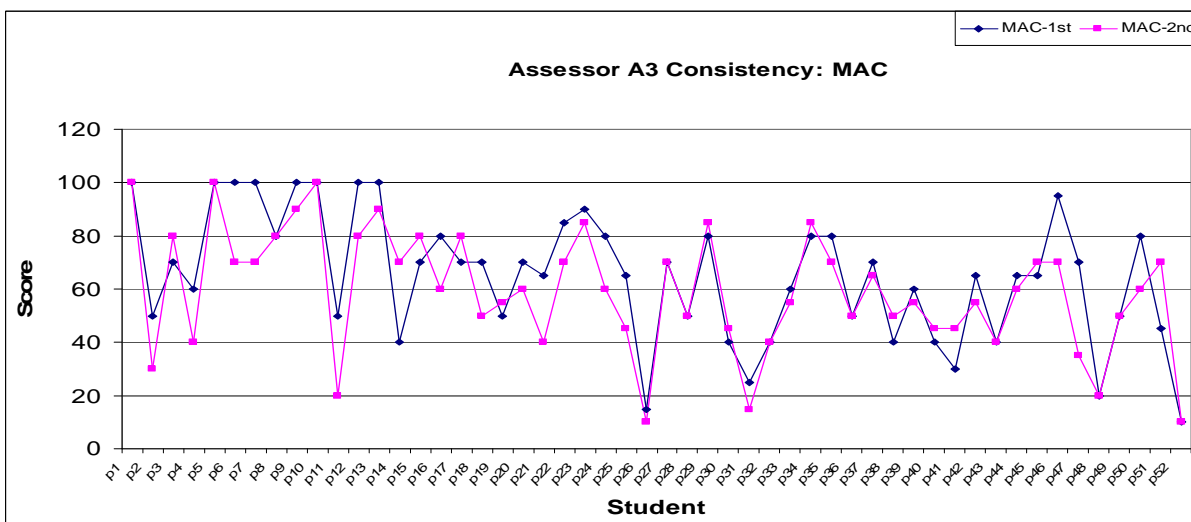


Figure G15: Assessor A3 1st and 2nd assessment scores for the performance factor, MAC

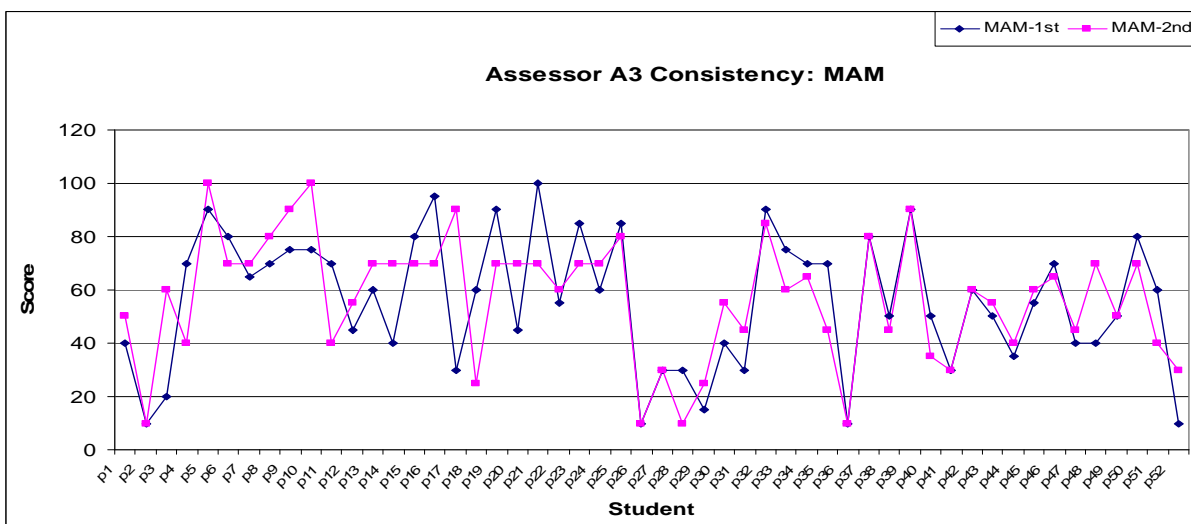


Figure G16: Assessor A3 1st and 2nd assessment scores for the performance factor, MAM

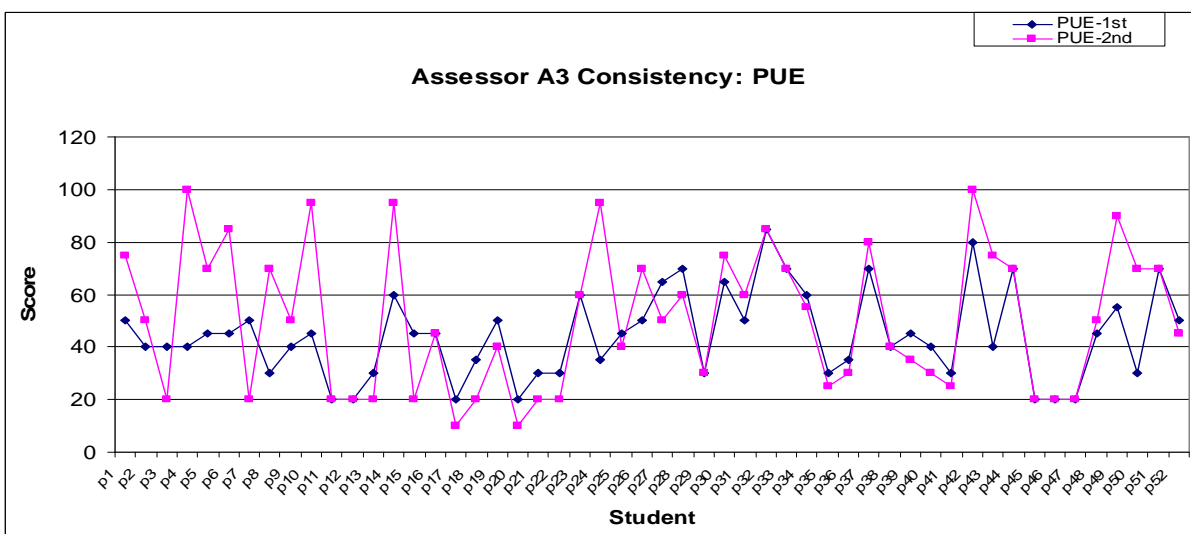


Figure G17: Assessor A3 1st and 2nd assessment scores for the performance factor, PUE

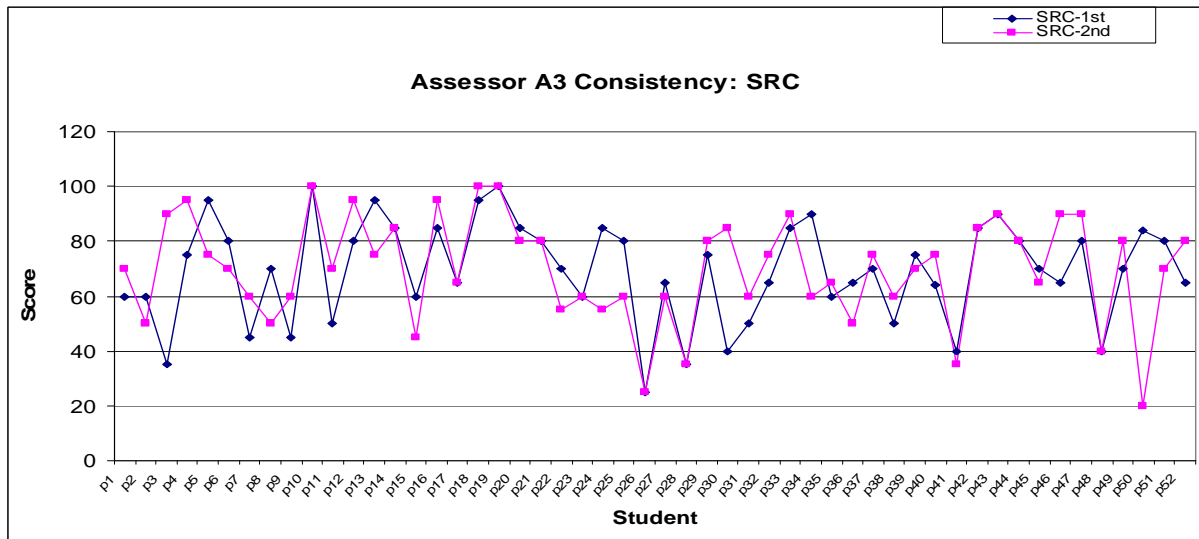


Figure G18: Assessor A3 1st and 2nd assessment scores for the performance factor, SRC

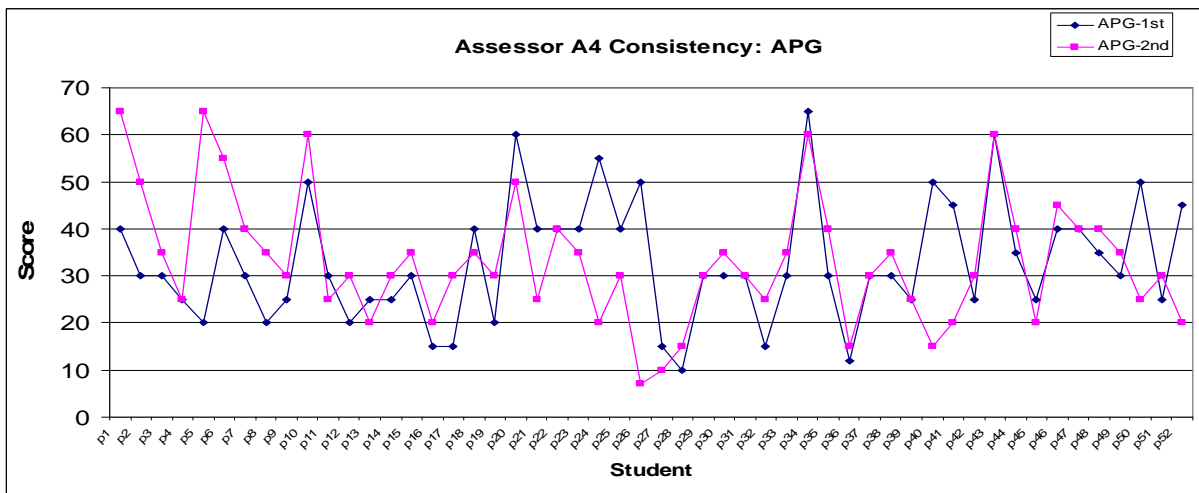


Figure G19: Assessor A4 1st and 2nd assessment scores for the performance factor, APG

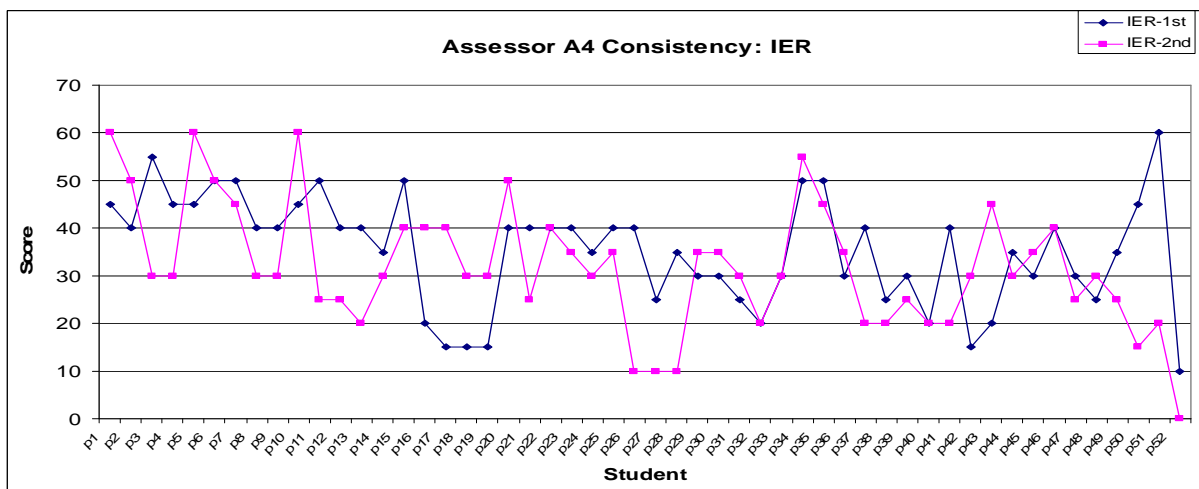


Figure G20: Assessor A4 1st and 2nd assessment scores for the performance factor, IER

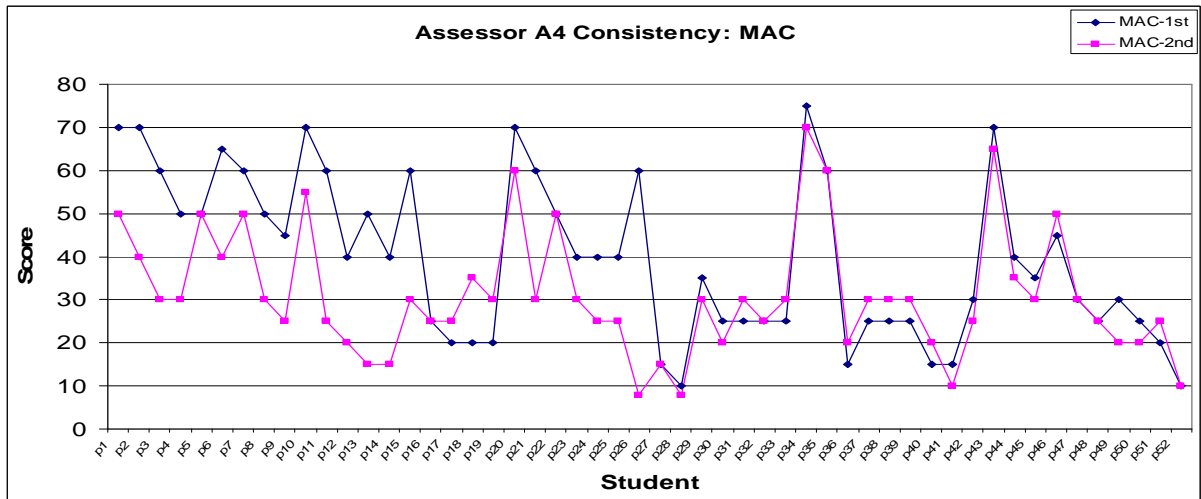


Figure G21: Assessor A4 1st and 2nd assessment scores for the performance factor, MAC

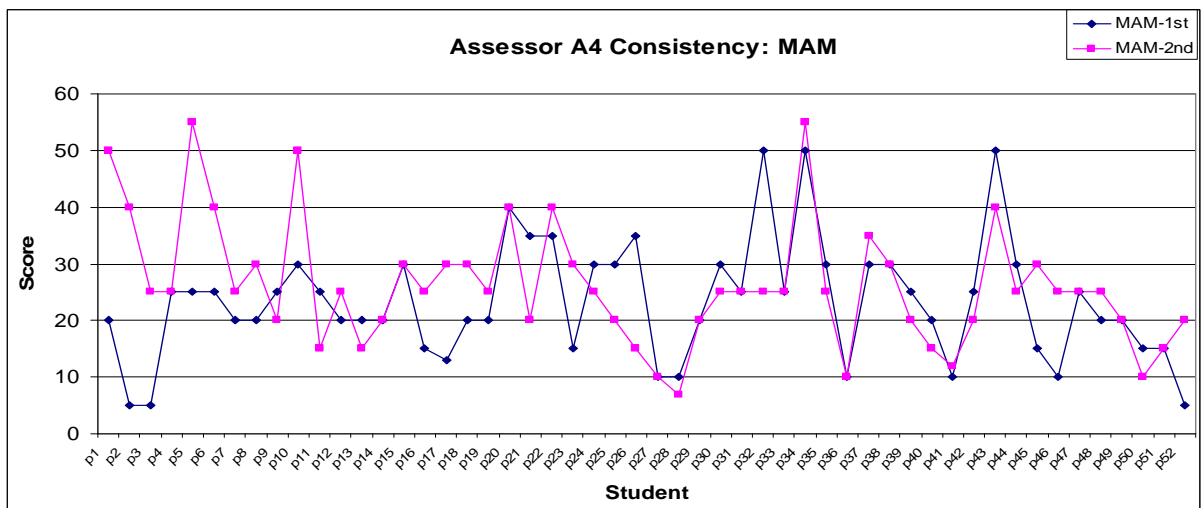


Figure G22: Assessor A4 1st and 2nd assessment scores for the performance factor, MAM

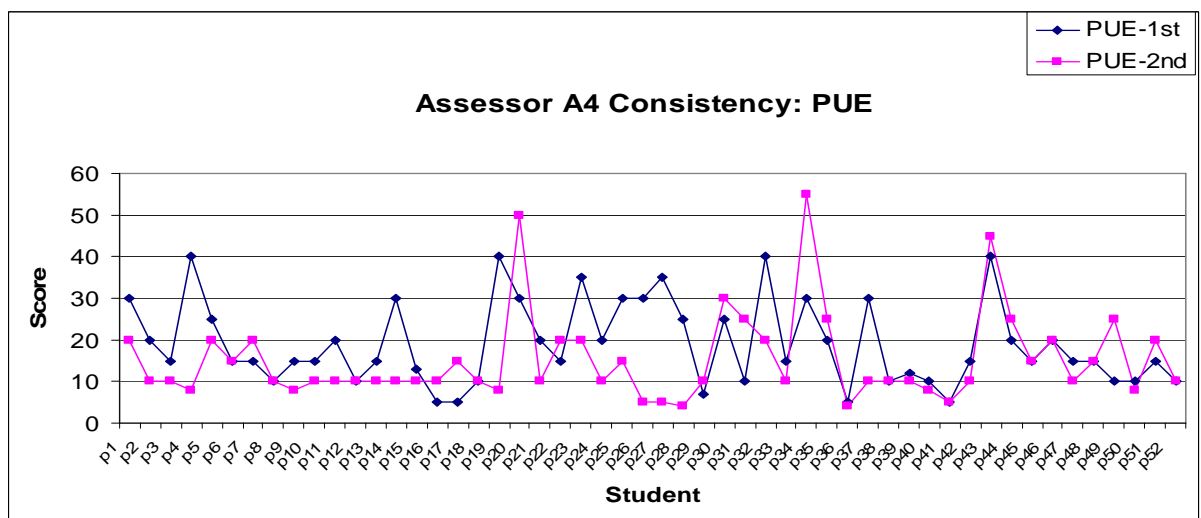


Figure G23: Assessor A4 1st and 2nd assessment scores for the performance factor, PUE

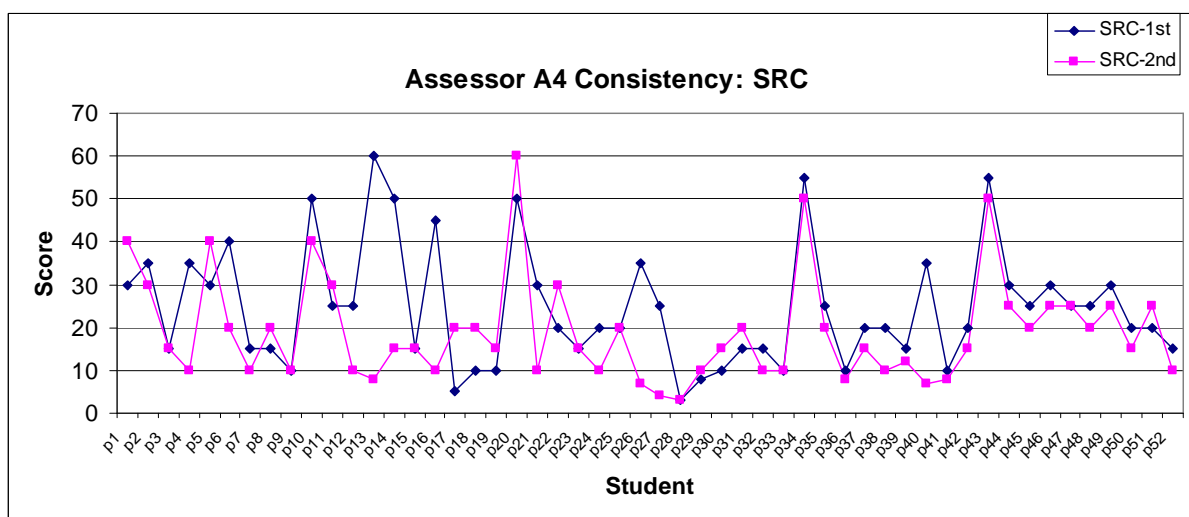


Figure G24: Assessor A4 1st and 2nd assessment scores for the performance factor, SRC

Table G1: Assessor Consistency Correlation factors and p-values

Performance Indicator	A1	A2	A3	A4
pue	$r = 0.53721$, $p = 0.00004$	$r = 0.50878$, $p = 0.000118$	$r = 0.644733$, $p = 0.000000$	$r = 0.33141$, $p = 0.016399$
src	$r = 0.53849$, $p = 0.000038$	$r = 0.67247$, $p = 0.000000$	$r = 0.53225$, $p = 0.000049$	$r = 0.51075$, $p = 0.000110$
wwc	$r = 0.56537$, $p = 0.000013$	$r = 0.84319$, $p = 0.000000$	$r = 0.530401$, $p = 0.000052$	$r = 0.50414$, $p = 0.000139$
mam	$r = 0.56239$, $p = 0.000014$	$r = 0.58962$, $p = 0.000004$	$r = 0.700819$, $p = 0.000000$	$r = 0.41007$, $p = 0.002539$
ier	$r = 0.61670$, $p = 0.000001$	$r = 0.82784$, $p = 0.000000$	$r = 0.645452$, $p = 0.000000$	$r = 0.31272$, $p = 0.024006$
mac	$r = 0.61238$, $p = 0.000001$	$r = 0.82342$, $p = 0.000000$	$r = 0.823206$, $p = 0.000000$	$r = 0.67701$, $p = 0.000000$
pg	$r = 0.61461$, $p = 0.000001$	$r = 0.65969$, $p = 0.000000$	$r = 0.727328$, $p = 0.000000$	$r = 0.35832$, $p = 0.009103$



APPENDIX H: INTER ASSESSOR AGREEMENT

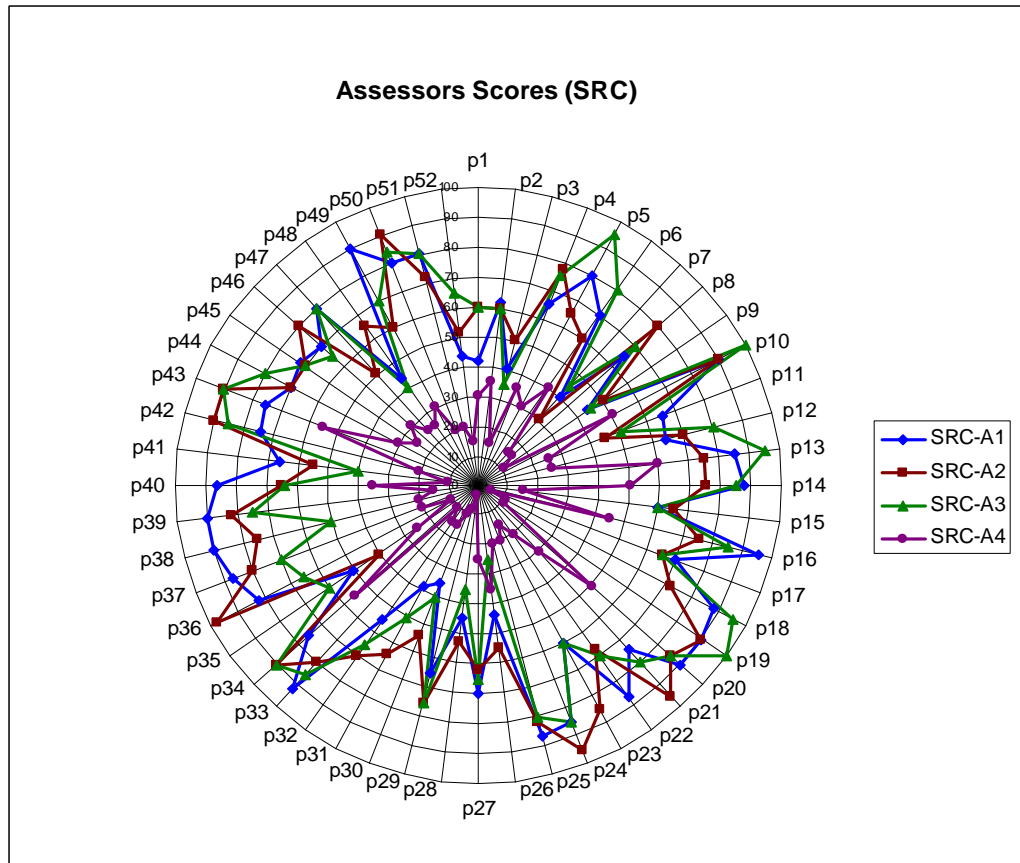


Figure H1: Inter Assessor Agreement: Assessors scores for the performance indicator, *SRC*

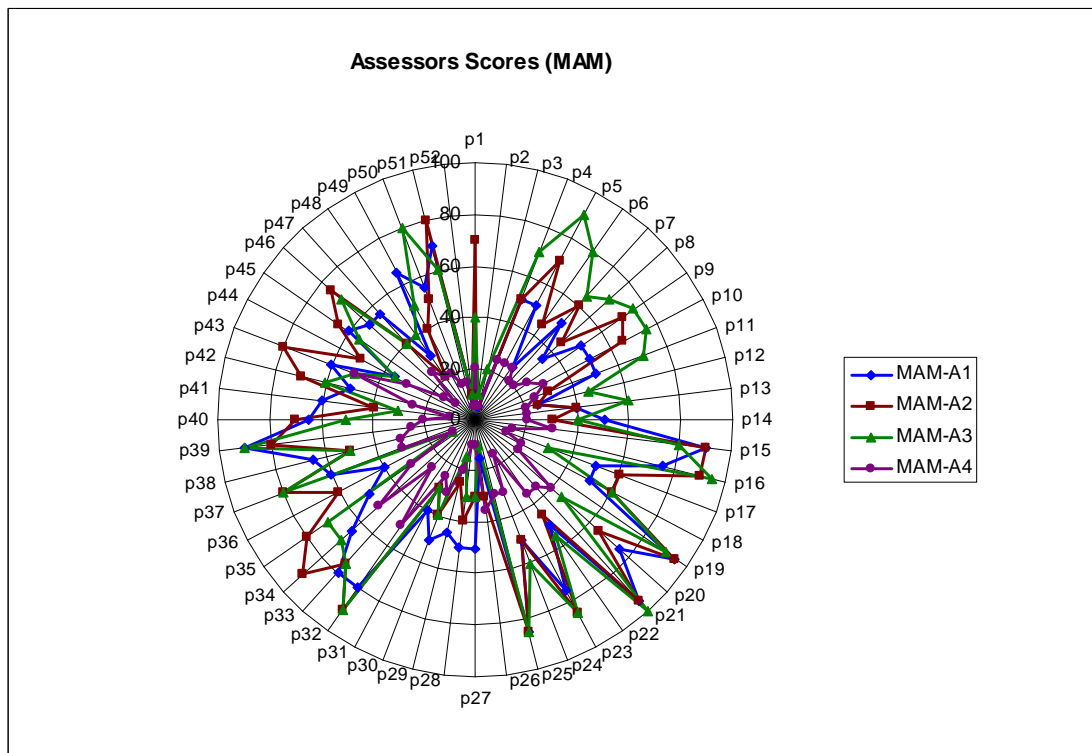


Figure H2: Inter Assessor Agreement: Assessors scores for the performance indicator, *MAM*

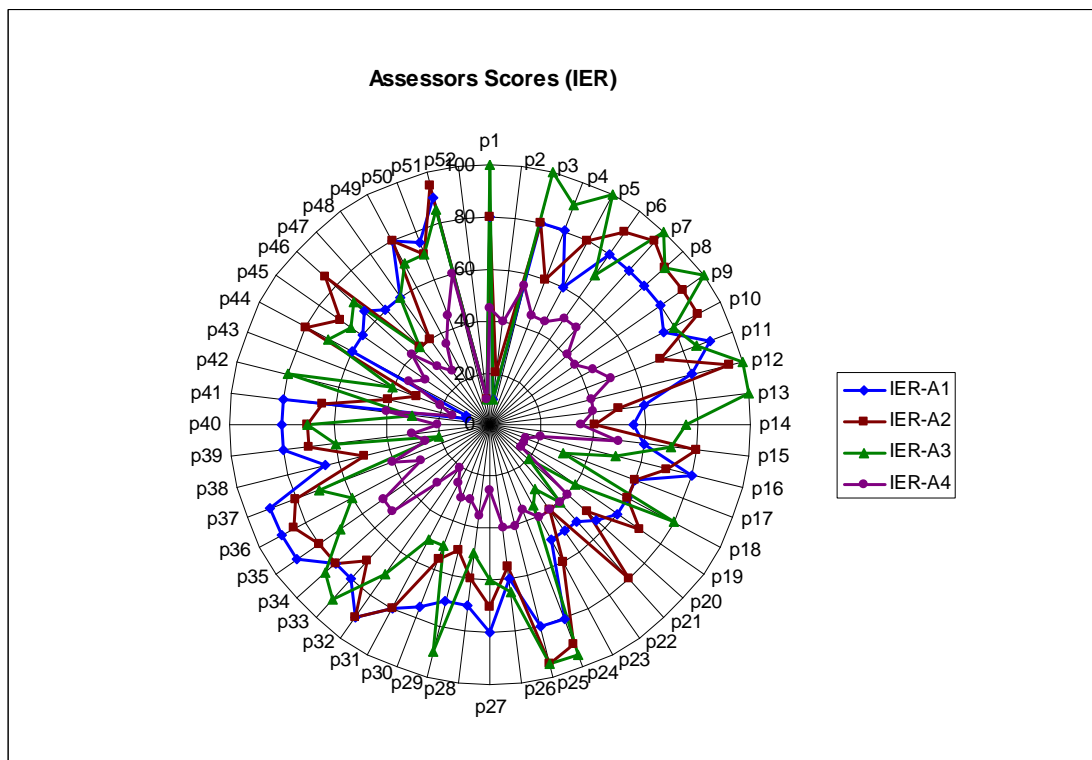


Figure H3: Inter Assessor Agreement: Assessors scores for the performance indicator, *IER*

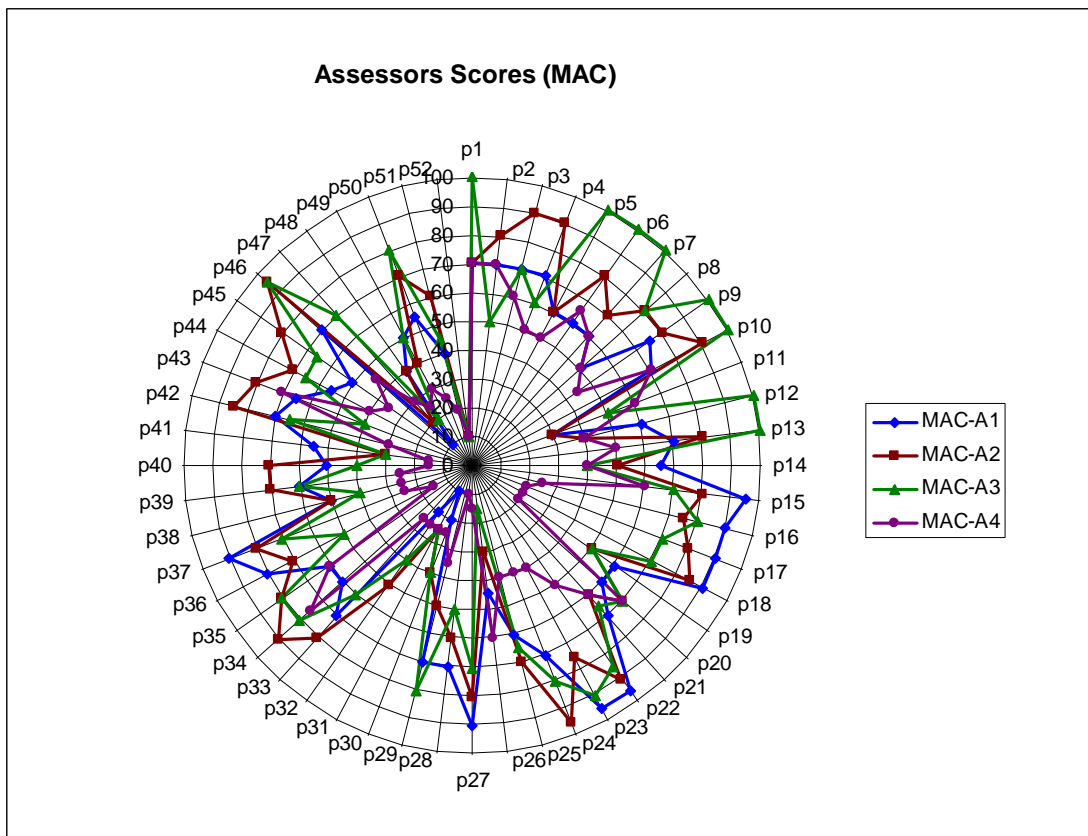


Figure H4: Inter Assessor Agreement: Assessors scores for the performance indicator, *MAC*

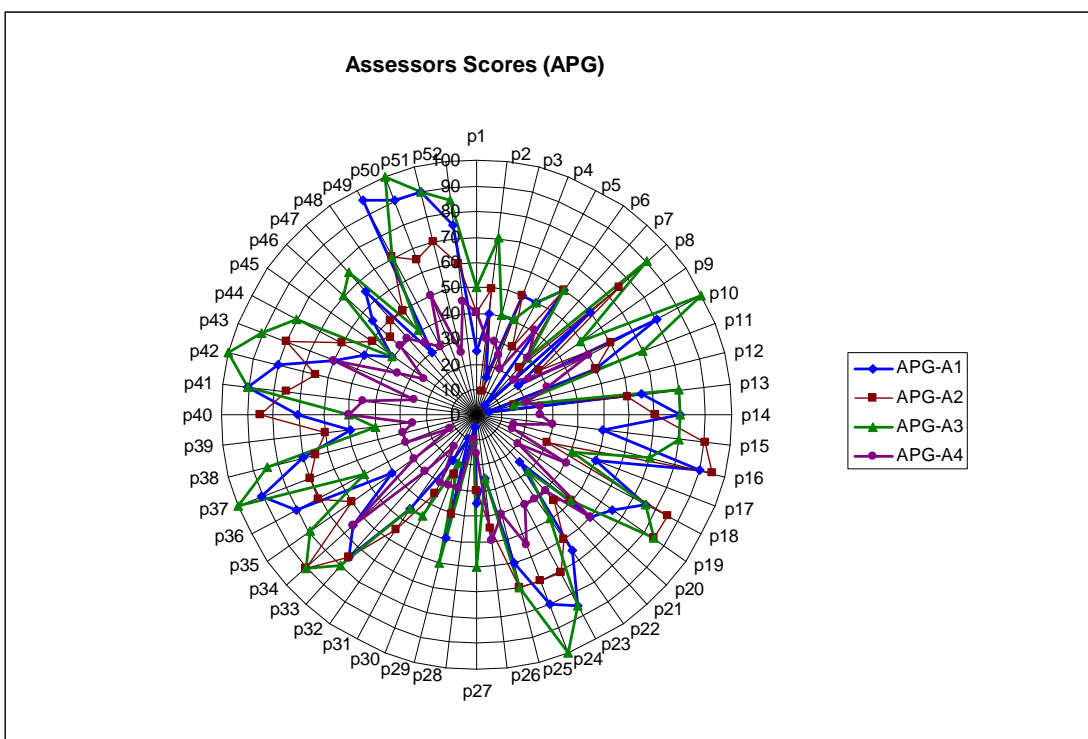


Figure H5: Inter Assessor Agreement: Assessors scores for the performance indicator, *APG*



APPENDIX J: HUMAN ASSESSORS VS MODEL

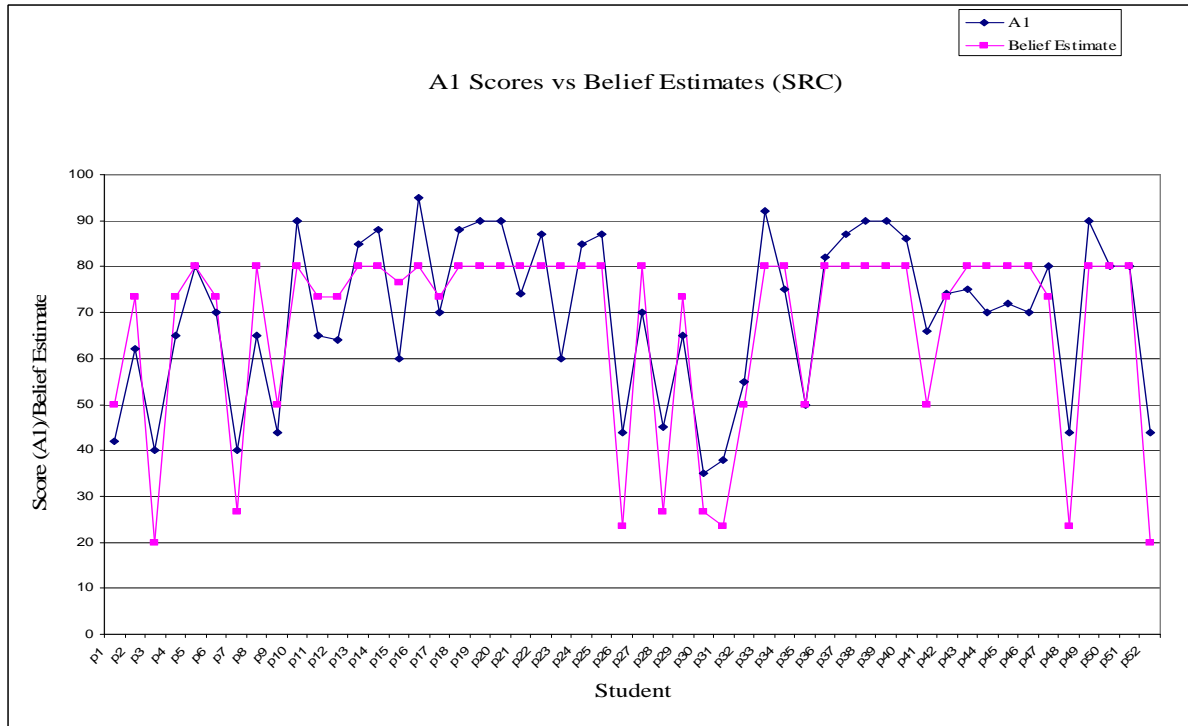


Figure J1: Human-Model Agreement--Assessor A1 scores vs. model belief estimates for SRC.

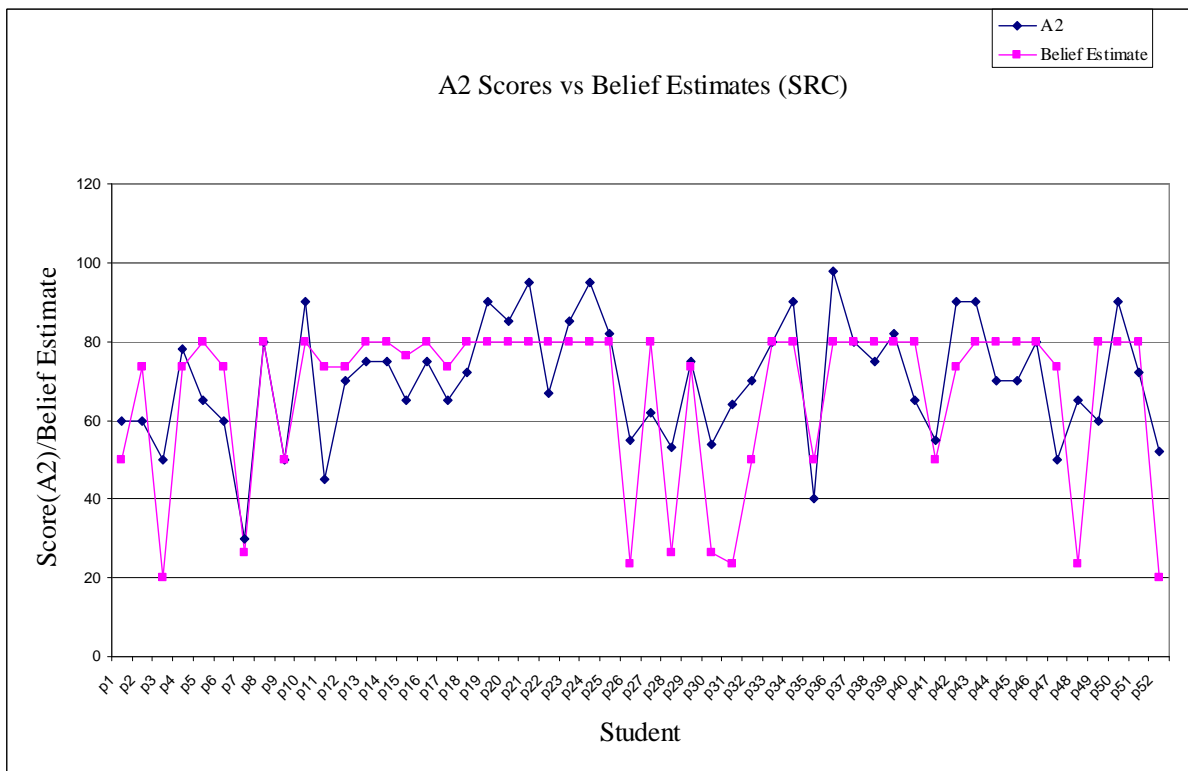


Figure J2: Human-Model Agreement--Assessor A2 scores vs. model belief estimates for SRC.

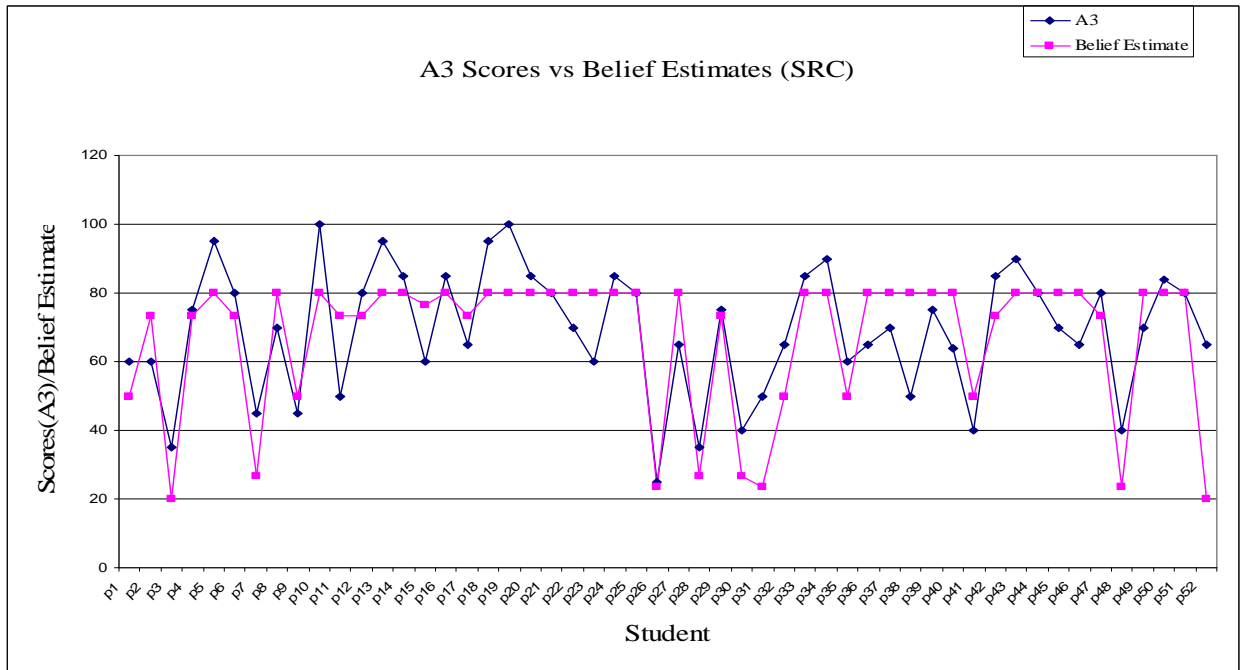


Figure J3: Human-Model Agreement--Assessor A3 scores vs. model belief estimates for SRC.

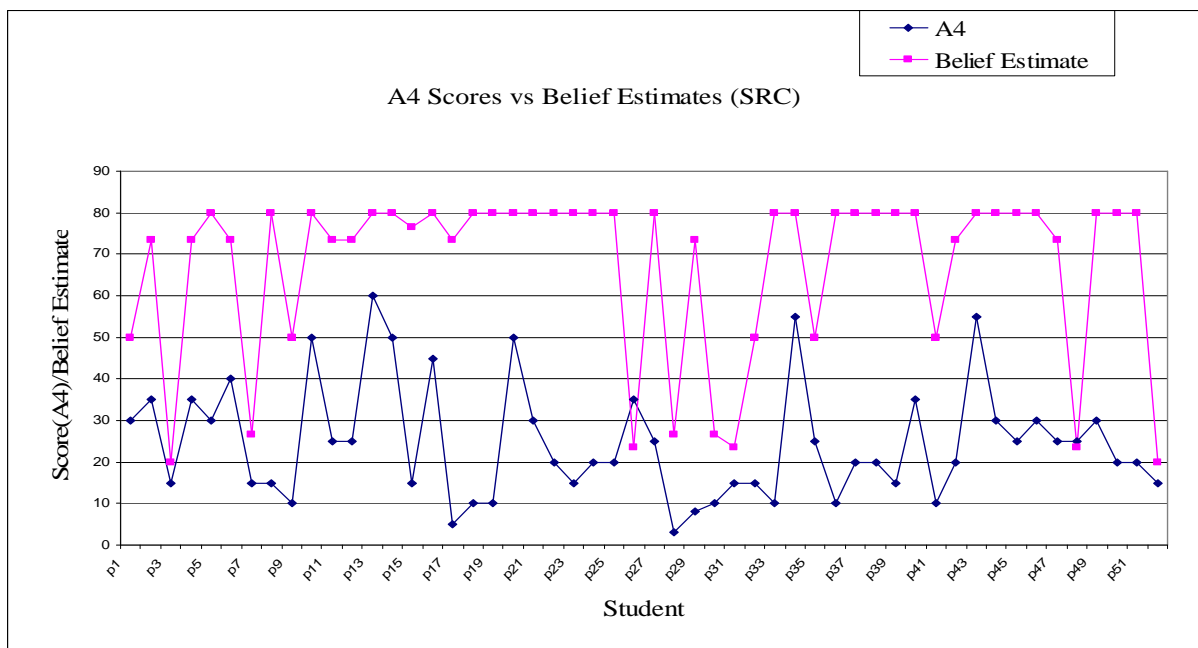


Figure J4: Human-Model Agreement--Assessor A4 scores vs. model belief estimates for SRC.

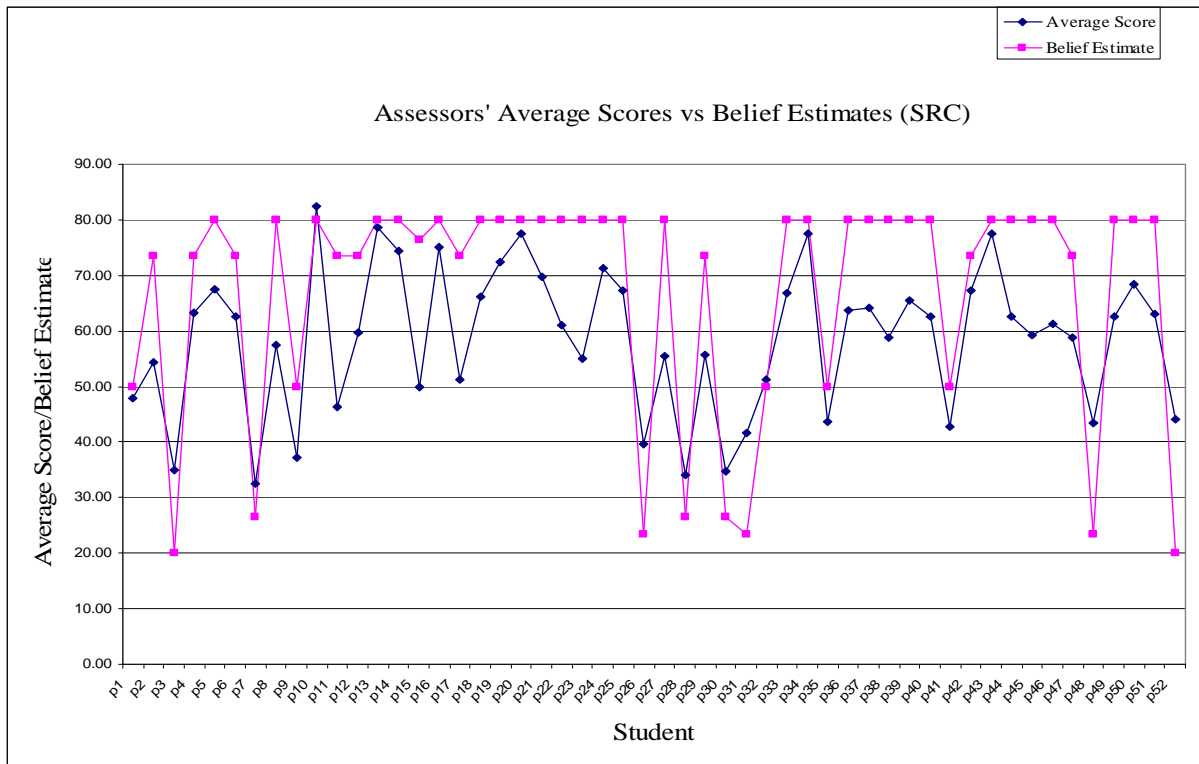


Figure J5: Human-Model Agreement--Assessors average scores vs. model belief estimates for SRC

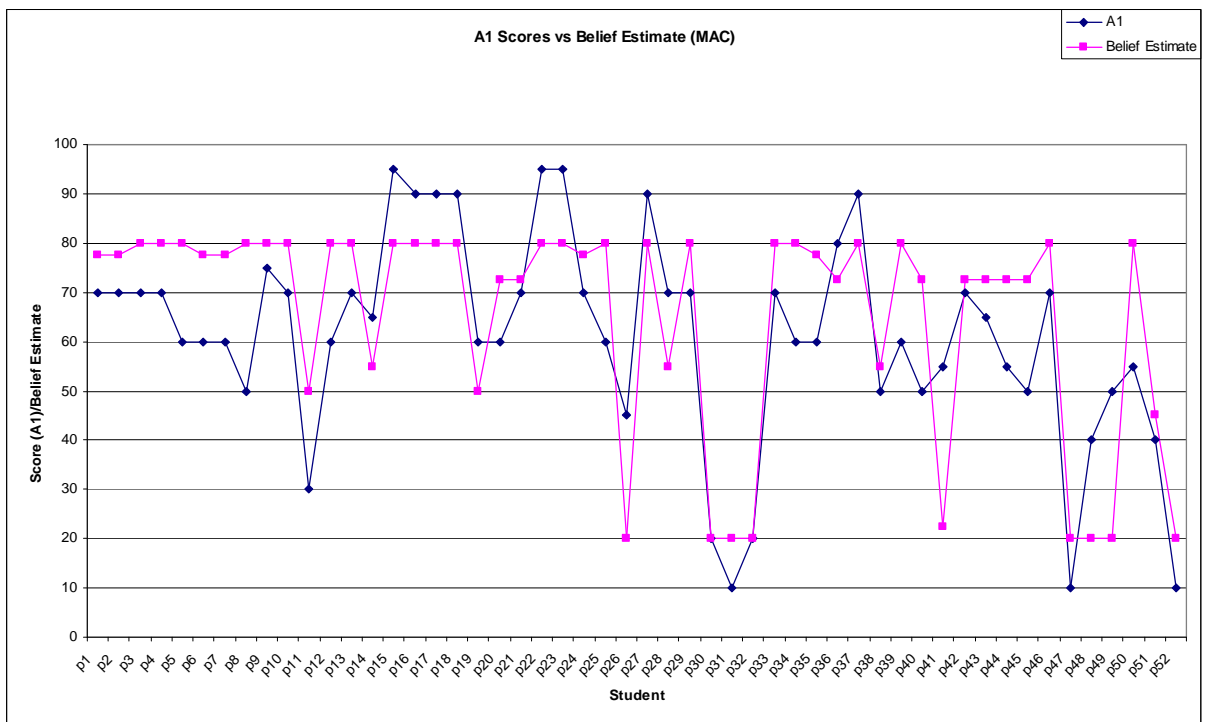


Figure J6: Human-Model Agreement--Assessor A1 scores vs. model belief estimates for MAC.

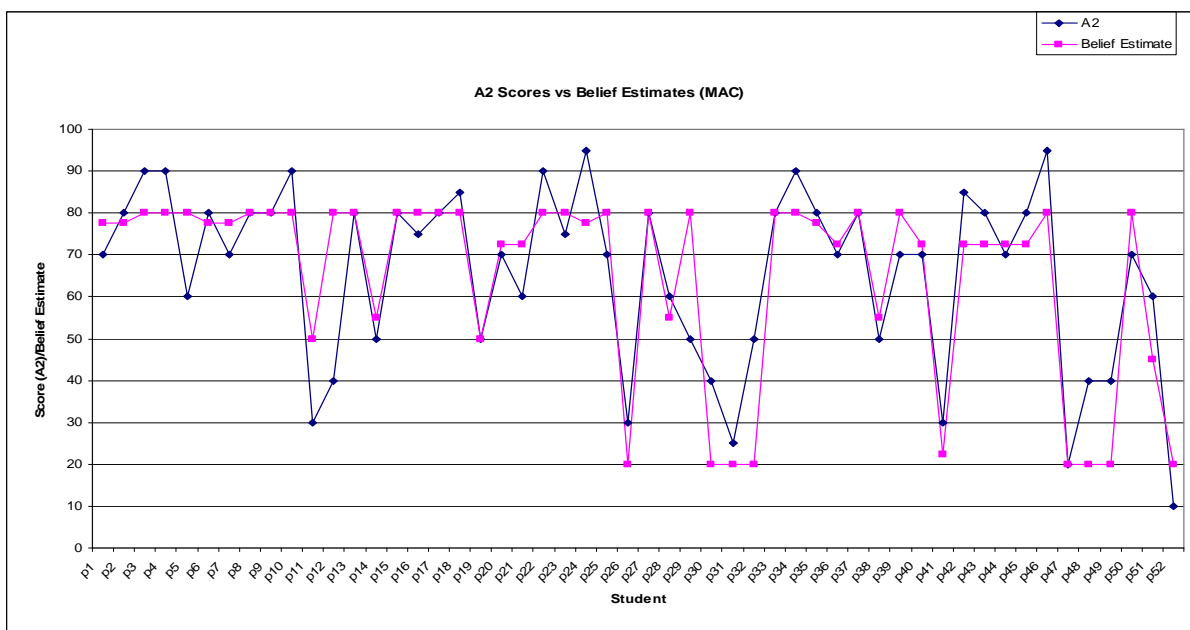


Figure J7: Human-Model Agreement--Assessor A2 scores vs. model belief estimates for *MAC*.

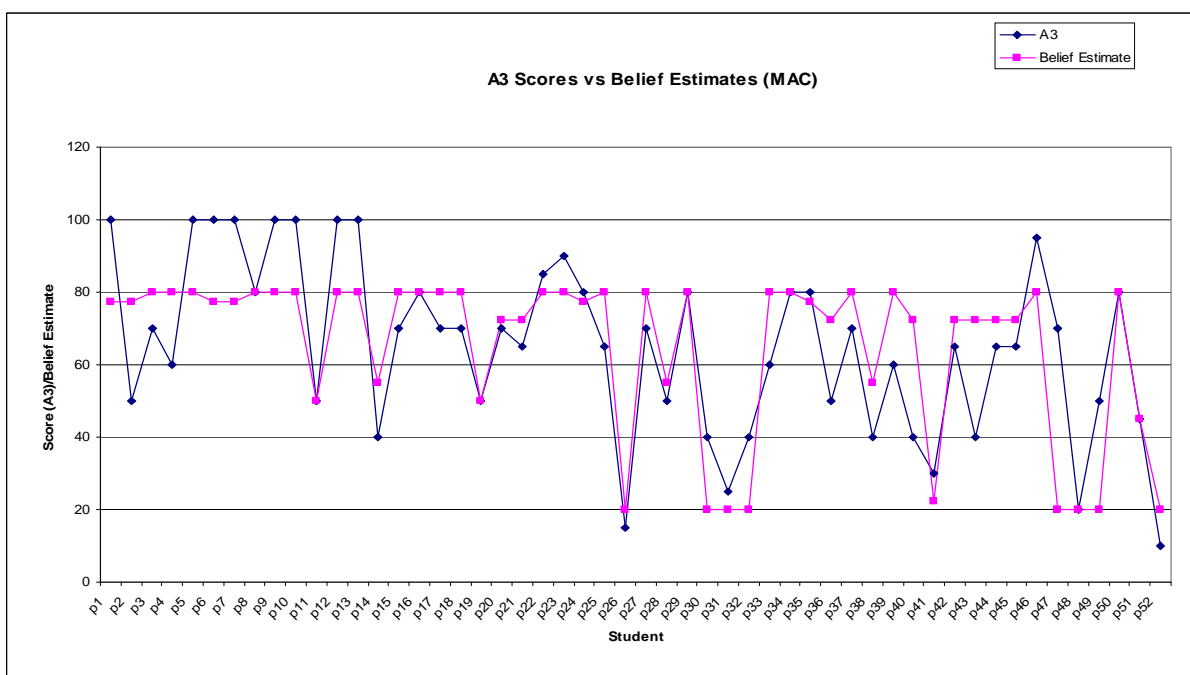


Figure J8: Human-Model Agreement--Assessor A3 scores vs. model belief estimates for *MAC*.

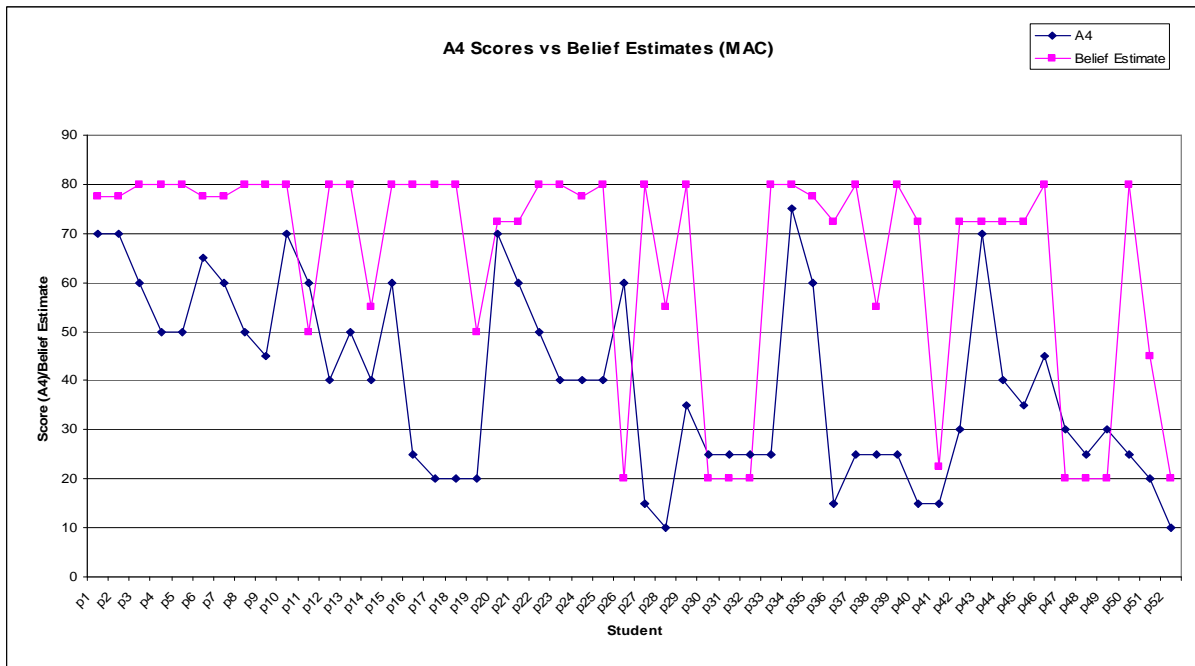


Figure J9: Human-Model Agreement--Assessor A4 scores vs. model belief estimates for *MAC*.

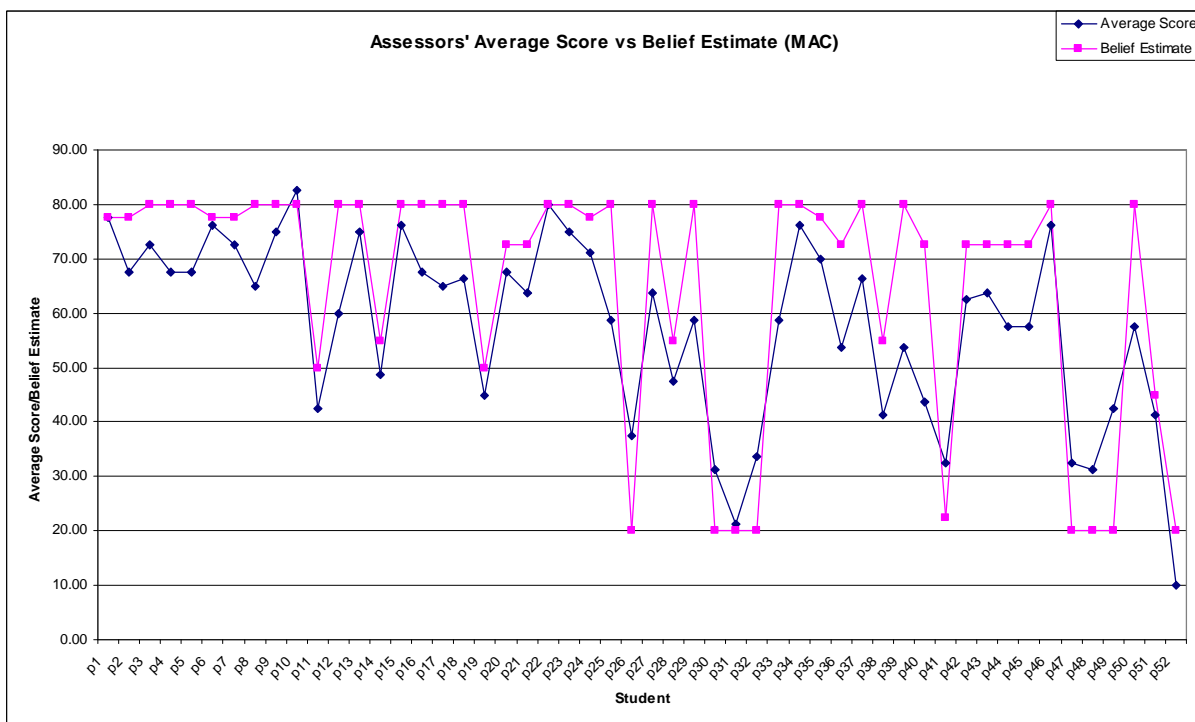


Figure J10: Human-Model Agreement--Assessors average scores vs. model belief estimates for *MAC*