

A SMARTPHONE-BASED ADAPTIVE LEARNING APPROACH TO GAMIFIED CLINICAL TRAINING IN LOW-INCOME COUNTRIES



Timothy Tuti Ng'ang'a

Kellogg College,
University of Oxford

Supervisors: **Prof Niall Winters** and **Dr Chris Paton**

*Thesis submitted in partial fulfilment of the requirements for the award of the
degree of Doctor of Philosophy*

Trinity Term 2020

Thesis Abstract

Of the estimated 2.9 million neonatal lives that are lost each year globally, Sub-Saharan Africa (SSA) has the highest overall risk of death within the first 24 hours of life, accounting for 37% of global deaths. Severe workforce shortages, coupled with a health workforce skill imbalance and maldistribution, and a lack of training opportunities are likely to be key contributors to this high mortality rate. Given the lack of capacity to run face-to-face training at scale, new digital approaches that leverage the growing ubiquity of smartphone devices are being considered for training across SSA's health systems. However, little evidence exists on the implementation of digital learning interventions that are: (a) relevant to the context of SSA; (b) consider healthcare providers' diverse initial and continuing clinical training needs; and (c) adapt learning content as learners continue to develop their knowledge. This is despite research – from non-clinical, high-resource contexts – showing that digital learning platforms that use either gamified approaches or adapt to learner needs have a significant positive effect on learning outcomes.

This thesis aims to study how gamified approaches to smartphone-based learning can be configured to support scaffolded Self-Regulated Learning (SRL) for clinical training in low-income countries (LICs). Through studying the contextual, self-directed learning needs of healthcare providers, this thesis builds a systematic understanding of the nature and pattern of use of gamified digital learning interventions and their impact on clinical learning outcomes.

The research questions addressed are: (a) What is the effectiveness of offering adaptive versus standardised feedback on learning gains on neonatal care emergency scenarios?; (b) What is the effect learner characteristics and SRL behaviours had on individualised learning gains?; and (c) What are the knowledge tracing approaches that are most suited for this context and subject domain?; (d) What are the learning experiences and perceptions of healthcare providers regarding gamified smartphone-based learning?, and (e) How do these learning experiences influence their SRL efforts? The methodology used to answer these questions was divided into three phases incorporating semi-structured in-depth interviews, Nominal Group Technique, survey methodology and a randomised experiment. The target population was healthcare providers of all cadres from low-income countries, from students to consulting specialist physicians.

The findings detail how adaptive learning can be designed and implemented to leverage the SRL behaviours of healthcare providers. They shed light on why within-platform learning feedback (that elaborates on the consequences of [in]action or erroneous action) is preferred to reflective feedback, and illuminate clinicians' expectations of learning experiences. Overall, the findings suggest that digital learning platforms should help to build a learner's confidence in the knowledge gained, so as it can be applied in their routine clinical setting. By doing so, future improvements in clinical outcomes in LIC health systems may be realised as part of holistic workplace capacity development strategies.

Acknowledgements

I must first thank my supervisors – Prof. Niall Winters and Dr Chris Paton for their encouragement and guidance throughout the DPhil degree. Undertaking a DPhil degree would not have been possible without the Oxford studentship fully funded by the Economic and Social Research Council.

Furthermore, I am grateful to the Kenya Medical Research Institute- Wellcome Trust Research Programme (through Prof. Mike English) for allowing me to use their professional networks and resources during the conduct of this thesis.

I would like to thank my friends from St. Ebbe's Church, especially those from Thesis Table 4, and my friends from the Madaraka Real Group in Nairobi, especially those from the Waiyaki Way branch. Their encouragement and support have been pivotal to the completion of my DPhil studies.

Finally, special recognition goes to some of my closest friends during my time in the UK: Christoph and Stephanie Schott, Lauren Delapenha, Rick and Sarah Lange, Matt Cherukara and LaRae Ferguson. They made the UK feel a bit like home, reminiscent of *the city under the sun*.

“...⁸For we do not want you to be unaware, brothers and sisters, of the affliction we experienced in [Asia]. For we were so utterly burdened beyond our strength that we despaired of life itself. ⁹Indeed, we felt that we had received the sentence of death. But that was to make us rely not on ourselves but on God who raises the dead. ¹⁰He delivered us from such a deadly peril, and he will deliver us. On Him we have set our hope that he will deliver us again...”

Publications

Three articles on the thesis methodology and results chapters have been peer-reviewed and published as journal articles, book chapters or conference proceedings (and three more articles are currently under review). The data used in these publications was entirely the product of the thesis study, and the analysis presented in the thesis was done by me, the student. The corresponding references to the in-review and published works are as follows:

1. Tuti, T., Paton, C., & Winters, N. (2020, March). Learning to represent healthcare providers knowledge of neonatal emergency care: findings from a smartphone-based learning intervention targeting clinicians from LMICs. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge* (pp. 320-329).
2. Tuti, T., Winters, N., Muinga, N., Wanyama, C., English, M., & Paton, C. (2019). Evaluation of Adaptive Feedback in a Smartphone-Based Serious Game on Health Care Providers' Knowledge Gain in Neonatal Emergency Care: Protocol for a Randomized Controlled Trial. *JMIR research protocols*, 8(7), e13034.
3. Tuti, T., Paton, C., English, M., & Winters, N. (2019, September). Building a Learner Model for a Smartphone-Based Clinical Training Intervention in a Low-Income Context: A Pilot Study. In *European Conference on Technology Enhanced Learning* (pp. 55-68). Springer, Cham.
4. Tuti, T., Winters, N., Edgcombe, H., Muinga, N., Wanyama, C., English, M., & Paton, C. (2020). Evaluation of Adaptive Feedback In A Smartphone-Based Serious Game On Healthcare Providers' Learning Gain: A Randomised Experiment. *J Med Internet Res*. 2020 Apr 8. DOI: 10.2196/17100.
5. Tuti, T., Paton, C., & Winters, N. (2020). The Counterintuitive Self-Regulated Learning Behaviours Of Healthcare Providers From Low-Income Settings. *[Manuscript in Review]*.
6. Tuti, T., Muinga, N., Wanyama, C., English, M., Nagraj, S., Paton, C., & Winters, N. (2020). Gamified Smartphone-Based Self-Regulated Learning Experiences Of Healthcare Providers. *[Manuscript in Review]*.

Contents

Thesis Abstract	i
Acknowledgements	ii
Publications	iii
List of Figures	ix
List of Tables	xi
List of Abbreviations	xii
1 INTRODUCTION	1
1.1 Adaptive learning in low-income countries	1
1.2 Clinical learning needs in Sub-Saharan Africa	2
1.3 The learning platform used in this thesis	5
1.3.1 The user journey in the LIFE gamified smartphone application	6
1.4 Thesis contribution	11
1.5 Thesis Outline	13
2 LITERATURE REVIEW	17
2.1 Educational perspectives to adaptive learning and gamified learning	17
2.1.1 Media perspectives to gamified learning	18
2.1.2 Clinical perspectives on digital-based training and gamified learning	26
2.2 Approaches to adaptive learning on digital platforms	31
2.3 Pathways of adaptive instruction – How the source is translated into target	34
2.4 Summary	37
3 THEORETICAL BACKGROUND	38
3.1 Constructivist-based learning and cognition in clinical education	39
3.2 Theorising learning relative to instructional design	40
3.3 Self-Regulated Learning theory and adaptive learning on digital platforms	42
3.4 Approaches to measuring Self-Regulated Learning in adaptive platforms	50
3.4.1 Bayesian Knowledge Tracing Details	53
3.4.2 Additive Factor Models (AFMs) and Performance Factor Models (PFMs)	55
3.4.3 Knowledge Tracing and Item Response Theory	58
3.5 Scaffolding instructional support in Self-Regulated Learning	58
3.5.1 Zone of Proximal Development, the flow channel and adaptive learning	58
3.5.2 Motivation within SRL-driven feedback scaffolding	61
3.6 Summary of literature and theoretical review	63
3.7 Justification of research	65
3.8 Research questions	68

4	METHODOLOGY	71
4.1	Overall methodological design	71
4.2	Study participants and settings	75
4.3	Participant timeline	75
4.4	Ethical considerations	75
4.5	Student-step data collected from LIFE application	77
4.6	Summary of methodology	78
5	EFFECT OF ADAPTIVE FEEDBACK ON HEALTHCARE PROVIDERS SELF-REGULATED LEARNING GAINS: A RANDOMISED EXPERIMENT	81
5.1	Recap of Objectives	81
5.2	Methodology	81
5.2.1	Study design	82
5.2.2	Eligibility criteria	82
5.2.3	Study setting and recruitment	82
5.2.4	The intervention	83
5.2.5	Implementation of BKT mechanism for adaptive feedback	88
5.2.6	Outcomes	92
5.2.7	Participant timeline	94
5.2.8	Sample size calculation	95
5.2.9	Randomisation	96
5.2.10	Statistical methods and analysis	97
5.2.11	Data Management	98
5.2.12	Informed Consent	98
5.3	Results	98
5.3.1	Baseline characteristics of the participants	98
5.3.2	Effect of adaptive feedback on learning gains across study arms	102
5.3.3	Effect of adaptive feedback on individual learning gains	103
5.4	Summary of findings	109
6	SELF-REGULATED LEARNING BEHAVIOURS OF HEALTHCARE PROVIDERS FROM LOW-INCOME COUNTRIES:	110
6.1	Recap of Objectives	110
6.2	Methodology	111
6.2.1	Study context, participants, and procedure	111
6.2.2	Study instruments, variables, and data management	113
6.2.3	Statistical analysis	113
6.2.4	Post-hoc analysis	115

6.2.5	Summary.....	115
6.3	Results.....	116
6.3.1	Baseline characteristics of the participants.....	116
6.3.2	Summary of OSLQ tool responses	118
6.3.3	Latent Profile Analysis (LPA) Results	122
6.3.4	Influence of SRL strategies on healthcare providers' learning gains.....	127
6.4	Summary of findings	130
7	SELF-REGULATED LEARNING PROCESS OF HEALTHCARE PROVIDERS FROM LICS	132
7.1	Recap and Objectives	132
7.2	Methodology	133
7.2.1	Study design, setting and participants	133
7.2.2	Study Variables.....	133
7.2.3	Statistical Methods, Missing Data, And Sensitivity Analyses	134
7.2.4	Summary.....	135
7.3	Results.....	135
7.3.1	Summary of LIFE play game statistics.....	135
7.3.2	Model Performance: AFM versus PFM	137
7.3.3	Moderation effect of level of feedback given previous learning opportunity.....	138
7.3.4	Model diagnostics	143
7.4	Summary of findings	144
8	LEARNING TO REPRESENT HEALTHCARE PROVIDERS KNOWLEDGE OF NEONATAL EMERGENCY CARE.....	146
8.1	Learning to represent healthcare providers knowledge of neonatal emergency care: Methodology	146
8.1.1	Prior work on knowledge tracing.....	146
8.1.2	Study Design, Setting And Participants	152
8.1.3	Study Variables.....	153
8.1.4	Model Training	153
8.1.5	Model Evaluation	156
8.1.6	Statistical Methods, Missing Data, And Sensitivity Analyses	157
8.1.7	Summary.....	159
8.2	Results.....	159
8.2.1	Baseline characteristics of this study participants and variables.....	159
8.2.2	Summary of DKT model performance	160
8.2.3	DKT model accuracy by learning spacing, healthcare provider features and overall performance	163

8.2.4	Adaptive feedback’s impact on knowledge mastery and wheel-spinning.....	166
8.3	Comparison of knowledge tracing approaches so far.....	169
8.4	Summary of findings	171
9	EXPLORING HEALTHCARE PROVIDERS PERCEPTIONS OF GAMIFIED SELF-REGULATED LEARNING EXPERIENCES	173
9.1	Recap and Objectives	173
9.2	Methodology	174
9.2.1	Study design, setting, and participants and timelines	174
9.2.2	Data collection and analysis procedures	176
9.2.3	Summary.....	177
9.3	Results.....	178
9.3.1	Baseline characteristics of the qualitative study stages	178
9.3.2	Results from Nominal Group Technique.	178
9.3.3	Learning task definitions, elaboration and performance imperatives.....	181
9.3.4	Context of practice, cognitive adaptations and goal setting	183
9.3.5	Feedback and motivation, and transfer to practice.....	186
9.3.6	Game experience and engagement.....	188
9.4	Summary of findings	190
10	DISCUSSION	192
10.1	Implications of findings.....	192
10.2	Summary of the key findings	200
10.3	Relation to published literature and studies.....	202
10.3.1	Experimental findings	202
10.3.2	Self-regulated learning of LICs healthcare providers on digital platforms.....	203
10.3.3	Effect of adaptive feedback on practice opportunities	204
10.3.4	Knowledge tracing for clinical learning interventions	205
10.3.5	Qualitative exploration healthcare providers perceptions on gamified self-regulated learning on smartphones	206
10.4	Thesis strengths and limitations	207
11	CONCLUSIONS	211
	REFERENCES.....	214
	APPENDIX	230
A.	Appendix A – Chapter 1.....	230
A.1	Knowledge Components, Student-Step and Opportunities	230
B.	Appendix B – Chapter 2.....	233
C.	Appendix C – Chapter 3.....	237

C.1	Constructivist approaches to education.....	237
C.2	The Bayesian Knowledge Tracing Model.....	238
D.	Appendix D – Chapter 4	241
D.1	Ethics Approvals	241
E.	Appendix E – Chapter 5.....	243
E.1	Randomisation algorithm	243
F.	Appendix F – Chapter 6.....	245
F.1	Online Self-regulated Learning Questionnaire (OSLQ) Tool	245
F.2	SRL Confirmatory Factor Analysis Model	247
G.	Appendix G – Chapter 7	249
G.1	A primer into cross-validation	249
H.	Appendix H – Chapter 8	251
I.	Appendix I – Chapter 9.....	253
I.1	Nominal Group Technique Data Collection Procedure	253
I.2	Guide For the Brief Open-Ended Semi-Structured Interview (Stage 2).....	254
I.3	Guide For the Open-Ended Semi-Structured Interview (Stage 3).....	255

List of Figures

Figure 1.3.1: LIFE start screen	6
Figure 1.3.2: Screenshot of example of scenario selection in LIFE	7
Figure 1.3.3: Navigation task provided with scenario background information	7
Figure 1.3.4: Navigation interaction using on the smartphone device	8
Figure 1.3.5: Multiple choice question illustration	8
Figure 1.3.6: Learning task assessing student's ability to perform bag-mask ventilation	9
Figure 1.3.7: Screenshot of example error feedback window in LIFE application	9
Figure 1.3.8: Learner's score card	10
Figure 2.1.1: Conceptual model of digital-game based learning	21
Figure 2.1.2: Flow channel as a representation of skill-versus-challenge learning maximisation process (Borrowed from Kiili et al., 2012)	22
Figure 2.2.1: The tripartite structure of adaptive instruction (M. Vandewaetere et al., 2011)	33
Figure 3.3.1 Winne and Hadwin SRL model (Winne & Hadwin, 1998)	46
Figure 3.4.1: Bayesian Knowledge Tracing model illustration	54
Figure 3.5.1: Preliminary conceptual model of adaptive self-regulated learning through LIFE platform	63
Figure 4.1.1: Overall study phases	73
Figure 4.1.2: LIFE components distinguishing those implemented independently and specifically to allow for data collection for this DPhil project from the original platform	74
Figure 4.5.1: Data collection procedure within LIFE.	78
Figure 5.2.1: The flow of feedback provision	87
Figure 5.2.2: Example of BKT learning trajectories from a single Knowledge Component (KC) for five different healthcare providers	90
Figure 5.3.1 Experiment's enrolment flow-chart	99
Figure 5.3.2: Change of student learning gains over learning sessions by study arm	107
Figure 5.3.3: Change of student learning gains over successive learning sessions by spacing options	107
Figure 6.3.1: SRL Item pairwise correlation representation.	121
Figure 6.3.2: Self-Regulated Learning profiles of healthcare providers in LICs based expected item means from OSLQ tool	123
Figure 6.3.3: SRL skills by learning platform	124
Figure 6.3.4: Learning gains of healthcare providers by SRL profiles	128
Figure 6.3.5: SRL profiles by clinician training status	129
Figure 7.3.1: PFM Learning curves of feedback effect by self-regulated spacing of learning repetition	140
Figure 7.3.2: PFM Learning curves by SRL profile	141
Figure 7.3.3: Pairwise correlation between SRL skills. (**) represents statistical significance. SRL skills are defined in detail in Chapter 3.3.1	142
Figure 7.3.4: Model diagnostic plots for the PFM model	143
Figure 8.1.1: LSTM memory cell illustration	149
Figure 8.1.2: LIFE's proposed DKT model	150
Figure 8.1.3: Architecture of the proposed Autoencoder	151
Figure 8.1.4: Illustration of LSTM Autoencoder model structure with hidden layers structure	155
Figure 8.1.5: Sliding window sample generation based on varying length of timesteps for a single learner	157
Figure 8.1.6: The DKT prediction setup for LIFE	158

Figure 8.2.1: DKT model performance over length of sequence	161
Figure 8.2.2: Visualisation of knowledge representation in the encoding layer of the DKT Auto-Encoder model	162
Figure 8.2.3: DKT model prediction accuracy across student performance and spacing	164
Figure 8.2.4: Prediction accuracy of performance by clinical role and years of experience	165
Figure 8.2.5: Prediction accuracy of the outcome of the next learning task based on SRL profile and level of training	166
Figure 8.2.6: Knowledge mastery by healthcare provider role	167
Figure 8.2.7: Wheel spinning by healthcare provider role	168
Figure 8.2.8: Knowledge decay rate by healthcare provider role	168
Figure 9.3.1: Illustration of the perceptions of self-regulated learning on gamified scenario-based training on digital platforms of the study participants	190
Figure A.1: Learning content representing clinical guidelines that need follow	231
Figure C.1 Corbett and Anderson Bayesian Knowledge Tracing model	238
Figure E.1: Random allocation procedure	243
Figure E.2: Random allocation ratio based on time of day (the seed used for the generator)	244
Figure F.1: SRL configuration for CFA model	247
Figure F.2: Three-profile model of SRL	248
Figure G.1: Illustration of cross-validation	249
Figure G.2: Different configuration of cross-validation	250

List of Tables

Table 3.3.1: Common levels of feedback within interactive digital learning platforms, <i>borrowed from (Holmes et al., 2015)</i>	48
Table 5.2.1: Feedback content provided in LIFE smartphone application for Neonatal Resuscitation Scenario A Training for the individual learning tasks.....	84
Table 5.2.2: Summary statistics of pilot data from LIFE game play	89
Table 5.2.3: BKT model parameters generated	90
Table 5.3.1: Tests of whether differences in allocation rate between study arms resulted in non-equivalent LIFE app use for all study participants included	100
Table 5.3.2: Baseline characteristics of study participants where demographic data is available.....	101
Table 5.3.3: Tests of whether dropout rate is different between study arms using baseline characteristics of study participants where demographic data is available.....	101
Table 5.3.4: Performance of learners within study arms	102
Table 5.3.5: Number of feedback messages provided.....	103
Table 5.3.6: Learning intervention effect on individual healthcare providers normalised learning gains	104
Table 5.3.7: Multicollinearity tests for the learning gain outcome	106
Table 5.3.8: Learners in study arms over each iteration	106
Table 5.3.9: Learning analysis of learning gain using LIFE	108
Table 6.3.1: SRL survey participants by platform.....	116
Table 6.3.2: Baseline characteristics of survey respondents	117
Table 6.3.3: Descriptive Statistics of SRL Questionnaire Tool Items.....	118
Table 6.3.4: Latent profile analysis results for Self-Regulated Learning.....	122
Table 6.3.5: Logistic regression results predicting profile membership	126
Table 6.3.6: Self-regulated learning skills by profiles	127
Table 7.3.1: Summary statistics of data from LIFE game play sessions	136
Table 7.3.2: Results from PFM model of probability of getting the attempt correct on first try based on learning opportunity	138
Table 7.4.1: Adaptive feedback effect on learning gain comparison	145
Table 8.2.1: Learners excluded due to insufficient timestep data needed for the DKT model.....	159
Table 8.3.1: Model Metrics Comparing Different Knowledge Tracing Approaches	170
Table 9.2.1: Summary of the Qualitative Research Stages	175
Table 9.3.1: Summary characteristics of study participants.....	178
Table 9.3.2: Summary of design components derived through consensus from item - generation phase of the Nominal Group Technique.....	179
Table 9.3.3: Weighted rank of items generated from the voting phase of Nominal Group Technique	180
Table B.1: Summary of common learning theories used in the implementation of Intelligent Tutoring Systems.	233
Table E.1: Participants breakdown by region	244
Table E.2: Learning spacing between first two sessions, N (%)	244
Table F.1: SRL profiles by digital platform	248
Table H.1: Learning Spacing By Running Learner Performance.....	251
Table H.2: Multivariate MANCOVA analysis	251
Table H.3: Univariate MANCOVA analysis	252

List of Abbreviations

AIED	Artificial Intelligence in Education
AFMs	Additive Factor Models
ANN	Artificial Neural Network
ANOVA	Analysis of Variance
AUC	Area Under a Curve
BGN	Bayesian Graph Network
BKT	Bayesian Knowledge Tracing
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
CME	Continuous Medical Education
DKT	Deep Knowledge Tracing
ETAT+	Emergency Triage, Assessment and Treatment plus admission care
IRT	Item Response Theory
ITSs	Intelligent Tutoring Systems
KCs	Knowledge Components
LDHF	Low-Dose-High-Frequency
LFA	Learning Factor Analysis
LICs	Low-Income Countries
LIFE	L ife-saving I nstructions F or E mergencies
LPA	Latent Profile Analysis
LSTMs	Long-Short Term Memory
KEMRI	Kenya Medical Research Institute
NGT	Nominal Group Technique
OSLQ	Online Self-Regulated Learning Questionnaire
PFMs	Performance Factor Models
RMSEA	Root Mean Square Error of Approximation
RNN	Recurrent Neural Network
ROC	Receiver Operating Curve
SRL	Self-Regulated Learning
SRMR	Standardised Root Mean Square Residual
SSA	Sub-Saharan Africa
TLI	Tucker-Lewis Index
VR	Virtual Reality
ZPD	Zone of Proximal Development

1 INTRODUCTION

1.1 Adaptive learning in low-income countries

Adaptive learning can be defined as a range of learning experiences, instructional approaches, and academic support strategies designed to address the specific learning needs of individual students (Holmes, Anastopoulou, Schaumburg, & Mavrikis, 2018). There is a scarcity of evidence on the use of computer-based adaptive learning interventions for learning in low-income countries (LICs) like those in Sub-Saharan Africa(SSA) (Conn, 2017). This is despite evidence from non-clinical, high-resource contexts that demonstrates that computer-based adaptive learning can significantly outperform (1) teacher-led large-group instruction, (2) non-adaptive computer-based instruction, and (3) paper-based instruction in producing learning gains (Ma, 2017; Ma, Adesope, Nesbit, & Liu, 2014; WHO, 2011). However, research into learning on digital platforms in settings such as SSA has been limited and has tended to focus on the following areas:

- Sustainability and relevance of platforms and how to integrate local decisions around education and local context (S. Barteit et al., 2015; Long, Pariyo, & Kallander, 2018; Nye, 2015),
- How to increase access and awareness of platforms for the intended recipients (S. Barteit et al., 2015),
- Increasing the quality of learning content and adapting available learning materials to the local context (S. Barteit et al., 2018)
- Usability challenges for digital learning interventions while failing to consider infrastructural and socio-cultural and economic contexts (S. Barteit et al., 2018; Bediang, Bagayoko, Raetzo, & Geissbuhler, 2011; Bediang et al., 2013; Nye, 2015).

This is posited to have resulted in a situation where the available research evidence lacks appropriate representation from low-income contexts, reflecting wider socio-cultural imbalances in Artificial Intelligence in Education (AIED) research production (Blanchard, 2015; Conn, 2017).

On the other hand, opportunities for adaptive learning in LICs are becoming increasingly possible primarily due to growth in low-end smartphone ownership (Nye, 2015; Poushter, 2016; Silver & Johnson, 2018). Thus, mobile-based, context-relevant learning interventions that aim to evaluate novel instructional strategies are advocated as a step forward for adaptive learning in LICs (Conn, 2017; Nye, 2015). This step forward still remains largely unexplored to date but is nevertheless still warranted (Blanchard, 2015; Buliva, 2019; Nye, 2015).

It is crucial to consider the educational context when evaluating barriers to adoption of adaptive learning approaches on digital platforms; if progress is to be made, the subject-domain cannot be disassociated from the learning platform (Nye, 2015). Barriers that are more likely to affect educational technology adoption in LICs, such as (1) if and how pedagogy using adaptive learning approaches can supplement traditional instructional modes, (2) the limited training time and resources available, (3) subject-domain restrictions about “allowable” instruction approaches, need to be addressed (Blanchard, 2015; Nye, 2015). This thesis seeks to address the lack of research evidence into adaptive learning on mobile platforms in contexts like SSA, with the aim of addressing such barriers, identifying challenges and potential solutions within the clinical subject-domain to improve learning processes and outcomes. In the next section, I detail why the thesis focused on the clinical subject domain.

1.2 Clinical learning needs in Sub-Saharan Africa

Of the estimated 2.5 million neonatal lives that are still being lost each year globally, SSA has the highest overall neonatal risk of death, with more than half of the deaths in children aged under five years occurring in the neonatal period (UNICEF, 2018). In LICs such as those in the SSA, the need for health workers to provide care by themselves is more frequent than middle- and high- income settings, and is associated with negative neonatal outcomes (UNICEF, 2018). Severe health workforce shortages, coupled with the health workforce skill imbalance and maldistribution, and a lack of

training opportunities (Crisp, Gawanas, & Sharp, 2008; UNICEF, 2018) are key contributors to the high neonatal mortality rate. Additional training is needed to better prepare health providers in these regions to provide effective emergency paediatric and neonatal care (Hategekimana et al., 2016; WHO, 2011) but unfortunately, face-to-face training cost between \$80 to \$300 per person per day and is difficult to deliver at scale (Wall et al., 2009). Only a small fraction of trained health providers with the basic new-born resuscitation requisite skills are therefore found in this context (Chaudhury et al., 2016; Lawn et al., 2014). Traditional learning approaches in the clinical domain, especially those requiring face-to-face interactions, tend to be time-consuming, costly and inaccessible (Reeves, Fletcher, McLoughlin, Yim, & Patel, 2017). New novel strategies are therefore required to improve training access for over 1 million health providers across SSA. Any such novel approach needs both to allow for timely and efficient updates as guidelines change in light of new evidence and (ideally) capture data on how many health workers are able to train to them (Edgcombe, Paton, & English, 2016; Frenk et al., 2010).

Use of mobile technologies (smartphones and tablet computers) have been proposed as a solution to this problem (Edgcombe et al., 2016). While they have shown potential, implementation of learning interventions on them has been technocentric in nature, being too reliant on transfer and diffusion models (Dearing, 2009; Winters et al., 2019; Winters, Oliver, & Langer, 2017). As discussed by Lyytinen and Damsgaard (Lyytinen & Damsgaard, 2001), implementation through transfer and diffusion is known to be problematic for the following reasons: their decoupling from educational learning theories, the assumption that the technological innovations are well-defined and understood, and that the population that adopts them is homogenous in properties and characteristic. They also assume that there is sufficient information concerning the characteristics of the social system, and that the communication process explains the adoption decisions (Lyytinen & Damsgaard, 2001). Research into gamified learning by healthcare providers from LICs can hardly be judged to have addressed these concerns (Edgcombe et al., 2016). A key evidence gap here is the understanding of

how digital-based learning technologies for supporting clinical training in SSA can be implemented in a socially embedded manner.

There is scant evidence on the implementation of digital learning interventions that are relevant to LICs, take into account health workers' initial and continuing clinical training needs, and adapt learning content in light of skill mastery and performance as learners continue to develop knowledge (Bollinger et al., 2013; Brusilovsky & Peylo, 2003; Car et al., 2019; Edgcombe et al., 2016; S. V. Gentry et al., 2019; Greenhalgh, 2001; Ma, 2017). In the clinical domain, in high-income contexts, digital education has been demonstrated to be at least as effective as traditional learning and more effective than no intervention in terms of knowledge gain (Car et al., 2019). However, evaluations of the effectiveness of interactive, simulation-based, spaced, and mobile-delivered forms of digital learning for healthcare professions are lacking (Car et al., 2019). In high-income settings, there is evidence of the benefit of using instructional design variations in healthcare that consider learners' characteristics and the data generated during the learning process to adapt training (Fontaine et al., 2019). These adaptive design variations in the instructional techniques are meant to move learners progressively toward stronger understanding and, ultimately, greater independence in their learning process, a technique referred to as scaffolded learning (Maybin, Mercer, & Stierer, 1992).

Within LICs, the investigation of such tailored instructional approaches that are informed by learner models (cognitive models that try to model observed student learning behaviours) is sorely lacking as highlighted by recent evidence reviews (Car et al., 2019; Fontaine et al., 2019; Ma, 2017). The approaches used for healthcare training applications (apps) that have been developed for low-income contexts like SSA and can broadly be divided into two categories:

(1) Some simply replicate existing teaching strategies 'on a screen', for example, by providing questions and answers for exam practice or displaying textbook graphics, i.e. 'drill-and-practice' pedagogical approach;

(2) Others take advantage of features specific to digital devices, examples of which include the ability to respond with different pathways to user choices or the use of animations with which the user can interact, i.e. a more learner-centred pedagogical approach (Edgcombe et al., 2016).

Serious games, which are games with a specific, applied purpose (other than entertainment) that can be played on mobile phones, are one such way of providing interactive, simulation-based, self-paced, and mobile-based learning (Mohan, Schell, & Angus, 2016). The rationale for using gamified learning is that, in a similar way to ‘first-person’ computer games, emergency care training should enable health workers to follow highly structured pathways (such as clinical care algorithms) with pieces of information (cues) sought at each step that determine the correct actions to perform. With both clinical training and performance in computer games, executing cue-response sequences perfectly, rapidly, and automatically, with minimal help (e.g. from corrective feedback and hints) demonstrates mastery. This type of mastery has been shown to improve clinical care delivery behaviours (S. V. Gentry et al., 2019). However, when using traditional learning approaches, the required frequency of rehearsal as that in from gamified digital-based learning is difficult and expensive to maintain for face-to-face training (Fitzgerald, 1997; Reeves et al., 2017). By using a gamified approach, users may be more motivated to repeatedly play the serious game, using incentives such as rewards, increasing difficulty and scores – techniques that have been successfully used to encourage repeated gameplay in non-serious computer games (S. V. Gentry et al., 2019). There is a scarcity of evidence from evaluation and assessment of serious gaming approaches in smartphones for healthcare training in LICs. The need to bridge this gap has been highlighted in the most recent systematic reviews in this area (S. V. Gentry et al., 2019; Maheu-Cadotte et al., 2018).

1.3 The learning platform used in this thesis

My research utilises the Life-saving Instruction For Emergencies (LIFE) platform. LIFE is a smartphone-based clinical training intervention; an Android-based application developed to train healthcare providers from low-resource settings to manage neonatal and paediatric clinical emergencies (Oxford

University, 2015). The LIFE app is a gamified platform intended for use with low-cost smartphones to provide training in the care of very sick neonates with plans of expanding it to include other clinical care scenarios. It evolves scenario-based teaching where the components being assessed emphasise the tenets of paediatric critical care with early recognition of children who need immediate care. This is achieved by using game-like training techniques to reinforce the key steps that need to be performed by a healthcare worker to manage an emergency; an approach commonly referred to as serious gaming (Bergeron, 2006; Wang, DeMaria Jr, Goldberg, & Katz, 2016). Whereas game-based learning approaches involve the design of fully-fledged (serious) games, LIFE uses gamified learning approaches which focus on augmenting or altering an existing learning process to create an experience that learners perceive as game-like (Sailer & Homner, 2019). The LIFE app uses scenario-based teaching, where the components being assessed emphasise the tenets of paediatric critical care, including early recognition of children who need immediate care. Consequently, it follows a specific ordering of clinical care-giving algorithms, with each learning task being timed.

1.3.1 The user journey in the LIFE gamified smartphone application

When the learner initiates the smartphone app, after tapping on the start button, they get a choice of learning scenarios. The user may select a scenario, and background information on the scenario which relates to the learning context and clinical care being trained on is provided (Figure 1.3.1 – 1.3.3).

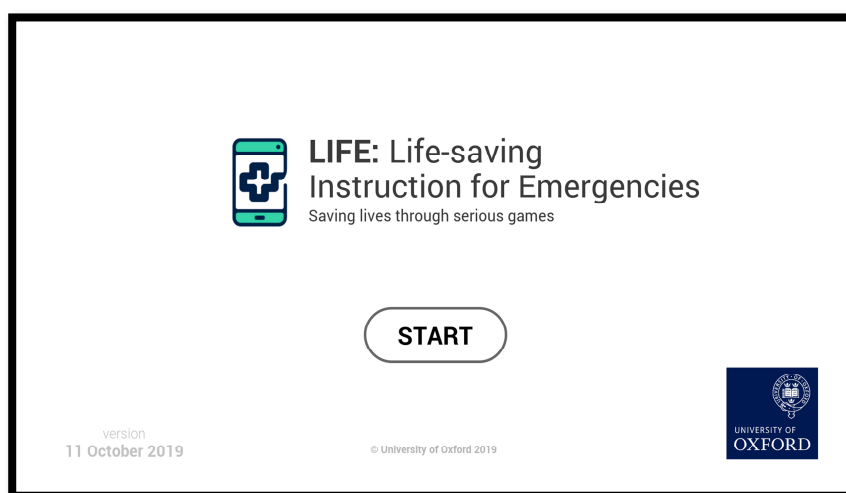


Figure 1.3.1: LIFE start screen

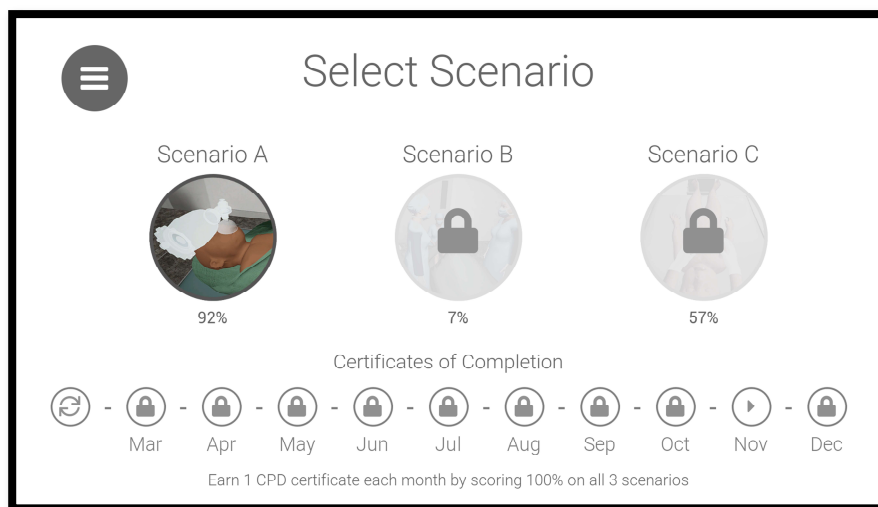


Figure 1.3.2: Screenshot of example of scenario selection in LIFE



Figure 1.3.3: Navigation task provided with scenario background information

The learner navigates to the next learning task by “tapping” on the *next* button on the smartphone (illustrated on the bottom right of Figure 1.3.4). Learning interactions on the LIFE smartphone application are primarily through finger press. There are three main types of learning interactions within LIFE:

- a. Navigation interactions, where the learner must “navigate” the learning space to locate medical equipment, tools or patients, similar to a first-person game (Figure 1.3.4),



Figure 1.3.4: Navigation interaction using on the smartphone device

- b. Multiple choice questions where the learner gets to choose the next steps in the clinical care giving process or identify the clinical criteria needed for decision (Figure 1.3.4).

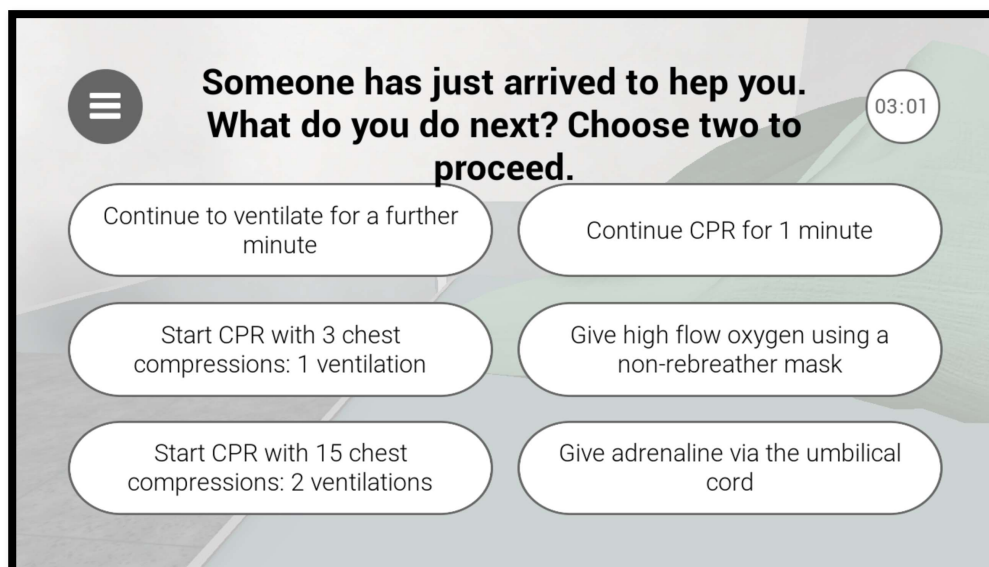


Figure 1.3.5: Multiple choice question illustration

- c. Technique learning tasks, where the learner is assessed on their understanding of how to perform a clinical task, e.g. how to bag-ventilate a baby and perform bag presses at the correct rate consistently. (Figure 1.3.6).

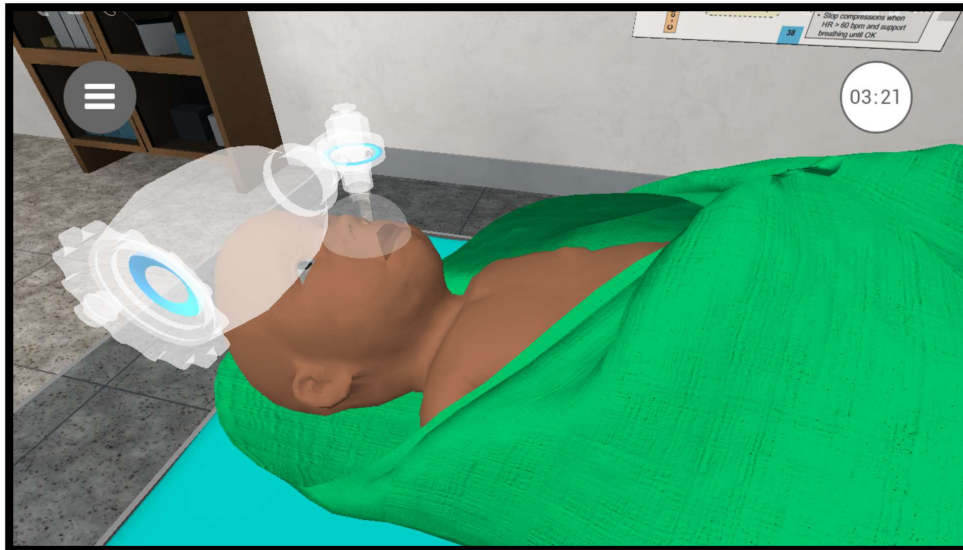


Figure 1.3.6: Learning task assessing student's ability to perform bag-mask ventilation

Each learning task elicits input from the learner. Each learning task is timed (Figure 1.3.6 top-right corner) but the learner has control over how long to contemplate over the question before attempting to answer it. The elicited input can be assessed as being correct or incorrect. For each incorrect learner attempt, standardised feedback is provided, along with the option of seeing more information (Figure 1.3.7).

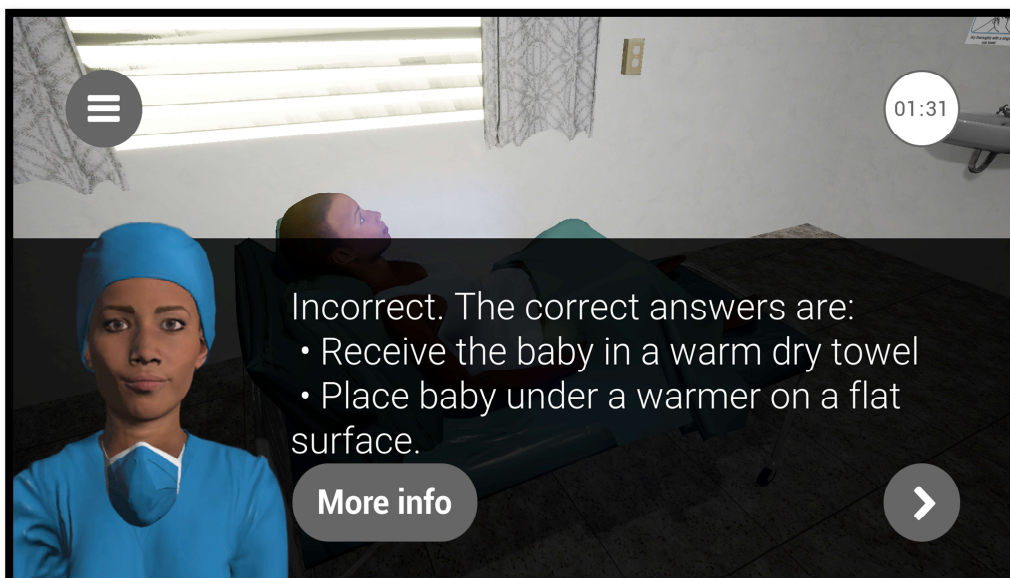


Figure 1.3.7: Screenshot of example error feedback window in LIFE application

The standardised feedback on the first incorrect attempt requests the learner to repeat the task. The feedback on the second incorrect attempt provides a detailed explanation of the correct choices to

select (Figure 1.3.7). The learner is forced to repeat the learning task until they make a correct attempt before they can then proceed to the next task. The end of the scenario is signalled by a crying baby, indicating that the baby is now breathing, and the neonatal scenario has been completed. A breakdown of scores is then provided to the learner (Figure 1.3.8).

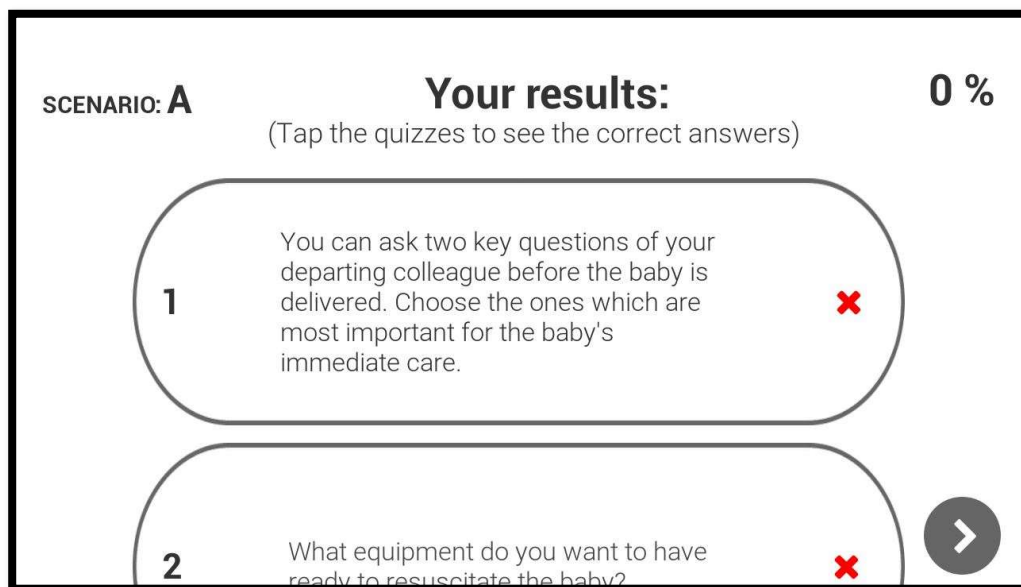


Figure 1.3.8: Learner's score card

Performance is calculated as the total score from all questions attempted where the learner was correct on the first try, without any feedback or help. The learner must attain a perfect score in Scenario A for the next scenario (in this case Scenario B) be *unlocked* (i.e. be given access to the content of the next scenario).

What the learner does during the learning sessions is captured by the number of learning tasks attempted, the number of tries per learning task, the amount of time they spend on each learning task, the level of feedback provided with each try, and the final performance score. Each quiz is a learning task representing a unique knowledge component that addresses a learning "...concept, principle, fact, skill, schema, production rule, misconception..." (Kurt VanLehn, Jordan, & Litman, 2007) within the clinical care-giving algorithm, details of which are provided in **Appendix A**.

The scenario model that is used is one that replicates content from the “Emergency Triage, Assessment and Treatment plus admission care” (ETAT+), a validated face-to-face training programme (Ayieko et al., 2011; Irimu et al., 2008) that is approved by the World Health Organisation (WHO, 2016). The ETAT+ content adapted for LIFE has been used to train healthcare workers and medical students across Eastern and Southern Africa, and now East Asia on evidence-based, best-practice care for most common conditions in very sick children in low-income settings (Ayieko et al., 2011; Irimu et al., 2008; WHO, 2016).

LIFE has been developed to be accessible at scale by healthcare providers and able to function off-line on low-end smartphone devices. The learner gets to choose when to use the learning app: that decision is not prescribed by the LIFE team. A key objective of LIFE is to provide self-regulated training opportunities akin to continuous professional development to learners. LIFE’s design emulates the low-dose high-frequency (LDHF) training model to encourage repeated refreshing of emergency care knowledge but without any face-to-face facilitation or classroom learning (Sutton et al., 2011). In low resource settings, this model, although typically face-to-face, has been shown to be effective for clinical care training in SSA (Atukunda & Conecker, 2017). The development of the simulated scenarios is an iterative user-centred design approach with continuous feedback from new and old cohorts of study participants on training content, platform interactivity, user experience and design parameters for future consideration. The LIFE team continually evaluates the platform with different cohorts of healthcare providers in East Africa.

1.4 Thesis contribution

While there is a consensus on the instructional potential of gamified learning, there persists a lack of methodologies to guide tool design and support analysis and assessment (S. Arnab et al., 2015). Additionally, pre-existing conceptual models that are using student-step trace data from digital learning platforms in attempts to map pedagogical patterns to learning offer little evidence on how adaptive learning might influence the users’ learning experience given:

- The limited evaluation of learners' skills and knowledge application -based on an adaptive approach- **beyond** traditional pre-post assessments (Car et al., 2019).
- Lack of consensus on how-to and which parameters to use to assess the experience of the learning process (S. Arnab et al., 2015).
- Specific approaches to learning analytics (e.g. Deep Knowledge Tracing) are still in their infancy and might still be limited in demonstrating adaptive learning effects (Baalsrud-Hauge et al., 2015; Dawson, Joksimovic, Poquet, & Siemens, 2019).

Furthermore, even though gamified learning offers skill development where there are limited opportunities for practice on rare events, e.g. in emergency care etc. there have been limited attempts to determine if gamified learning is perceived by LICs practitioners to be a sound pedagogical tool in health domain (dit Dariel, Raby, Ravaut, & Rothan-Tondeur, 2013; Stuckless, Hogan, & Kapralos, 2014). Consequently, the following key emergency care training aspects that might address the weak knowledge base and skill imbalance of the health workforce remain unaddressed:

- (1) Relevance to the contextualised needs of healthcare providers from SSA (Edgcombe et al., 2016)
- (2) Accessibility to far-flung healthcare providers in need of initial and continuing clinical training (Frehywot et al., 2013; Kinfu, Dal Poz, Mercer, & Evans, 2009)
- (3) Adaptive to the learning needs and experiences of the healthcare providers as they continue to develop knowledge through it (Bollinger et al., 2013; S. V. Gentry et al., 2019).

Building on LIFE's work (Oxford University, 2015), this thesis represents *an evaluation of an adaptive learning approach* in an interactive, simulation-based, self-paced, gamified mobile-delivered learning for healthcare professions in LICs (Car et al., 2019; S. V. Gentry et al., 2019). This thesis aims to study how implementation approaches to gamified learning can be configured to support scaffolded Self-Regulated Learning (SRL) for clinical training in low-income countries on smartphone devices. Through studying the contextual self-directed learning needs of the individual healthcare providers as they continue to interact with the digital learning platforms, this thesis aims to address the current

lack of systematic understanding of the situated use of such learning interventions including their impact on learning outcomes and clinical practice. My contribution to the LIFE project was implementing the adaptive learning mechanism within LIFE and a data pipeline. This will be explained in detail in the methodology chapter (Chapter 4.1).

1.5 Thesis Outline

[Chapter 1](#) has opened with setting up the scene for the need for learning interventions for clinical training for Low-Income Countries (LICs) in aid of reducing avoidable hospital deaths. It includes the description of a gamified smartphone learning intervention that is envisaged to be implemented to aid in addressing this global health challenge. This chapter has also highlighted the state of computer-based adaptive learning in LICs and the promise it holds for addressing clinical training needs, which prompted me to consider undertaking this thesis.

In exploring how such an intervention might be implemented and scaled-up to support self-directed adaptive learning that is cognisant of the resource and practice context, the literature review is provided in [Chapter 2](#). The literature review explored how media perspectives into gamified learning, clinical perspectives into digital-based training, and educational perspectives into adaptive learning intersect. The theoretical framing of this thesis is captured in [Chapter 3](#) and brings together the different elements of the literature review (media perspectives within gamified learning, clinical perspectives within digital-based training, and educational perspectives within adaptive learning) and discusses the links between educational theory about SRL learning and statistical measurement theory with knowledge tracing approach introduced and discussed in some detail. The summary of findings from the literature and theoretical review identify (1) the main matters of debate in implementing adaptive learning for clinical training in light of the desired patient and clinical outcomes hoped to be achieved, (2) the learning behaviours that need to be addressed for the desired outcomes to be achieved, (3) how to evaluate the knowledge and skills healthcare providers need to facilitate better targeting of these learning behaviours, and (4) how this can be practically implemented for healthcare providers from Low-Income Countries (LICs) using smartphones to learn.

From [Chapter 2](#) and [Chapter 3](#), it becomes evident that to explain and support the nature and behaviour of SRL on an adaptive serious-gaming mobile platform for clinical training in low-income contexts, it is necessary to study in more detail *how* and *why* healthcare providers from LICs engage in SRL practices and how adaptive support can be better designed to enhance this. Thus, the latter section of the theoretical background review introduces the concept of the zone of proximal development where the engagement and motivation of the learner are posited to be enhanceable on digital platforms by leveraging adaptive mechanisms. The summary from the literature and theoretical review is articulated in detail in the justification for the research section ([Chapter 3.7](#) and [3.8](#)) in this thesis. [Chapter 3](#) ends with providing the research questions that this thesis aims to address.

[Chapter 4](#) opens with the overarching design of this thesis broken down into three phases involving:

- I. Exploratory quantitative data collection for the development of a knowledge tracing mechanism for adaptive feedback support, coupled with qualitative data collection through focus group interviews and semi-structured in-depth interviews;
- II. An experimental study evaluating adaptive feedback that uses the knowledge tracing mechanism; evaluation of the effect of learning opportunities and level of adaptive feedback on healthcare providers' learning curves using student-step data from the learning intervention; and the quantitative evaluation of self-regulated learning using a validated modified version of the Online Self-Regulated Learning Questionnaire (OSLQ) tool;
- III. The use of quantitative student-step data collected in phase 2 to comparatively evaluate different knowledge tracing approaches in efforts to address the challenges highlighted in the theoretical chapter ([Chapter 3](#)). It also includes additional qualitative data collection using semi-structured in-depth interviews with participants from the experiment in Phase II.

The methods employed in each phase and the rationale for the individual study design choices are described in their respective sections within the methodology chapter: [Chapter 3.6](#), [Chapter 4.9](#) for Phase I; [Chapters 4.6-4.8](#), for Phase II, and [Chapters 4.9 and 4.10](#) for Phase III. The research questions do not map exactly to one phase, nor does the ordering of the phases coincide with the ordering of

the research questions. The ordering is primarily to enhance the readability of the methodology and results chapter by combining the findings for the qualitative study together.

[Chapter 5](#) begins the series of chapters of this thesis where the research findings are presented: It discusses the findings from the experimental study in Phase II ([Chapter 5](#)), which was aimed at addressing the first part of the first research question (*“Within a gamified smartphone-based learning platform for clinical training, what is the effect of adaptive feedback on learning gains compared to standardised feedback”*). Further findings from Phase II are presented in later sections: [Chapter 6](#) contains results from Latent Profile Analysis of self-regulated learning of healthcare providers from LICs using digital learning platforms. It is aimed at addressing the second research question (*“How do healthcare providers’ SRL strategies manifest and vary by individual learner characteristics? What influence do these strategies have on healthcare providers’ learning gains from using LIFE?”*). [Chapter 7](#) contains results from Learning Factor Analysis and is aimed at addressing the second part of the first research question (*“How is the effect of the adaptive feedback mechanism moderated by (a) the opportunity for knowledge mastery, and (b) the level of adaptive feedback support provided?”*). It includes the SRL profile variable for healthcare providers which was derived from findings in [Chapter 6](#), as a key predictor of their learning curves from Learning Factor Analysis. [Chapter 8](#) presents the findings related to the fourth research question: *Given a healthcare provider’s sequence in the emergency care tasks that they have prescribed; which knowledge tracing approach is better at predicting the outcome of the next caregiving task? How does this prediction vary by the healthcare provider’s spacing of their own learning, experience and clinical role?* To answer this question, Deep Knowledge Tracing model was employed to evaluate the hypothesised knowledge tracing approach in comparison to approaches used in [Chapter 5](#) and [Chapter 7](#).

Detailed analyses of the qualitative data from Phase I and Phase III are presented in [Chapter 9](#) and focused primarily on addressing the third research question: *What are the perceptions of healthcare providers on how to enhance the design and implementation of more engaging learning experiences*

on smartphone-based gamified emergency care training? How do these healthcare providers perceive such contextualised learning experiences to influence their SRL efforts as they embark on bridging their knowledge gaps? [Chapter 9](#) brings together the qualitative findings from across all the phases and seeks to explore how a gamified mobile training intervention might be better designed to support healthcare providers self-regulated learning of neonatal emergency-care management in low-resource settings.

The final [Chapters 10](#) and [11](#), bring together the key findings of this thesis, comparing the findings of this thesis that might have had similar objectives, and delves into the meanings and implications of the key findings (on SRL, knowledge-tracing, gamification and clinical training in LICs) presented in this thesis. Discussion about the strengths and limitations of this thesis are provided in [Chapter 10.4](#), with Conclusion section ([Chapter 11](#)), providing an indication of the future direction of research in this interdisciplinary scope of implementation science.

2 LITERATURE REVIEW

This chapter discusses the literature on the use of digital learning platforms in healthcare training, learning in gamified environments and how it is linked to psychological and learning theories, and the challenge of linking gamification effects on digital platforms to learning outcomes. Approaches to learning from three domain areas, media, clinical and education are detailed from their perspective of how prior work informs the interdisciplinary nature of this thesis. Specifically, *Section 2.1.1* focuses on contributions to gamified learning from the media studies field. It gives special attention to the role of gamification on learner experience, motivation and self-regulation habits with consideration of how they are informed by psychological and learning theories. *Section 2.1.2* reviews empirical studies from the clinical field that describe the gamified approach to digital training, reflecting on the fields' perceptions of learning and assumptions of how learning effectiveness is evaluated and validated. Building on the gaps highlighted in *Section 2.1* and concepts discussed in *Section 2.2*, *Section 2.3* reviews evidence from the educational field on how the concept of adaptive learning within digital-based platforms is understood and has been implemented. It reflects the key theoretical underpinnings of the use of digital platforms for adaptive learning specifically for gamified learning approach.

2.1 Educational perspectives to adaptive learning and gamified learning

Digital-based education, of which digital gamified learning is a part of, has traditionally primarily focused on instructional design, with a particular interest in developing learning experiences and environments that promote and support personalised knowledge or skill acquisition (M. Vandewaetere, Desmet, & Clarebout, 2011). This might be attributable to the capacity in digital learning environments to capture, analyse and cater for differences in learner characteristics -such as cognitive, affective and behavioural features- which are considered important for driving individualisation in the learning trajectories that can be fashioned as a flow channel, a concept that is well detailed in the literature on gamification from media studies. This lends itself well to the ubiquitous and pervasive increase in digital devices, and with it, a radical increase of the possibility of

adopting design frameworks for technology-rich learning due to the upsurge in the capacity of capturing and processing more aspects of digital learning environments (Luckin, 2010, pg. 35), including in the African context (Kumar, Paton, & Kirigia, 2016). The following section ([2.1.1](#)) provides a summary of literature as it pertains to how adaptive learning mechanisms borrow from and mimic concepts from gamification and where pertinent research gaps lie.

2.1.1 Media perspectives to gamified learning

2.1.1.1 *Effects on cognitive learning gains*

Serious-games and gamified processes are learning artefacts that arouse meaningful immersive learning experiences from continuous interaction between the learner and an environment which implement game-mechanics (Dewey, 1998; Kiili, De Freitas, Arnab, & Lainema, 2012). Game mechanics are methods designed for interaction within a game and can include storytelling, increasing challenge with increasing levels, having different game modes (e.g. co-operative versus competitive), risk and reward choices etc. An example of an extensive list of game-mechanics is detailed elsewhere (Al, 2012). Where learning builds upon pre-existing knowledge, serious-gaming and gamified learning produce unique learner experiences on the basis of flow (extreme focus with active control on learning task where the skill and difficulty are roughly proportional), immersion (loss of self-awareness, time distortion), presence (merging of learning action and awareness) and arousal based on subjective mental maps (Kiili et al., 2012). Central to flow is the idea that there should be an optimal match between the skills an individual possesses and the challenges presented by a learning activity (Nakamura & Csikszentmihalyi, 2014), the experience should be intrinsically rewarding, and involve deep concentration and personal learner control, have clear goals and provide direct and immediate feedback (Sherry, 2004).

Considerable amounts of research on digital serious games and gamified learning exist, with much of it being aligned to educational research (Sailer & Homner, 2019). Evidence shows digital serious games have consistently shown positive effects with regard to learning and motivational gains (Clark, Tanner-Smith, & Killingsworth, 2016). However, the same research has struggled to isolate the specific

combinations of game characteristics that are responsible for these positive findings (Imlig-Iten & Petko, 2018). And while evidence points to motivational, interactive, fun and multimedia elements having a causal link to positive effects, these attributes are too general and do not offer guidance on how varying their combinations affect knowledge gains (Abdul, Azita, & Felicia, 2015).

The limited explanation for positive effects of gamified learning is argued to stem from the limited availability of comparable studies which tend to have poorly reported study designs (All, Castellar, & Van Looy, 2014). But of bigger implication is the assumption that learning with serious games is treated as comparable to learning with conventional methods, as opposed to comparing different modalities and applications of serious-gaming: Serious-gaming and non-gaming treatments differ in too many respects making them hardly comparable, especially where moderation and mediation components are dissimilar (Imlig-Iten & Petko, 2018). This treatment of game characteristics on digital learning environments as mere add-ons obscures the following key aspects such as: (1) game characteristics that influence changes in behaviour and attitudes, (2) game elements that affect behaviours and attitudes which moderate instructional effectiveness, and (3) the relationship between game elements and learning outcomes which are mediated by behaviours and attitudes (Landers, 2014).

These unlike comparisons of serious-gaming and conventional training tends towards a conflation of mediation, moderation and intervention effects (Landers, 2014; MacKinnon, 2011). The implication here is that to better evaluate the factors contributing to successful gamification –which remain somewhat unresolved especially for cognitive learning outcomes (Sailer & Homner, 2019)- (a) the comparative gamification effects of the learning process/platform need to be evaluated *on the same platform*, (b) there is need to adopt methodological approaches that can control for potential *lurking and confounding* factor(s) and establish a causal relationship, (c) that the evaluation of the gamification effects or game elements needs to have low *granularity*, and (d) how learner behaviours and characteristics moderate learning outcomes needs to be accounted for.

Studies in using gamification have traditionally drawn from psychology and media studies and tend to utilise well-received theories, examples being the concept of flow (Nakamura & Csikszentmihalyi, 2014) or self-determination theory (Deci & Ryan, 2011). This is in addition to drawing from theoretical frameworks that explain further the psychological aspects that underly the use of digital media (Kiili et al., 2012; O'Brien & Toms, 2008), including engagement and emotions in games (Boyle, Connolly, Hainey, & Boyle, 2012). These frameworks and theories emphasise the need for competence development while encouraging relatedness of the serious-gaming/gamified activity to real-world experience as a necessary aspect of learning (Deci & Ryan, 2011). Evaluation of whether learners can attest to competence development that has a bearing on the real-world application is necessary and has implications on the value they attach to gamified learning approaches.

Gamified learning is posited to (1) increase representational competence, (2) be associated with the development of other thinking skills, (3) enhance inductive rule-discovery through repeated hypothesis testing, and (4) improve learner's tactics of attention distribution (Kiili et al., 2012; Prensky, 2001, pg. 40-46). These gamification-related learning behaviours, which tend to moderate instructional effectiveness (Landers, 2014), are explained in detail in the following section.

2.1.1.2 The new archetype of gamer-generation learning

Where learning happens in an interactive environment, learners are posited to be active participants in their own learning and exercise agency on learning activities (Kiili et al., 2012; Prensky, 2001, pg. 46). While accused of short attention spans, they strategically divide attention into different learning activities while simultaneously offering concentration bursts after selectively determining what counts as learning (Prensky, 2001, pg. 48-50; Sherry, 2004). In addition to Jane Healy's views that among the gaming generation, sustained attention to traditional media of learning is considered far less appealing than faster-paced stimuli, Clifford Stoll opines that "...serious gaming substitutes quick answers and fast action for reflection and critical thinking..." (Prensky, 2001, pg. 50-51). This is a practice of "*on-action*" reflection (afterwards) as opposed to "*in-action*" (during), but it might not be a conscious action by the learner: Only when a learner consciously processes their own learning

experience, can they adapt learning strategies and thereby construct a hypothesis to test (Kiili et al., 2012). Thus, activities related to learning (quizzes and assessments) and activities related to controlling the gamified experience (goal-setting, strategic learning etc.) should be distinct and adequately targeted – this is rarely the case (Abdul et al., 2015).

With serious-gaming learning approaches, there is a preference for a higher speed of content delivery in a non-sequential structure that can be parallelly processed and lends itself to rule discovery through hypothesis testing of selected choices (Prensky, 2001, pg. 35-64). This non-sequential structure does not need to be strictly adhered to in a gamification approach (as opposed to a serious gaming approach) (Sailer & Homner, 2019). Additionally, media studies looking at gamified learning posit that learning should be offered through an interactive environment that can be figured out through trial and error to support logic and complex puzzles tasks and gives agency to the learner on learning activity (Kiili et al., 2012; Prensky, 2001, pg. 35-64). This positioning is informed by how gamers ideally go through *adventure* video games and is illustrated in Figure 2.1.1 below.

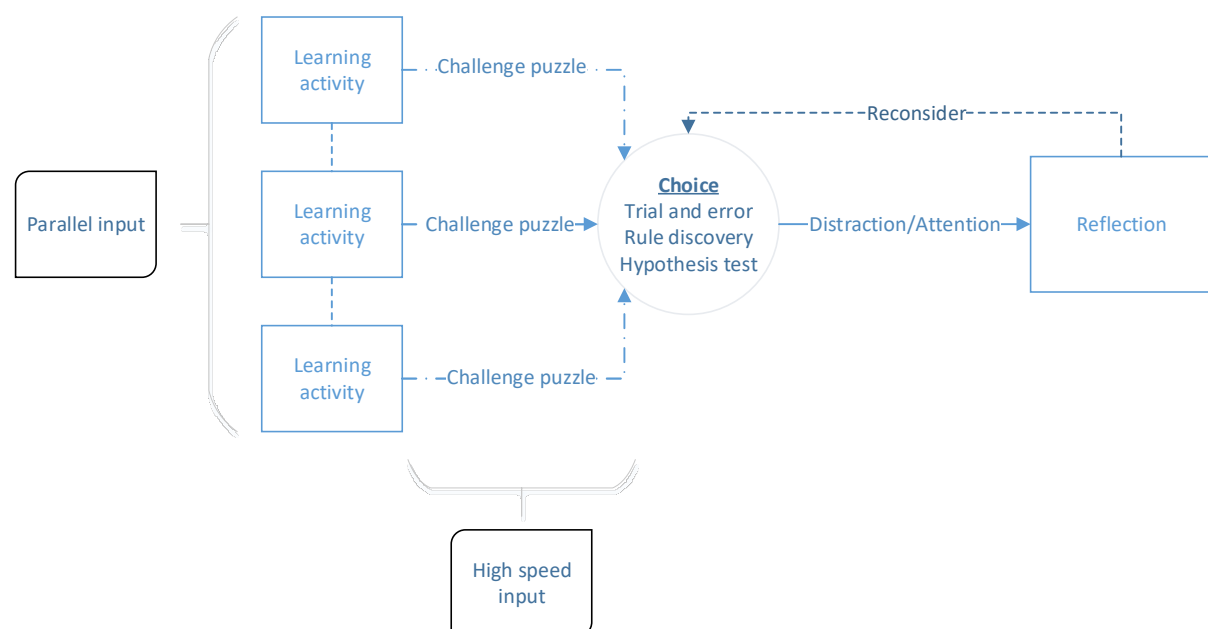


Figure 2.1.1: Conceptual model of digital-game based learning

Finally, within gamified learning, there is an expectation of a payoff for the effort exerted while learning which creates a huge intolerance for unrewarded learning (Deci & Ryan, 2011; Prensky, 2001,

pg. 61). This affects learner's motivation to continue engaging in learning processes and interactions, with higher motivation leading to deeper engagement (Luckin, 2010, pg. 29-31). A supportive mechanism in a gamified learning model that encourages reflective adaption of learning strategies for better learning outcomes is feedback loops: **there is no learning without feedback** (Hattie & Timperley, 2007). A challenge within game-based learning experiences has been that feedback models are often generalised rather than adaptive even though there is enough socio-technological advancement to support the modelling of user performance (Kiili et al., 2012). Additionally, the concept of flow (Nakamura & Csikszentmihalyi, 2014, pg. 74) (Kiili et al., 2012) illustrated by Figure 2.1.2 can be extended by providing some guidance to the learner to support competence development in gamified learning, while ensuring an intrinsically rewarding experience that is supported by direct and immediate feedback (Sherry, 2004). The implication of this challenge is discussed in detail in the next section.

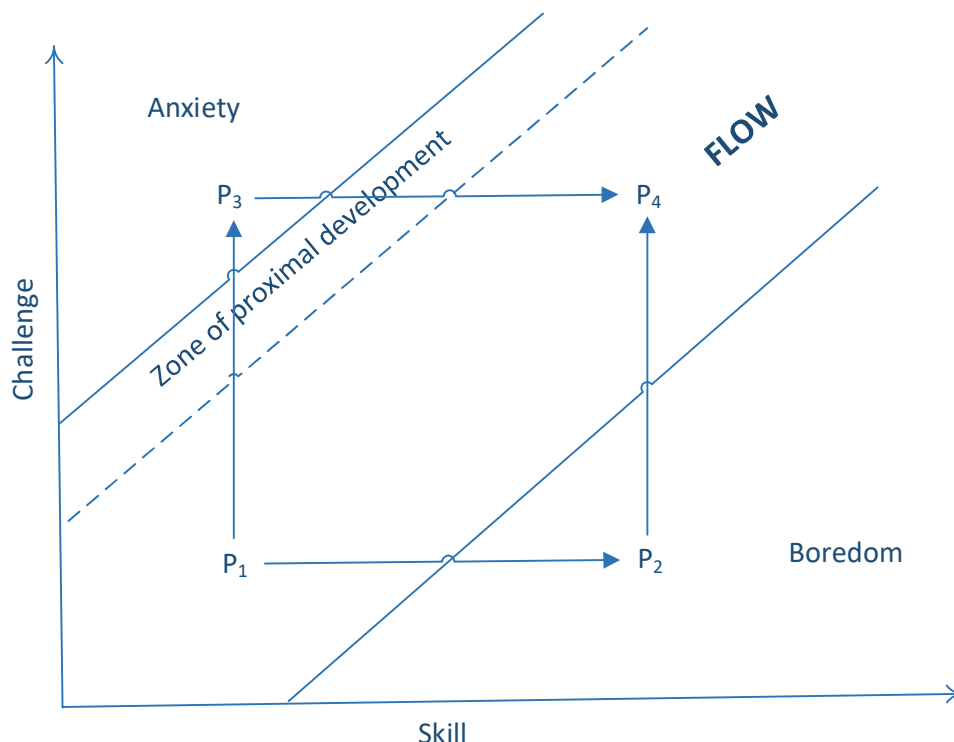


Figure 2.1.2: Flow channel as a representation of skill-versus-challenge learning maximisation process (Borrowed from Kiili et al., 2012)

2.1.1.3 *Commonly overlooked gamified learning concepts*

Learning principles that games incorporate are strongly linked to research in cognitive science where good serious-games are those that offer on-demand just-in-time contextual information within the goals of the current learning task (Gee, 2003). Additionally, "...[they] operate on the outer and growing edge of a player's competence, remaining challenging but do-able...[becoming] pleasantly frustrating, which is a motivating state [for learning]..." (Gee, 2003) arguably scaffolding through the Zone of Proximal Development (ZPD). Scaffolding is a teaching technique that involves providing learners with the support needed to complete a task or facilitate their learning of new concepts by systematically building on previous experiences and knowledge (Olenka, 2009). It focuses on bridging the gap between what a learner has already mastered and what the learner can achieve when provided with educational support (Heather, 2009) by gradually scaling back the assistance provided to learners as they progress through increasingly complex learning tasks (Luckin, Holmes, Griffiths, & Forcier, 2016, pg. 25). The bridgeable gap is what is referred to as ZPD, and its relation to the flow concept is illustrated in Figure 2.1.2.

The ability to adapt game progression to suit the varying levels of ability of the learners and patterns of playing while adjusting for the growing competence of the learners lends itself well to learning theories (Chaiklin, 2003, pg. 39-64). As illustrated in Figure 2.1.1, game-play designed to allow the learner to discover rules and patterns embedded within the game, and make testable hypotheses usable in future more complex phases of gameplay builds inference making and problem skills in the learner (Gee, 2003). Serious gaming implicitly offers adaptive learning environments geared towards a flexible, individualised, engaging learning process. From previous evidence synthesis, knowledge representation used to establish a differentiated interaction between the learner and the [gamified] digital adaptive learning platforms can be characterised through the use of following models (Luckin et al., 2016, pg. 19; M. Vandewaetere et al., 2011):

- a. ***Pedagogical model:*** the theoretical approaches to instructional strategies that are considered appropriate. It defines what can be adapted, when and how it is to be adapted.

- b. **Domain model:** the subject matter that is being taught. E.g. for LIFE, it is domain expertise in emergency care delivery competence for neonates in LICs settings.
- c. **Learner model:** What we know about the learner and what the learner knows. i.e. previous learning achievements and challenges, affective states, and engagement with the learning task.

While these models (i.e. *pedagogical-domain-learner*) have been represented as distinct, the learning process within the adaptive digital environment is usually an interaction of these domains of knowledge, e.g. based on learner capabilities and needs (*learner model*), domain and pedagogical models determine the next most appropriate learning interaction, with their resultant outcomes being continuously fed back to the learner model to make it richer and more personalised in supporting learning effectively. This offers the ability to reflect on learners' critical thinking approaches and how they regulate it (Luckin et al., 2016, pg. 20-21). For gamified learning on digital platforms, a key gap here which arguably is rarely, if ever, attended to in clinical education literature, lies in mapping out the pedagogical and learner models that expose the needs of the learners.

There is an insufficient evidence base to allow the implementation of this typical *pedagogical-domain-learner* models for achieving adaptivity in education in LICs (Nye, 2015; Poushter, 2016; Silver & Johnson, 2018). Specifically, there is insufficient research on learning theories and student models that might inform the implementation of self-directed digital learning platforms in LICs for healthcare providers. From some of the most referenced comprehensive systematic reviews on adaptivity in digital education and AleD, the lack of empiric work representing the LICs learning context is self-evident (Holmes et al., 2018; Holmes, Mavrikis, Hansen, & Grawemeyer, 2015; Luckin et al., 2016; Ma, 2017; Nye, 2015; Mieke Vandewaetere, Cornillie, Clarebout, & Desmet, 2013; M. Vandewaetere et al., 2011). This is a non-trivial issue: we cannot presume that this model applies to LIC context given arguably most if not all of its evidence base is from high income countries and non-clinical domain settings, where the resources, culture, expectations and context of digital-based learning are

demonstrably different from LICs (Blanchard, 2015; Conn, 2017; Nye, 2015). Therefore, while the *pedagogical-domain-learner* models is considered to be the desideratum of adaptive learning in education, because of the scant evidence of its application or relevance to LICs, this thesis is not focused on applying it, rather it is geared towards generating formative evidence on *pedagogical-domain-learner* building blocks (i.e. the individual models) in a contextually cognisant way, as a preamble for future implementations of digital-based learning interventions in LICs like SSA.

Gamification offering an adaptive learning environment can be equated to an Intelligent Tutoring System (ITSs) which delivers scaffolded learning activities based on learner's self-regulated cognitive ability to appropriately challenge and support learning by targeting the flow channel (Figure 2.1.2) (Luckin et al., 2016, pg. 25). Intelligent Tutoring Systems (ITSs) are software programs that dynamically model learners' psychological states in order to provide them with tailored individualised instruction (J. Lee & Park, 2008, pg. 469-484; Ma, 2017).

This also presents an opportunity for exploring within LIFE, how the shift from 'stop-and-test' to a dynamic formative evaluation of students while they learn would look like, given the situational context and subject domain. Additionally, it would allow for the exploration of how learners arrive at answers they provide given the learning interactions and changes in their engagement and motivational states throughout the learning process mediated by the learning environment. This is posited to help develop the best teaching practice approaches for individualised practice-based learning in the future (Luckin et al., 2016, pg. 34-36). Literature acknowledges that technological advances in education have produced distinctive approaches to individualised instruction while ameliorating computational limitations that might have hampered efforts to provide adaptive learning to the learner in the past (Dror, 2008; M. Vandewaetere et al., 2011).

Consequently, there is a growing call for an individualised learning experience in the clinical domain (Fontaine et al., 2019). But this individualised-learning approach entails major challenges, both for the learners who typically have to self-direct their learning, and the tutors who have to support them

(Holmes et al., 2018). It remains unclear how technology aids these efforts in the presence of different learning contexts (e.g. LICs resource-constrained settings) and specifically for clinical subject domains (e.g. continuous professional development in clinical training). Additionally, it is also unclear if and how effectively these technological gains can be leveraged for adaptive learning when using low-end mobile devices that are typically used in an offline mode (i.e. without internet) in LICs setting. What clinical education literature emphasis in respect of digital education is detailed in the next section ([2.1.2](#)).

2.1.2 Clinical perspectives on digital-based training and gamified learning

Serious gaming for health professional education is seen as an innovative educational intervention whose primary focus is facilitating “...the adoption of new ways of understanding and developing learning...” (Sangrà, Vlachopoulos, & Cabrera, 2012). However, the studies that I describe in this section raise key challenges in defining (and measuring) what the utility of gamified learning as a teaching strategy is (S. Gentry et al., 2016). There is a proposed consensus framework that on learning outcomes, demonstrable non-inferiority to conventional training should be the irreducible benchmarks for gamified learning in healthcare and that measurement is well-founded and corresponds accurately to the real world (Graafland et al., 2014). This framework is further strengthened by the adoption of standardised validity measures for medical education quality assessment of serious games defined by the following domains (Graafland et al., 2014):

- Content validity – The legitimacy of the subject matter being taught should be determinable
 - In many cases, this translates to externally validated evidence-driven medical practices and guidelines.
- Construct validity – Serious-gaming platforms and gamified learning should be able to measure some form of educational and learning realities, the example being the effect on learning outcomes.
- Concurrent validity – Performance in gameplay or gamified learning process should be a valid indicator of actual learning gains.

- Predictive validity – Whether performance in the game leads to better clinical and patient outcomes. This is considered the pinnacle of serious-gaming and gamified learning validity in the healthcare domain.

However, from the Graafland *et al.* framework, validity -while important- is still skewed towards didactic functionality and is not the only predictor of a gamified learning approach's success or its attractiveness to the learner; As detailed in [Section 2.1.1](#), gamification approach's success also depends on the level of flow (extreme focus with active control on learning task where the skill and difficulty are roughly proportional), immersion (loss of self-awareness, time distortion), presence (merging of learning action and awareness), and arousal based on subjective mental maps (Kiili et al., 2012).

A complementary approach to, and widely used with *Graafland et al.* 's validity measures is the Kirkpatrick's model (Kirkpatrick, 1996), commonly used for assessing the impact of technology-enhanced training in healthcare (Cheng, Lang, Starr, Pusic, & Cook, 2014; D. A. Cook et al., 2013; Wang et al., 2016). The Kirkpatrick's levels of evaluating technology-enhanced training, sequentially ordered, are: (1) reaction (a measure of participants initial reactions to the learning content), (2) learning (a measure of the amount of information that participants accrue), (3) transfer (a measure of the amount of material learned that participants actually use in everyday work) and (4) value to the organisation (a measure of the financial impact of the training course on the bottom line of the organisation). Learning and transfer levels of the Kirkpatrick's model can be loosely equated to construct, and predictive validity, respectively, as defined by *Graafland et al.*'s study. Its added input to the examination of the validity of gamified learning platforms in the health domain is the inclusion of considerations around learner's initial experiences, and sustainability of the results of the learning intervention over time.

These two benchmark criteria (which are the most commonly used criteria for assessing the quality of digital education interventions in healthcare) are pivotal to understanding perceptions of health

domain towards learning. They demonstrate a working assumption that a gamified learning approach's value is predicated on measurable cost-effective transfer to practice with implementation focusing on patient outcomes – arguably taking a utilitarian view to learning. Furthermore, they also adopt a pragmatist evaluative approach where the emphasis is primarily on 'what works', which loosely translates to the healthcare professional who used the intervention are assumed to learn if their practice performance improves. As a result, "to know" and "to perform" concepts are treated as one and the same (Corbett & Anderson, 1994). Under this assumption, explanations of what are the active mechanisms for the educational intervention, how they function and why -such as in a realist evaluation (Pawson & Tilley, 1997)- are consigned to perpetual under-theorisation and under-exploration. The moderation effects of gamification in learning which have been highlighted in [Section 2.1.1](#) also end up being overlooked.

While measurement methodology frameworks from Graafland and Kirkpatrick *et al.* provide a useful starting point for effective evaluation of technology-enhanced learning, they may not be easily implementable in the evaluation of digital-based clinical education in LICs. This is due to the significant challenge of:

(1) Producing evidence on 'effective' health care education where finding quality patient outcome data is hard; making it difficult to assess whether what is learned is "...implemented in practice beyond a practitioner's own perception of knowledge gain or confidence..." (J.D. Pickering & Joynes, 2016), and

(2) It is assumed that aspects that reflect technology-enhanced learning experience at an individual level -which may have a considerable impact on learner's behaviour and overall intervention effectiveness- are measurable (in a standardised way) in the same manner as individual-level pre-test post-test scoring (Nicoll, MacRury, van Woerden, & Smyth, 2018).

These assumptions have made attempts to summarise evidence on the effectiveness of technology-enhanced learning in general medical education challenging (D. A. Cook et al., 2013; D. A. Cook et al.,

2011; Okuda et al., 2009), and more specifically for resuscitation (Mundell, Kennedy, Szostek, & Cook, 2013), and for modalities using gamified learning (Wang et al., 2016). The challenges -which corroborate findings from the literature reviewed in *Section 2.1.1* - include: (1) high inconsistency in findings due to high variation (without drilling down) on simulation modes, clinical topics, learner groups, instructional designs, research methods, and outcome measures (Cheng et al., 2014; D. A. Cook et al., 2013; D. A. Cook et al., 2011; Mundell et al., 2013); (2) which arises primarily from included studies having undefined contextual nature, most likely contributing to large between-study inconsistencies (D. A. Cook et al., 2013; D. A. Cook et al., 2011; Mundell et al., 2013); and (3) the poor the quality of studies available for inclusion tends to limit the credibility of findings from these evidence syntheses (Cheng et al., 2014; Mundell et al., 2013; Wang et al., 2016).

From these previous studies in healthcare training, key recommendations predicated on a realist evaluation (Pawson & Tilley, 1997) - which take a fundamentally different research design to those discussed above- insist on the need for future work to clarify the mechanisms of effective technology-based learning: when it comes to instructional design features for specific topics, what works, for whom, in what contexts and why (D. A. Cook et al., 2013). The comparative benefits of gamified learning approach to resuscitation training are less clear (Mundell et al., 2013; Wang et al., 2016), even though they are postulated to enhance learning effectiveness on the basis of cognitive science frameworks (Drummond, Hadchouel, & Tesnière, 2017), and considered to be cost-effectiveness when on mobile platforms (Edgcombe et al., 2016; Kumar et al., 2016; Wang et al., 2016). The impact of a gamified learning approach on emergency care training efficacy is yet to be thoroughly elucidated and requires its own study (Cheng et al., 2014; Wang et al., 2016). Additionally, from recent systematic reviews on digital learning for clinical training (Car et al., 2019; S. V. Gentry et al., 2019), none of the studies included:

- (1) Used mobile devices for delivery of learning interventions,
- (2) Evaluated novel educational modalities enabling simulated learning such as adaptive learning,
- (3) Provided all the essential methodological information necessary for comparability purposes,

- (4) Reported relative change in learning outcomes except one, and
- (5) Were from low-income settings such as Sub-Saharan Africa or South-East Asia where learning context, needs and perceptions on gamified learning might be substantively different.

In general, the perspective within medical field towards learning using serious-gaming and gamified learning tends to be devoid of explicit use of learning theories, making it difficult to explain how knowledge is absorbed, processed, and retained during learning, in what context the gamified approach is most effective and why, whether the perceived [in]effectiveness is domain-specific, and how differences in learners affect its usefulness as a learning approach. Subsequently, Graafland's and Kirkpatrick's frameworks are implicitly biased towards -and exhort- the use of cognitive and behaviourist learning approaches at the expense of social, humanistic, motivational, reflective and transformative learning theories (Taylor & Hamdy, 2013). This preference, I argue, is not underpinned by evidence. The limited explicit utilisation of learning theories limits gaining a deeper understanding of the active ingredients that underlie digital-based serious-gaming. This is evidenced in the decoupling of theory from the structure and content of most common tools used for appraising methodological rigour of medical education interventions (e.g. *Medical Education Research Study Quality Instrument* (MERSQI) and *Newcastle–Ottawa Scale-Education* (NOS-E) scores) (D. A. Cook & Reed, 2015).

Furthermore, from clinical education literature, the learner's experiences on digital platforms and the dynamic nature of pedagogical and learner contexts are rarely, if ever, attended to by use of theory. This limits the development of an understanding of how during game-play, changes in learners' emotive states affect their learning experiences and whether these changes affect their motivation, strategy for self-directed learning and ultimately, the learning outcomes (Nakamura & Csikszentmihalyi, 2014). How adaptive learning can be used to harness learning trajectories from digital platforms needed for evaluating learning processes, experiences, and outcomes is discussed in the next section.

2.2 Approaches to adaptive learning on digital platforms

Adaptive learning environments are conceptualised as “...digital learning environments that adapt teaching, learning approaches and materials to the capabilities and needs of individual learners...” (Luckin et al., 2016, pg. 18). Such digital platforms are designed to interact with the learner by supporting capabilities and intelligent behaviours that would be considered essentially human, and are considered to be *Artificial Intelligence*-based systems: they depend on knowledge about the world and algorithms to intelligently process that knowledge (Luckin et al., 2016, pg. 14).

Learning trajectories are artefacts from possible learner interactions provided by the digital learning environment; digital learning cannot be separated from platforms they are situated in (Luckin, 2010, pg. 1-3). This definition assumes that in a digital learning environment, learning pathways (which are dynamic and responsive to learner’s passage) and learning interactions are mutually constitutive; Digital contexts are made up of these pathways; and that learning trajectories can be good measures of learning over time (Luckin, 2010, pg. 6-18), and can be interpreted as representations of the flow channel (*Chapter [2.1.1.2](#) - [2.1.1.3](#)*). Learning here refers to the adaptive process of accumulating partial solutions to frequently encountered learning tasks whereas context can be either internal (based on cognitive functioning) or external (based on learner actions around learning artefacts) (Luckin, 2010, pg. 6-18).

This would suggest that learning trajectories within digital environments are an interaction between learning behaviour and context (both internal and external) thereby attributing it with temporal, localised, and experiential characteristics. The digital learning platforms are, therefore, also capable of generating many learning trajectories due to the myriad of possible interactions. However, treatment of learning trajectories within digital learning environments is again undertheorized, underexamined, and underdeveloped (Luckin, 2010, pg. 6).

These adaptive learning environments deliver information processing that is fundamental to diagnosing learner’s specific learning needs during instruction, subsequently providing an appropriate

instructional prescription for identified needs, ideally within learning process (Mödritscher, Garcia-Barrios, & Gütl, 2004). They also dynamically determine which instructional treatments are most appropriate for a learner in a given situation. This is achieved using learner-generated within-task measures or learner characteristics such as motivation levels over time (M. Vandewaetere et al., 2011). Intelligent Tutoring Systems (ITSs) are examples of adaptive systems, using artificial intelligence techniques in order to provide tailored and on-time instruction to the learner (J. Lee & Park, 2008, pg. 469-484). Any digital learning approach, where the mechanisms to differentiate individual learner needs is in the very least not inclusive of a criterion of within learning task measures, would hardly qualify as an adaptive approach to learning. This would disqualify from being referred to as adaptive environments those that only (1) permit learners to move through the course at speed according to their ability and other demands upon their time, OR (2) use pre-task aptitude-based measures to *only* prescribe static learning sequences during the learning process – as proposed by *Vandewaetere et al.* These, in my view, tend toward oversimplification of complex relationships between individual learner differences and learning outcomes and overlook continuing knowledge formation from interaction with the learning environment.

Adaptive learning platforms can be conceptualised as being tripartite in nature, consisting of (1) *source* of adaptation, (2) *target* of adaptation, and (3) *relationship* linking the *source* and *target* (M. Vandewaetere et al., 2011). This is illustrated in Figure 2.1.3 below. In this conceptual model, in order to be qualified as adaptive, digital environments ought to have variables that describe the learner's next task and the instructional intervention related to steps within each task, e.g. cascading levels of hints and feedback, by difficulty and complexity of task etc. (Kurt Vanlehn, 2006).

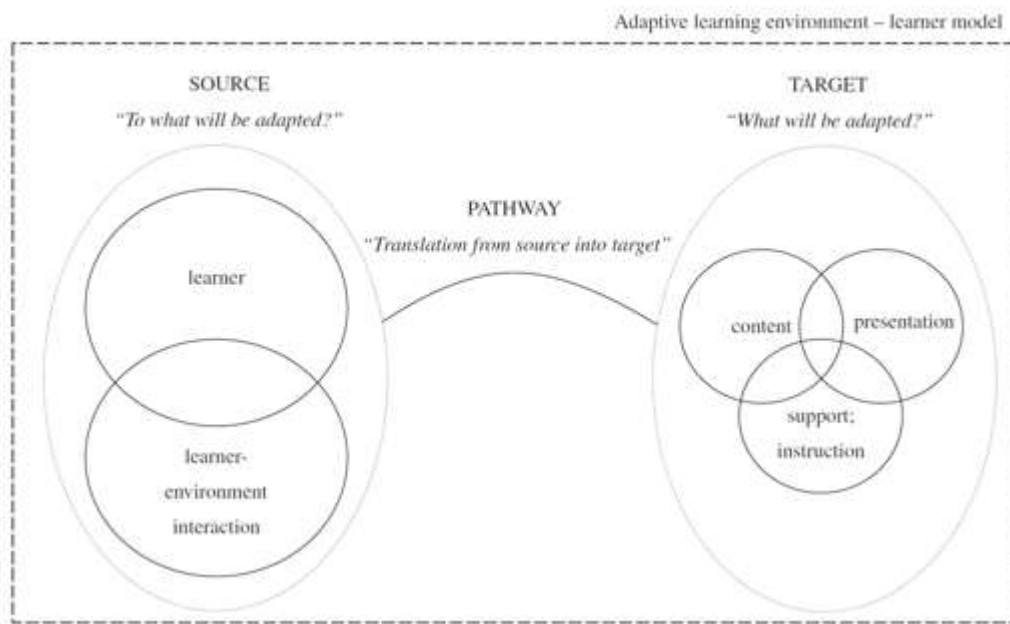


Figure 2.2.1: The tripartite structure of adaptive instruction (M. Vandewaetere et al., 2011)

Vandewaetere et al. posit that the *source* of adaptive instruction is categorised based on learner's *cognitive* (e.g. prior knowledge, learning orientation, learning goals etc.), *affect* (e.g. self-efficacy, frustration, motivation etc.) and *behaviour* (e.g. the number of tries per task etc.) characteristics. Given the existence of numerous studies demonstrating how *affect* and *behaviour* are a function of learner-environment interaction (Chaiklin, 2003; Luckin, 2010; Nakamura & Csikszentmihalyi, 2014), this conceptualisation is challenging: in an adaptive learning environment, the learner-environment interaction is based upon and cannot exist apart from the learner. How the environment is designed, contributes to *engagement* and *motivation* and ultimately, *behaviour*, which is discussed in detail in Chapter 3.3.

Learning from ITSs that personalise instruction –which is seen as the desideratum of today's digital-learning environment (Popescu, 2010)- does not outperform learning from individualised human tutoring or small-group instruction (Ma, 2017). However, the latter instructional models are highly unlikely in the SSA context given the urgent high demand for skilled health workforce and the resource constraints to support such mode of learning. In non-clinical, high-resource contexts, adaptive

instructional support has been shown to significantly outperform teacher-led large-group instruction, non-adaptive computer-based instruction, and paper-based instruction in producing learning gains (Fontaine et al., 2019; Ma, 2017; Ma et al., 2014; WHO, 2011). Within low-resource contexts, investigation of learner models (cognitive models that try to model observed student learning behaviours) needed to support such tailored instructional approaches in clinical training settings is necessary but unexplored (Car et al., 2019; Ma, 2017).

2.3 Pathways of adaptive instruction – How the source is translated into target

Theoretically, adaptive and intelligent digital learning environments are different concepts representing two different approaches. Adaptivity refers to the ability of the learning platform to handle learner models so as to “*behave*” differently based on individual learner capability, while intelligent systems tend to refer to artificial intelligence techniques that better support learners, i.e. the mechanism by which adaptivity is achieved, e.g. using computational algorithms to provide learning content based on specific learner’s interactions with digital platform etc. (J. Lee & Park, 2008, pg. 469-484; M. Vandewaetere et al., 2011). Of concern is the lack of distinction between these two concepts, to the extent that both systems are treated as equivalents (Brusilovsky & Peylo, 2003). The challenge here is that it minimises research into differences in different mechanisms for mapping the pathways illustrated in Figure 2.1.3 and their comparative effectiveness on the learning process and outcomes. Some of the commonly used sources of adaptive instructions include (Brusilovsky, 2007, pg. 263-290; M. Vandewaetere et al., 2011):

1. Stereotypical clustering: Learners are clustered in groups based on common character traits and receive the same instruction adapted to the group profile. As previously stated, the use of static prescriptive learning sequences barely constitutes adaptive instruction, given that the learning needs of individual students are likely to change during the learning processes. This approach lends itself to oversimplification of group learning in the following sense: individual learner needs changing impact upon the credibility of “static” group identity and the

subsequent instruction prescription based on group profile might end up being rendered unfit.

Given the just outlined reasons, I do not consider stereotypical clustering as an adaptive learning mechanism.

2. Feature-based: Changes to specific individual learner attributes, e.g. learner's prior knowledge, resources utilised and learning goals etc. are dynamically tracked so that the model can be updated during the learner's interaction with the digital environment, and thus cater for changing learner's characteristics during interaction with the learning environment (Brusilovsky, 2007, pg. 263-290).
3. Constraint-based: Focuses on erroneous knowledge state of individual learner: Considers the recognition and correction of performance errors and faulty knowledge as core activities in the learning process. The key difference between the default feature-based and constraint-based pathways is "...that the former focuses on *procedural* knowledge (knowing how to do something) and evaluates learners' actions, while the latter tends to focus on *declarative* knowledge (knowing that something is the case) in the form of constraints and evaluates the problem state a learner has arrived at ..." (Mitrovic, Mayo, Suraweera, & Martin, 2001; M. Vandewaetere et al., 2011). Both can be modelled as feature-based pathways of adaptive-instruction, only the emphasis on what are the core activities of the learning processes differ. With regards, to a constraint-based pathway to adaptive instructions, its aims for the learning process seem most representative of the general aims of serious gaming in health for training health workers (Mohan et al., 2016).

A key challenge with the typological differences from the above classification criteria is the indistinctiveness of the adaptive mechanism boundaries. An arguably more informative approach is looking at the pathway of adaptive mechanisms based on how decisions on the dynamic updating of the target (Figure 2.1.3) are arrived upon grounded in learning assumptions. According to a recent systematic review, one way to untangle the complexity of adaptive mechanisms is "...to consider personalised learning in terms of its **multiple dimensions** ..." which include:

1. The personalisation of **why** something is to be learned (e.g. the learning aims);
2. The personalisation of **how** it is to be learned (the learning approach);
3. The personalisation of **what** is to be learned (the learning content and pathways);
4. The personalisation of **when** it is to be learned (the learning pace);
5. The personalisation of **who** is involved in the learning (the learning group);
6. The personalisation of **where** the learning takes place (the learning context) ...”

(Holmes et al., 2018).

A key shortcoming of research into approach and pathways to adaptive learning in education is that it tends towards evaluation using technological criteria as opposed to pedagogical effectiveness, partly due to challenges in the implementation of theory within a digital-based instructional design (M. Vandewaetere et al., 2011). This is evident in the ITSs models skew towards siloed technological propositions as opposed to augmenting it with empirical theoretically-driven interdisciplinary evaluations (M. Vandewaetere et al., 2011). It is further compounded by the lack of a nuanced description of tripartite nature of learner models by previous reviews (Kulik & Fletcher, 2016; Ma et al., 2014; Steenbergen-Hu & Cooper, 2013, 2014; K. VanLehn, 2011); The interaction effect between the source, target and the mechanism chosen to support adaptive learning remains undetermined. How the tripartite nature of adaptive instruction determines the learning effectiveness and varies with (1) learning theory used, (2) reliability of evidence quality and (3) nature of subject-domain, also remains unclear.

Within the SSA context, there is arguably limited evidence of how in using a gamified learning approach, adaptive learning might be effectively implemented to support the learning of procedural knowledge (“to preform”) and declarative knowledge (“to know”) in emergency care. As to what factors about the SSA context should be factored into the decision process on the estimation of such knowledge or learning progression on digital platforms, there is no evidence direction. This is further compounded by lack of evidence of what types of data reflecting semiotic practices (which are useful

for medical educators) would ideally be needed for understanding directed self-learning using serious gaming in within this context.

2.4 Summary

The literature reviewed highlights why this thesis -positioned as research into adaptive gamified learning in clinical training in LICs- is necessary in understanding how the instructional design of the learning process that incorporates gamification can be evaluated, while controlling for potential *lurking and confounding* factor(s) in order to establish a causal relationship, and account for contextualised learner behaviours and characteristics which tend to moderate learning outcomes. Chapter [2](#) also focused on the concept of validity in healthcare-based technology-enhanced learning, -which, while important, is still skewed towards didactic functionality even though it is not the only predictor of a gamified learning approach's success or its attractiveness to the learner; Together with the mechanism by which adaptivity is achieved, the interaction between the learning trajectories within digital learning environments, and learner behaviours remain undertheorized and underexamined in clinical training. And thus, evidence on the best approach for implementing a gamified adaptive learning mechanism that encourages reflective adaption of learning strategies for better learning outcomes using feedback loops for contexts like SSA remains scant.

In the next chapter, I provide an in-depth discussion of the theoretical and conceptual link between adaptive mechanisms on digital platforms, self-regulated learning theory and the appropriate measurement theories. I also detail the research questions that this thesis aims to address and provide their justification provided, having discussed the theoretical approaches that I considered for this thesis.

3 THEORETICAL BACKGROUND

In this chapter, I explain my reasoning behind the adoption of a constructivist approach to learning as the theoretical basis for the design of LIFE's adaptive-learning mechanism to support learners as they co-construct their own understandings of the educational content. It is implausible for any single overarching theoretical framework to account for how people learn in all situations. The emphasis of this chapter is, therefore, to consider a theoretical approach that makes connections between different psychological and pedagogical theories highlighted in [Chapter 2](#). I will demonstrate how in taking up a self-regulated learning theoretical approach,

- (1) it is possible to explore an adaptive learning approach that relies on the model building of the individual learner and is strongly based on interactivity with the learning platform.
- (2) It is possible to generate learning trajectories that mimic the flow channel as resulting from cognitive processes where the experience of the subject domain is assimilated and filtered through prior knowledge,
- (3) The way in which a learner actively constructs an understanding of a subject domain is both informed by the nature of the subject domain, the social context of learning and produces measurement implications of how to track the individual learning trajectories.

Controlling for potential *lurking* and *confounding* factor(s) of a gamified adaptive learning approach imposes an experimental approach, which is useful in determining the causal relationship of the mechanism to learning outcomes. However, this chapter also sets out to provide an argument for the theoretical approach that underpins the measurement of the contextualised learner behaviours and characteristics which tend to moderate learning outcomes and the cognitive processes that produce such behaviours. And why such a self-regulated learning-linked measurement approach matters for formative evaluation in the clinical domain.

3.1 Constructivist-based learning and cognition in clinical education

In constructivist-based learning, the learners construct knowledge for themselves as they learn, individually and socially where meaning-making is presumed to be learning; there is no other kind (George E.H., 1991). Consequently, its primary focus is on the learner when designing learning, not subject pedagogy, and it assumes the non-existence of knowledge independent of the meaning attributed to experience (constructed) by the learner. However, the assumption that the sum of what is knowable is only what the learner can construct for themselves ignores the generative mechanisms of evidence-based practise and knowledge in the healthcare domain (e.g. impact of erroneous knowledge on efforts to minimise mortality risk), exactly because constructivist learning can be argued to not prioritise the impact of poor knowledge of the subject matter. This risk-averting extrinsic motivation towards learning in the healthcare domain ensures the learner's interest is maintained by promoting increased competence and subject-domain relatedness (Ryan & Deci, 2000). It necessitates the ability to formatively evaluate whether the learning processes are building the learner's *competence* relative to *subject-domain relatedness*.

Learning in the healthcare domain can be argued to be developing an understanding of the "true" nature of things, as we continue to uncover them (a structuralist paradigm of critical realist view), as opposed to a personal and social construction of meaning out of the bewildering array of sensations which have unordered structure apart from explanations which we fabricate for them (a relationalism view) – which is how knowledge of social phenomena is approached within constructivist theories. This implies that learning in the healthcare domain would take *a close-ended approach* to instructional design that deviates from a typical gamification's open-ended learning approaches approach.

This thesis research does not look for evidence of learning solely in the 'minds' of the LICs healthcare providers. Instead, I argue that their knowledge is expressed in their abilities to (1) *interact* with external tools -such as learning platforms- and (2) *integrate* their knowledge into the flow of how they would provide clinical care (Säljö, 2010).

3.2 Theorising learning relative to instructional design

The design of adaptive learning in health on the gamified digital platform must understand the situational position of the learner, the nature of the medical education domain, and how these influences the pedagogical and learner model (Smith & Ragan, 1999, pg. 12). Specifically, in accounting for individual differences among the learners and their learning trajectories, it is necessary to explore instructional design issues around scaffolding and related learning theories (Ertmer & Newby, 2013), and in particular:

1. The foundation for the use of scaffolding in engaging strategy selection.
2. The integration of scaffolding into the gamified learning context
3. Reliable prediction of learning progression needs and outcomes within directed self-learning.

When considering the gamified learning archetype (Figure 2.1.1), cognitive processing required by the learner is typically evaluated against the flow channel concept (Figure 2.1.2). To ground this research's strategy for instructional design within gamified learning in health in learning theory, the remainder of this chapter will focus on *how learning occurs, factors influencing learning, the role of memory, how transfer occurs, types of learning optimal for the theory, basic assumptions of the theory relevant to instructional design and structure of instruction to facilitate learning* (Dale, 1991; Ertmer & Newby, 2013).

Medical education, and particularly around emergency care, rightfully focuses on the importance of consequence of learning performance, with reinforcement being emphasised where the learning performance is gauged as sub-par (Pusic, Boutis, Hatala, & Cook, 2015). The cognitive abilities of the health worker are intended to grow towards automaticity in acquiring skills, thereby rendering the learner reactive to the learning environment when their learning peaks (Pusic et al., 2015). In the context of neonatal emergency care in SSA, this approach of deliberate practice leveraging forgetting curves lends itself well where: (1) mortality is high (implying a clear goal), (2) its causes are known through evidence-based medicine (implying close-ended solutions), (3) requires chained procedures

to avert death, but ultimately (implying conceptual-procedural knowledge required), (4) the quality of care provided is consistently poor (implying need for deliberate practice for skill acquisition). As a starting point for learning, the emphasis on sequence of instructions and cues anchored around evidence-based medicine is needed to ensure mastery of required steps for common preventable conditions (Ertmer & Newby, 2013; Nolan et al., 2001). Learning is demonstrated where all skills around a specific knowledge component are consistently applied, and transfer occurs when the same pattern of sequential procedural care knowledge is applied to other similar caregiving scenarios and is self-monitored (IntraHealth (PRIME II Project) & JHPIEGO, 2002; Pusic et al., 2015).

Crucially in LIFE's case is the conscientisation of learner's knowledge gaps, problem-solving capacity, and reflective practice: It is concerned with what learners know, how they acquire this knowledge, and how they improve upon it, with an emphasis on the digital environment's facilitation of learning (Ertmer & Newby, 2013). Specifically, the role of adaptive feedback and its influence in helping the learner exercise agency in adopting an appropriate learning strategy is emphasised. To this end, it is paramount to understand learner's experiences and context which impact learning outcomes, effectively organise and structure education content to leverage the learner's previous knowledge to increase its likelihood of assimilation (Ertmer & Newby, 2013).

Given that the learners' knowledge needs are anticipated to change with the growing familiarity of the training content -which arguably presents a case for a constructivist approach to instructional design, LIFE's approach is based on the reality that the subject matter (domain model) is well-formed, the requirements of the tasks to be learned are evidence-based, tightly regulated, and already externally validated (Ertmer & Newby, 2013), and the content should be consistent to all clinical carders involved in neonatal emergency care. It, therefore, combines both cognitivist and constructivist underpinnings to learning.

3.3 Self-Regulated Learning theory and adaptive learning on digital platforms

Intelligent Tutoring Systems (ITSs) that use adaptive learning approaches target knowledge representation in learning, and have historically taken on a constructivist approach to knowledge formation and dynamic updating, usually assuming the ability to transfer knowledge to similarly patterned multiple contexts, the ability to represent abstractions, and the ability to represent expertise (Saad & Zaghloul, 2002).

These adaptive learning mechanisms in ITSs are designed to manage and regulate learning experiences within a specified domain, and a confined pedagogical approach and are *effective* in helping individuals gain new knowledge and learn problem-solving procedures by emphasising metacognitive learning strategies (S. Lajoie et al., 2014). Recently attempts to enhance adaptive learning mechanisms within ITSs has been focused on the inclusion of tools and methods that promote ***Self-Regulated Learning (SRL)*** by incorporating strategies linked to metacognitive awareness and regulation (S. Lajoie et al., 2014). To this end, these adaptive mechanisms implicitly target (1) declarative and procedural knowledge dealing with the interaction between knowledge of the level of ability and the strategies used to successfully perform learning tasks given their nature, i.e. meta-cognition and (2) actions related to “...goal selection, planning, monitoring, control, and reflection...” of the learning process, i.e. regulation (S. Lajoie et al., 2014). This adaptive-learning on digital platforms lends itself well to SRL theories (as explained from Chapter [3.3](#) to [3.5](#)), which I adopted as the theoretical framework to guide the design and implementation of adaptive feedback.

Self-Regulated Learning (SRL) conceptual frameworks are used to understand how learner apply their *will* (motivation) and *skill* (strategies) to effectively learn. Self-regulation is a process that is initiated by learners to control their learning through: (1) using cognitive strategies to acquire, store, and retrieve of information (e.g., elaboration); (2) using cognitive strategies to plan, monitor and regulate their learning process in order to accomplish a goal; (3) using resource management strategies to manage time, help, effort etc. when organising one’s study (Pintrich, 2000). Learning is viewed

considering learners' control over their own behaviour guided by personal reflection on their own thinking, strategic approach to learning tasks, and motivation to learn. SRL is a socio-cognitive perspective, where the situational context influence on self-regulation of learning is directed at individual regulation; ***the individual is the analysis unit.***

Feedback is considered the key mechanism for SRL behaviour where the generation and processing of internal feedback and external feedback respectively by learners are useful for monitoring engagement and attainment of their learning goals for specified learning tasks (Butler & Winne, 1995). Learners are posited to *actively* manage their own learning and monitor it mainly through cognitive strategies facilitated primarily through feedback. Distinctions between personal goals and current performance are acknowledged and profiled. Situational, contextual and motivational aspects of learning are tracked over time and in real-time, using feedback for supporting agency in learning (Panadero, 2017).

LIFE, designed to be a single player-single user gamified learning platform, lends itself well to this conceptual framing and thus I focus on SRL, which is a socio-cognitive model that tends toward socially mediated cognition and motivation at the individual level. However, it is my assertion that given that LIFE is a complex intervention, differing approaches in different learning theories (such as socio-cultural and socio-cognitive approaches), are invariably interwoven together in this case, and can be seen to be in productive tension – a mixture that is not uncommon given “*...no single learning theory has enough explanatory and predictive power to inform the range of practices found in medicine...*” (Bleakley, 2006; Dearing, 2009).

And while there are many variants of this conceptual model, I adopted *Winnie and Hadwin* SRL perspective (Winne & Hadwin, 1998) in this section to address how, to learn better, users are posited to adapt to the technology-enhanced learning environment. This is because it focuses on the interrelation of cognitive and motivational processes at the individual level while accounting for volitional strategies in the performance – which is reasonably aligned to the aims and claims of

adaptive mechanisms use on digital platforms. ***Learner's cognitive processes generate experience that learner can inspect***, e.g. level of effort applied and whether successfully (Winne & Hadwin, 1998). Also, these processes can be self-controlled (agency in regulation). ***Cognition is theorised to have limited capacity*** making learners to bundle learning content into step-wise declarative knowledge (Brünken, Seufert, & Paas, 2010); this, in turn, makes the learners prone to making errors in complex or time-limited tasks (e.g. delivery of emergency care in new-borns) and they respond by becoming motivated to regulate which content they work with to avoid being overwhelmed (Winne H.P., 2011). ITSs in such an instance functions as a More Able Partner (MAP) to match and effectively challenge the cognitive capacity while balancing complex learning tasks (Luckin, 2009, 2010).

Additionally, procedural setup representing knowledge that accomplishes tasks (e.g. steps to resuscitate a new-born) and declarative format for representing knowledge (which can be probabilistic) represent a form of cognition. "...A standard trajectory transforming declarative knowledge for action (a list of steps) into ...actions that absorb cognitive resources (at novice stage) but transitions to automated action (expertise) where there is little demand on limited cognitive resources ***require extensive, deliberate practice with feedback...***" (Winne H.P., 2011). Put succinctly, how learners think reflects what they know, and directly influences oncoming performance which is automatically interpreted in relation to prior knowledge (Winne H.P., 2011). Cognition changing from implicit to explicit is exposed in the presence of errors, and traceable by learning events across time (Winne H.P., 2011). This imposes a knowledge measurement strategy that recognises the latent and observable cognition in learning tasks, a weighted effect of knowledge level by frequency and sequence of errors and that this operates as a continuous learning stream.

This research considered evidence from published literature on SRL strategies which are expected to enhance learning on digital learning environments (Boud, 2013; Schunk, 2005; Zimmerman, 2000; Zimmerman & Pons, 1986). In combination with Winnie and Hadwin's SRL model, which characterises

SRL as being weakly recursively (*through feedback*) and sequenced (Winne H.P., 2011) these SRL strategies illustrated in Figure 3.3.1 include:

1. Task definition: Learning tasks are in a context of affordances and constraints, either external (e.g. whether a goal is specified, time available, and source of help support) or internal (e.g. including knowledge about the topic, motivation to engage with it, predicted consequences of [un]successful task completion, and available task strategies).
 - a. Task strategy: Arrangement of study tasks and activities to ensure maximisation of learning process and outcomes and the effort used in learning (Effeney, Carroll, & Bahr, 2013; Richardson, Abraham, & Bond, 2012; Zimmerman & Pons, 1986);
2. Goal setting: Goal setting: Framing of learning objectives to gauge the required effort needed to realise them; Learners intensely but implicitly use this to weigh their hopefully increasing level of understanding.
3. Engagement: Learners' use of tactics to gauge the achievement of learning goals they've set for themselves, monitored through attributes such as effort, time and utilisation of instructional cues provided.
 - a. Self-evaluation: How learners actively monitoring and judge their own progress and performance against personal learning objectives;
 - b. Help-seeking: How learners are soliciting help from fellow students, informational resources or a tutor (Aleven, McLaren, Roll, & Koedinger, 2006).
4. Metacognitive adaptations: Learners' reflect on the success of the strategies used, with the objective being their satisfactory transfer to similar learning tasks in the future.
 - a. Strategic learning: They schedule, learning activities, spacing and ordered accomplishment of learning tasks needed to realise learning goals;
 - b. Elaboration: They establish meaning from the learned knowledge concepts by combining new knowledge the learner gained with what they already knew (Niemi, Nevgi, & Virtanen, 2003; Weinstein, Acee, & Jung, 2011);

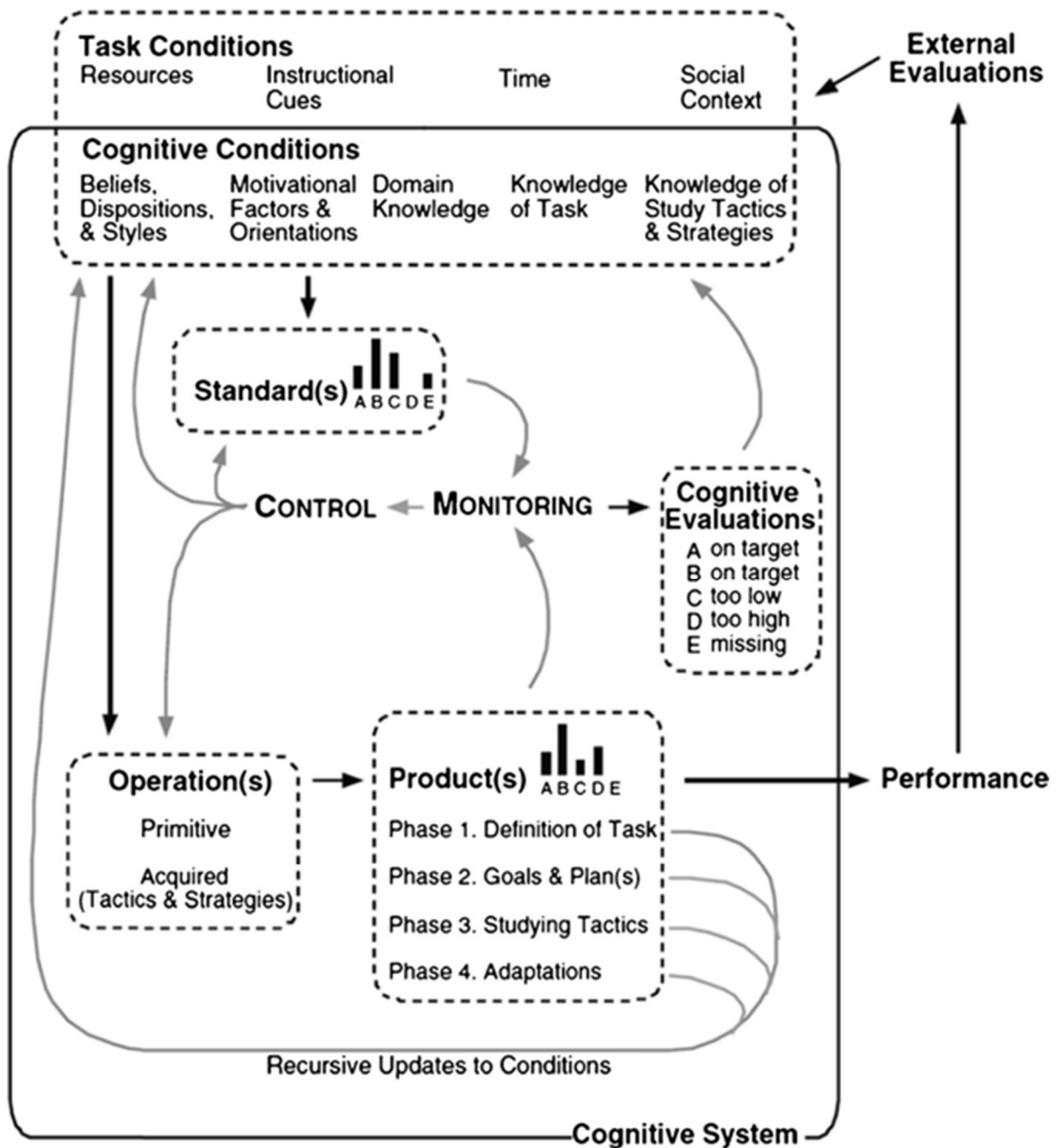


Figure 3.3.1 Winne and Hadwin SRL model (Winne & Hadwin, 1998)

LIFE's gamification instructional design derived from the context of clinical practice in SSA and based on evidence-based findings from empirical research is designed to promote learning based on learners cognitively mediated strategies by "...application of knowledge and skills to choose whether and how to use features of instructional design to optimize achievements relative to their goals..."(Winne et al., 2006). This echoes adaptive learning's need to reflect learner's application of cognitive strategies

by fully embedding modelling approaches and pedagogical strategies that reflect how learners use continuous feedback loops to regulate resources and emotions when learning (S. Lajoie et al., 2014).

Throughout this thesis, (meta)cognition has only been referenced insofar as it relates to self-regulated learning as a key building block within this learning theory. In relying explicitly on the framing of cognition as conceptualised by the Winnie and Hadwin's SRL model (Winne et al., 2006), the rest of this thesis emphasises the evaluation of SRL behaviours, and explores which SRL skills of healthcare providers on digital learning platforms might need fostering: It does not focus on exploring the relationships between metacognition and domain level learning within SRL as this exploration presumes that the relationships between metacognition, motivation and affect have both been established (qualitatively or quantitatively) and can be validly measured. That is not the case in the context of healthcare training in LICs. The groundwork for empirically evaluating such relationships within SRL in the LICs context is yet to be laid. Instead of delving into an in-depth analysis on how metacognition in general is linked to AIED, of primary importance for this thesis is establishing how SRL behaviours manifest and are perceived by the LICs healthcare providers, both qualitatively and quantitatively, which would *ideally* serve as the starting point for any future research that aims at delving deeper into metacognitive evaluations.

In LIFE game, the speculation is that SRL scaffolding through employing an adaptive mechanism is achievable by focusing on three important families of SRL strategies: monitoring (e.g. cognitive evaluations such as goal setting), information-seeking behaviour (e.g. instructional cues such as hints and more information), and standards (e.g. probabilistic determination of domain and task knowledge). Given the nature of the learning content (neonatal emergency care), the ultimate learning goal is assumed (and prescribed) to be a perfect score (100%) in the shortest time without any support from the system, repeatedly. Data collected from piloting LIFE with a bit of tweaking can be leveraged to investigate SRL role in scaffolded learning.

Insofar as feedback is the fundamental mechanism for fostering better SRL behaviours (Butler & Winne, 1995), it is posited to enhance learning when “...it provides further information to correct or modify action through the construction and activation of more appropriate [action sequences]...” (Greenwood, 1995). The information provided ought to move the learner to a deeper understanding of the learning task (Rn, 2000). However, more feedback does not always equate to better learning: “...The amount of information given to the student must be what the student can use, rather than the amount the [tutor] may wish to give...” especially in light of the limited capacity of cognition (Brünken et al., 2010; Fournier, 1996; Lam, DeRue, Karam, & Hollenbeck, 2011; Rn, 2000).

Feedback that is too elaborate is more likely to produce cognitive overload. On the other hand, if it has low specificity, it is more likely to be perceived by learners as useless (Brünken et al., 2010; Voerman, Meijer, Korthagen, & Simons, 2012). Effective feedback is posited to be specific but not too elaborate and presented in manageable units (Shute, 2008; Voerman et al., 2012). Its timing, specificity, frequency, and type has varying effects in enhancing learning (Lam et al., 2011)

Within ITSs literature, different feedback levels and purposes have been highlighted, and example being illustrated in table 3.3.1 below.

Table 3.3.1: Common levels of feedback within interactive digital learning platforms, <i>borrowed from (Holmes et al., 2015)</i>			
Level	Feedback	Purpose	Example
1	Socratic	To encourage students to verbalize possible solutions.	<i>“Have you changed the numerator or denominator?”</i>
2	Guidance	To remind students of key domain specific rules or of the system's affordances.	<i>“The denominator is the bottom part of the fraction. *”</i>
3	Didactic Procedural	To specify a next step in procedural terms.	<i>“Check that the denominator, the bottom part of your fraction, is 12.”</i>
4	Didactic Conceptual	To specify a next step, in terms of the concept currently being explored.	<i>“Check that the denominator in your fraction is correct.”</i>

From the explanation of the LIFE platform provided in Chapter 1.3 and Appendix A, its feedback primarily operates at the *didactic procedural* and *didactic conceptual* level (Table 3.3.1). The provision of in-context, real-time feedback to learners is not uncommon in ITSs studies. For example, feedback

within ITSs has been used to help learners to better self-regulate and achieve domain-level mastery (Aleven, Roll, McLaren, & Koedinger, 2016). While on-demand, principle-based feedback during tutored problem solving is no longer highly regarded within ITSs, its utility in learning adaptive environments for healthcare providers in LICs is yet to be ascertained. This type of feedback is still posited to be helpful under certain circumstances, with its usefulness being subjective, platform-dependent and domain-specific (Aleven et al., 2016).

Different approaches to gauging the usefulness of feedback have been highlighted prominently in ITSs literature, some of which include:

1. Data-driven methods: This approach uses a combination of machine learning and artificial intelligence approaches to automatically select the best feedback strategy for learners given the learning content and context (Gross, Mokbel, Hammer, & Pinkwart, 2015; Gutierrez & Atkinson, 2011).
2. Elicitation methods: These methods aim to elicit and evaluate the learners' engagement with feedback, including in terms of their cognitive or affective responses to the feedback received from their (intelligent) tutors (Fernández-Toro & Furnborough, 2014; Porayska-Pomsta, 2016).
3. Experimental approaches: This approach is more focused on developing a theory of tutoring which would link the various learning mechanisms to the types of information the learner needs and hence what the ITSs should provide (Barrow, Mitrovic, Ohlsson, & Grimley, 2008; Delaval, Michinov, Le Bohec, & Le Hénaff, 2017).

These different approaches are not necessarily mutually exclusive: there exists examples of knowledge elicitation methods within ITSs which combine data-driven and elicitation methods (Chi, Koedinger, Gordon, Jordon, & VanLahn, 2011; Gutierrez & Atkinson, 2011; Porayska-Pomsta, 2016). The list above is neither comprehensive nor representative of the landscape of how feedback has/might be used and evaluated in ITSs literature: there are as many approaches as there are experts (Carenini et al., 2014; Holmes et al., 2015; Molloy, Boud, & Henderson, 2020).

Where the consensus lies is in the centrality of feedback in facilitating self-regulated learning on digital learning environments, that the feedback's timing, content, and presentation are pivotal to enhancing learning outcomes on these platforms, and that their utility in aiding the learning process needs to be empirically evaluated in order to allow better matching of feedback to the learner's context and needs.

Given I have already made a case for the necessity of exploring how lurking and confounding factors might moderate or mediate gamified learning effects (Chapter 2.1.1), and that the dearth of application of learning theory in digital-based learning in clinical education (Chapter 2.1.2), and the scarcity of evidence (and data) on digital-based healthcare training in LICs (Chapter 1.1, Chapter 2.1), experimental approaches lend themselves well to exploring feedback approaches in LIFE context.

3.4 Approaches to measuring Self-Regulated Learning in adaptive platforms

Given that LIFE content material consists of a sequenced chain of activities to mimicking steps needed to deliver neonatal emergency care, primary from deliberate, repeated practice, consider the following:

*Suppose a healthcare provider has six opportunities to apply a **knowledge component** in LIFE™ and makes the following sequence of correct (1) and incorrect (0) responses.*

0 1 0 1 1 0

Has the healthcare provider learned the care-giving rule?

Box 3.1: Knowledge tracing evaluation challenge for a single task

Knowledge Components (KC) – which have been explained in detail before in Appendix A.1- can represent fine-grained learning "...concept, principle, fact, skill, schema, production rule, misconception..." (Kurt VanLehn et al., 2007). An expansion of Box 3.1 use-case scenario could be illustrated in the following way:

Suppose a healthcare provider has six opportunities to apply to each of three (in)dependent knowledge components in LIFE™ and makes the following sequence of correct (1) and incorrect (0) responses.

KC1: 0 1 0 1 1 0

KC2: 0 1 1 0 1 0

KC3: 1 0 0 0 1 1

Has the healthcare provider mastered the care-giving procedure?

Box 3.2: Knowledge tracing evaluation challenge for multiple tasks

As a learner answers exercises of a given concept correctly or incorrectly, measurement strategies, that can update knowledge probabilities with the outcome of each subsequent learner response, are preferred due to the lack of local independence given the conjunctive nature of student-step within the knowledge component. This formulation of the sequence (Box 3.1) considers whether a skill has been learnt from observed performance, and also seeks to contextualise learning characteristics such as guessing and slipping, apart from estimating problem difficulty (Corbett & Anderson, 1994). This skill assessment stems from an understanding that the cognitive processes generate the sequences in Box 3.1-3.2 which can be used to evaluate whether the learning processes are building the learner's *competence* relative to concept *relatedness* (Ryan & Deci, 2000; Winne H.P., 2011).

LIFE implicitly captures aspects of a student's knowledge mastery over a sequence of conjunctive ordered tasks, and the aim is to map and predict the learner input sequence $(x_1, x_2, \dots, x_T)$ to an output sequence $(y_1, y_2, \dots, y_T)$. That is, when a student attempts a learning task (step), a record is kept of whether the attempt was successful (y) linked to the learning task (x). Generally, the knowledge tracing task can be formalized as follows: given a learner's historical interactions $X = x_1, x_2, \dots, x_T$ up to time t on a Knowledge Component (KC), it predicts some aspects of their next interaction x_{t+1} (Corbett & Anderson, 1994).

Such a learning conceptualisation is common in the ITSs and adaptive learning literature, where while not necessarily cognisant of contexts like SSA, learner interactions within the digital learning platform tend to be tracked as a sequence of student-driven steps (Kurt Vanlehn, 2006). That is, when a student attempts a learning task (step), the ITS records whether it was successful, whether any system-initiated assistance was provided, and may provide instructional support based on the learner's performance. Student-step data are important in breaking down the level of skill mastery in dealing with any emergency care rapid response and specifically what constitutes skill mastery. These kinds of data points are what are used for student learning needs' modelling and subsequently adapting content (Chi et al., 2011). Inclusion of such metrics in reporting the rate of progress in gaining experience or new skills (i.e. learning curves) and learning outcomes in digital-based clinical training interventions is rather sparse (Mieke Vandewaetere et al., 2013). Such metrics would arguably help inform the successful implementation of adaptive learning on digital training platforms to help bridge the skills gap in clinical care provided in low-income settings where low-cost highly accessible tailored training opportunities are hard to come by.

Learning is governed by diverse properties -as outlined in my choice of the SRL model- many of which are difficult to quantify using observable exercises. Here, measurement is predicated on applying measurement strategy to the problem of predicting learner responses to exercises based upon their recent learning activity. Generally speaking, the interest is in piloting two types of predictions: (1) how unseen learners will perform on the observed learning content (*validity of the theoretical perspective, generalisable to population*), and (2) how the same learners seen in the observed data will perform on unseen content (*validity of the theoretical perspective, generalisable to learning content*) (Chi et al., 2011). Popular knowledge tracing techniques (i.e. measurement strategies) used for adaptive learning instruction based on locally dependent knowledge components such as those highlighted in Box 3.1-3.2 are discussed in the remainder of Chapter 3.4. Knowledge tracing in its general form is formulated as supervised learning problem of a series of ordered events where the goal is to infer a

latent state ('Knowledge state') for a specific skill being learned and predict the future performance of the individual learner (Sapountzi A., Bhulai, Cornelisz, & van Klaveren, 2018).

3.4.1 Bayesian Knowledge Tracing Details

Bayesian Knowledge Tracing (BKT)(Corbett & Anderson, 1994) is a learning modelling method extensively used in the area of Intelligent Tutoring Systems (ITSs). It is used to model student knowledge from student performance within any educational tool, where knowledge is considered a latent variable. BKT epistemic assumptions are:

1. Student knowledge is not directly assessable or observable. Student performance is.
2. Student behaviour on a specific task can be assessed as either being correct or incorrect
3. Each step/task, i.e. *problem*, is associated with one clear skill, i.e. *knowledge component*
4. A student does not forget a skill, once he or she knows it*

The goal of using BKT is to infer whether a student has mastered a skill from a sequence of right/wrong answers to tasks (or questions) that require that skill. The latent variable (*knowledge*) is updated every time the student has an opportunity to use the skill. Using this approach enables this research to test the accuracy of the prediction of future knowledge-based assessments on a post-test evaluation. It also assumes that a learner's knowledge is adequately represented as a Bernoulli distribution.

In adding context to the problem illustrated in Box 3.1, BKT considers the following two scenarios when estimating the belief that the student has acquired the skill:

- A. The probability that the student will *slip* and make a mistake when the student knows a skill
- B. The probability that the student will *guess* correctly when they do not know a skill.

BKT, therefore, is used to trace four skill learning parameters for each skill according to the Corbett and Anderson's model (Corbett & Anderson, 1994) outlined below:

- ❖ **pKnown (pL)** – running estimate of skill mastery, the probability the skill is already known before the first opportunity to use the skill in problem-solving.

- **pInit (pL_0)** – A priori mastery of skill, i.e. the initial probability of knowing the skill a priori. Can be supplied using a pre-test or auto-generated from within the educational tool.
- ❖ **pLearn (pT)** – probability the skill will transition into “known” state, i.e. skill will be learned at each opportunity to use the skill
- ❖ **pSlip (pS)** – probability known skill is applied incorrectly, i.e. the student will make a mistake if the skill is known
- ❖ **pGuess (pG)** – probability not known skill is applied correctly, i.e. the student will guess correctly if the skill is not known.

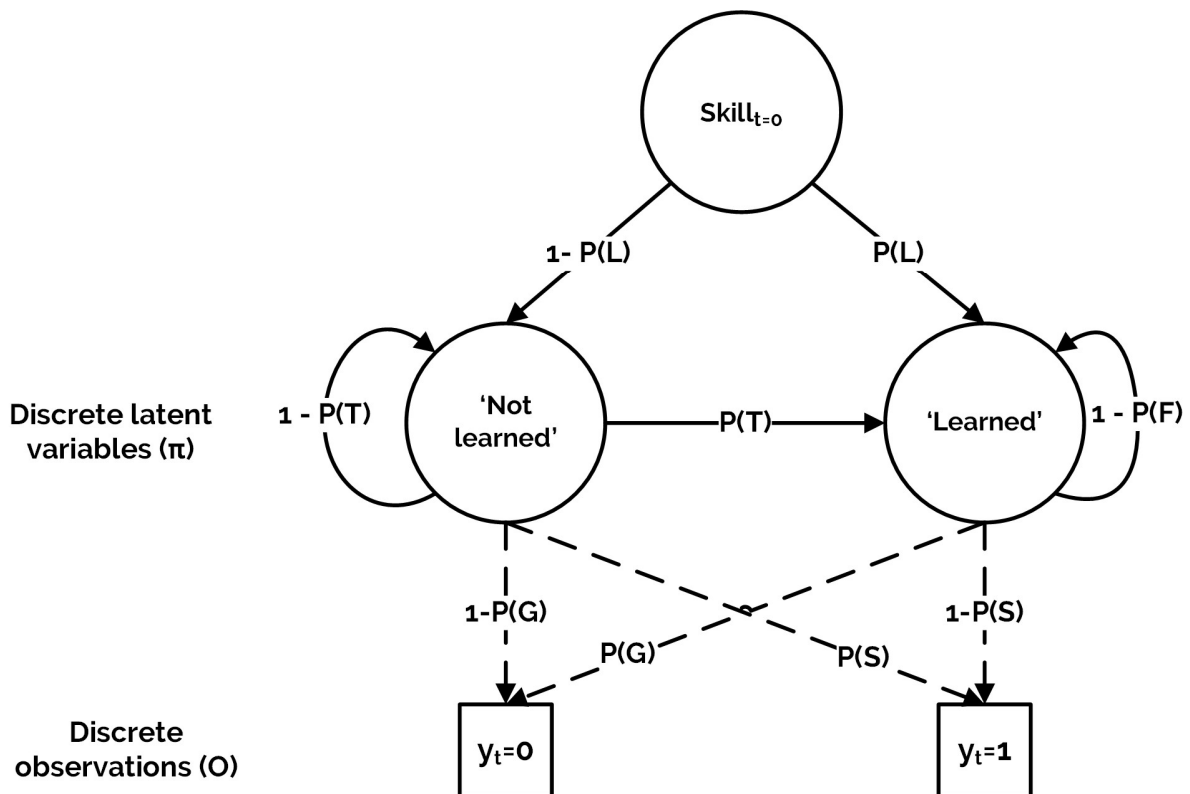


Figure 3.4.1: Bayesian Knowledge Tracing model illustration

Additional details about the BKT model can be found in [Appendix C.2](#). The BKT model assumes the latent states to be independence among the different skills in the learning module. Even within an individual skill, BKT assumes the probability of the next learning outcome to depend only on the latest previous outcome (The Markov assumption). Additionally BKT assumes that forgetting does not occur;

its typical implementation does not allow for learners to have different learning rates; it assumes that all students have the same probability of knowing a particular skill at their first opportunity (Pardos & Heffernan, 2010); and it suffers from the problem of multiple global maxima when trying to estimate model parameters (Gong, Beck, & Heffernan, 2011). While multiple-restarts estimation methods like Markov Chains Monte-Carlo estimations can help minimise local maxima challenges in estimating BKT model parameters, it is hard to eradicate the identifiability problem where given the same model structure and same data, there are multiple (differing) sets of parameter values that fit the data equally well (Gong et al., 2011). This is because in probabilistic terms, to find the global solution requires the model to iterate over all possible learner sequences. However, iterating over all learner sequences of varying lengths in order to maximize the objective function makes the number of iterations exponential to the size of the learning data, rendering it practically intractable.

To address some of the shortcomings of BKT models, Learning Factors Analysis (LFA) and their different modalities such as Performance Factors Models (PFMs) and Additive Factor Models have been proposed (Cen, Koedinger, & Junker, 2006, 2008; Pavlik Jr, Cen, & Koedinger, 2009). This is explained in detail in the next section. They model student knowledge states using logistic regression models in order to deal with the incorporation of multiple skills while estimating student ability into the model. They exploit the number of successes or failures of a learner's attempt at a KC to predict whether the learner has acquired understanding about the KC.

3.4.2 Additive Factor Models (AFMs) and Performance Factor Models (PFMs)

AFM is used to evaluate conjunctivity of skills in learning data. Its **additive** nature is due to a linear combination of skill parameters determining p_{ij} described in the equation below:

$$\ln \frac{p_{ij}}{1-p_{ij}} = \theta_i + \sum_k \beta_k Q_{kj} + \sum_k Q_{kj} (\gamma_k N_{ik}) \quad (3.4.1)$$

Where:

i represents student i

j represents step j

k represents knowledge component/skill k

p_{ij} is the probability that student i would be correct on step j .

θ_i is the coefficient for the proficiency of student i .

β_k is coefficient for difficulty of the knowledge component or skill k

Q_{kj} is the Q-matrix cell for step j using skill k .

γ_k is the coefficient for the learning rate of skill k

N_{ik} is the number of practice opportunities student i has had on the skill k

Q-matrices are used to represent the relationship between individual steps and knowledge components, typically encoded as a binary 2-dimensional matrix with rows representing knowledge components and columns representing steps (Chi et al., 2011). AFM posits that the probability of a learner getting a step correct is proportional to the amount of required knowledge they already know, together with skill difficulty and amount of learning opportunity they have been already exposed to (Cen et al., 2008). On the other hand, PFMs given by the equation below, seek to predict performance on the current item using the entire history of success and failures on previous items addressing the same student step (Galyardt & Goldin, 2014). It estimates the different effects of practising learning opportunities.

$$\text{logit}(p_{ijt}) = \theta_i + \beta_j + \alpha_j S_{ijt} + \rho_j F_{ijt} \quad (3.4.2)$$

Where:

i represents student i

j represents step j

θ_i is the coefficient for proficiency of student i .

β_j is coefficient for difficulty of the step j

X_{ijt} binary correct/incorrect, student i , step j , trial t

S_{ijt} count of previous success, up to trial t

F_{ijt} count of previous failures, up to trial t

p_{ijt} is $\Pr(X_{ijt} = 1)$

Due to PFA's linear structure, it may still yield implausible parameters, e.g. estimating that practice on a skill is associated with a decrease in the probability that the learner will correctly answer a problem on that skill: To address this challenge, parameters are artificially restricted (Gong et al., 2011). Additionally, using AFMs and PFMs with a small set of student-step data from a few learners might be challenging (and prone to overfitting) due to their need for large search space for parameterisation. This overfitting might be due to linear dependence (if the model matrix is of deficient rank), or from having too many predictors, i.e. small sample bias.

Also, although LFA-based models handle a learning task that is associated with multiple KCs better than the BKT approach, they are ill-equipped to deal with varied inherent dependency among KCs that are hierarchically related in the same learning content (Pavlik Jr et al., 2009). Furthermore, PFMs tracking of prior successful or failed practices counts disregards the ordering of those practices (Gong et al., 2011). AFMs, on the other hand, assume that all learners accumulate knowledge in the same manner while ignoring the outcome of their individual responses during each practice session (Chi et al., 2011).

AFMs and PFMs while highly predictive, are argued to be unable to optimise *subsets* of instructional content presented to learners according to their historical performance (Sapountzi A. et al., 2018) and are difficult to implement online let alone offline (Pelánek, 2016). Additionally, between BKTs, AFMs and PFMs, which of them is comparatively better for student modelling remains unclear and is posited to be context-specific (Chi et al., 2011). With these considerations in mind, I opted to use BKT as the mechanism for determining at what point to offer adaptive support- because of its ability to overcome the challenge of offline use of the model for prediction of performance using individual learner's data. This was achievable because a pre-trained standard BKT model can be implemented using a decision-tree structure that adjusts model parameters for each individual learner while completely offline (Figure 3.6.1).

3.4.3 Knowledge Tracing and Item Response Theory

AFMs and PFMs in nature are similar to parametric Item Response Theory (IRT) models (Cen et al., 2008). Like IRT models, they assume that there is one dominant latent trait being measured and that this trait is the driving force for the responses observed for each item in the measure (i.e. unidimensional); that as this latent trait level is increasing, the probability of a correct response also increases (i.e. monotonicity). However, they differ to IRT in two substantive ways: (1) Unlike IRT models, they do not strictly adhere to the local Independence assumption (i.e. Responses given to the separate items in a quiz are mutually independent given a certain level of ability), even though their attempts at representing local dependence is weak (Pavlik Jr et al., 2009).

IRT models are typically suited for coarse-grained skills where the learning process is slow (i.e. summative), making them undesirable for adaptive online learning purposes which would benefit more from fine-grained formative type assessments (e.g. using student-step data) (Sapountzi A. et al., 2018). Additionally, IRT models are inherently cross-sectional and typically, aim to explain observed differences in what has been learned; They are argued to be ill-suited for explaining how observed differences in learning between students came into existence, or what measures could reduce or alter these observed differences (Deonovic, Yudelson, Bolsinova, Attali, & Maris, 2018).

3.5 Scaffolding instructional support in Self-Regulated Learning

3.5.1 Zone of Proximal Development, the flow channel and adaptive learning

Zone of Proximal Development (ZPD) as a socio-cultural approach is based on a developmental process where through interactions between the learner, cultural environment, knowledge is internalized; However, it is conceivable that the cultural environment can be substituted for a digital environment when coupled with a More Able Partner (MAP) (Luckin, 2010, pg. 20). ZPD's possible implementations on digital platforms have been implemented in modalities such as (Luckin, 2010, pg. 27-29; Wood, Shadbolt, Reichgelt, & Wood, 1992):

- a. Instructional contingency: At what level will the next instructional intervention be.
- b. Domain contingency: Given the content the learner selected; what support should be provided.

c. Temporal contingency: How long should the learner be left struggling

ZPD is negotiated between the learner and a MAP such as a teacher, building on practical evidence from previously empirically tested learning technologies (Luckin, 2010, pg. 27-29; Wood et al., 1992), LIFE's adaptive mechanism had to be designed to mimic such a MAP in order to provide a context for a flow channel while trying to match the learner's passage with the appropriate amount of assistance (Kulik & Fletcher, 2016) based on learning progression as illustrated in Figure 2.1.2.

This MAP, conceptualised as a knowledge tracing model, has to adapt instructional support combined with a temporal contingency in the form of adaptive feedback provided to the learner based on their assessed mastery level, in order to encourage learner progression (Luckin, 2009, 2010; Luckin & Du Boulay, 1999).

SRL is a weakly recursively process that is facilitated by feedback at each stage (Butler & Winne, 1995; S. Lajoie et al., 2014; Winne & Hadwin, 1998) supporting learner's cognitive regulation of learning strategies to help them regulate resources and emotions while learning (S. Lajoie et al., 2014). In addition to regulation, cognition consists of declarative knowledge and deals with the interplay between knowledge of one's abilities to perform tasks, the nature of the learning task, and the strategies one can employ to successfully perform the task (S. Lajoie et al., 2014). It is theorised to have a limited capacity which renders the learners prone to making errors in complex or time-limited learning tasks (e.g. delivery of emergency care in new-borns) (Winne H.P., 2011). In the presence of such performance errors (e.g. slipping or guessing) due to cognition's limited capacity (Brünken et al., 2010), to guide SRL, a key objective becomes inference of the knowledge of the learning task being tutored. Feedback is posited to enhance learning when "...it provides further information to correct or modify action through the construction and activation of more appropriate [action sequences]..." (Greenwood, 1995). The information provided ought to move the learner to a deeper understanding of the learning task (Rn, 2000). However, more feedback does not always equate to better learning: "...The amount of information given to the student must be what the student can use, rather than the

amount the [tutor] may wish to give...” especially in light of the limited capacity of cognition (Brünken et al., 2010; Fournier, 1996; Lam et al., 2011; Rn, 2000). Feedback that is too elaborate is more likely to produce cognitive overload. On the other hand, if it has low specificity, it is more likely to be perceived by learners as useless (Brünken et al., 2010; Voerman et al., 2012). Effective feedback is posited to be specific but not too elaborate and presented in manageable units (Shute, 2008; Voerman et al., 2012). Its timing, specificity, frequency, and type has varying effects in enhancing learning (Lam et al., 2011). In the absence of flexibility in determining the instruction challenge level or stratifying learning pathways within the LIFE project (due to the efforts to standardise delivery of clinical guidelines training content), feedback remains the most promising intervenable theory-aligned concept for enhancing SRL using smartphones in settings such as SSA. The choice of adapting feedback based on learner progression and inferred level of knowledge is anchored on the centrality of feedback in sustaining SRL (Chapter [3.3](#)).

From previous studies (Chounta, Albacete, Jordan, Katz, & McLaren, 2017), uncertainty about student performance is high when a student model predicts the learner’s ability to make a successful attempt with a probability around 0.5. This probability threshold region, where the prediction error is considered high, has been referred to by one study as the “Grey Area” and conceptualised to approximate the student’s ZPD. The reasoning for selecting these thresholds as analogous to targeting the ZPD mechanism is that:

“...The space ‘Above the Grey Area’ denotes the region where the student is predicted to answer correctly (the fitted probability is considerably higher than the cut-off threshold) and consequently may indicate the area above the ZPD; that is, the area in which the student is able to answer a question without any assistance. Accordingly, the space ‘Below the Grey Area’ denotes the area where the student is predicted to answer incorrectly (the fitted probability is considerably lower than the cut-off threshold) and consequently may indicate the area below the ZPD; that is, the area in which the student is not able to carry out the task either with or without assistance...” (Chounta et al., 2017)

Admittedly, the symmetry and size of “the Grey Area” could change depending on the classification threshold and the learning objectives. The role of the MAP can be conceptualised as effectively managing the areas in and below “the Grey Area” in order to produce knowledge mastery. From previous work, evidence of mastery by which the MAP may also use to determine the level of adaptive support or instructional complexity to provide the learner with includes: the probability that a student grasped a concept is ≥ 0.95 , **OR** probability of getting a series of attempts correct in a row, **OR** the lack of “wheel spinning” in the attempts made where wheel spinning is defined as the learner spending a considerable amount of time or effort struggling to learn a concept without gaining any mastery (Beck & Gong, 2013). The assistance provided based on learner progression is predicated on the belief about the *current* level of knowledge on the subject matter of the learner, using a measurement approach such as those discussed in Chapter [3.4](#).

Bearing in mind that no single learning theory has enough explanatory and predictive power to inform the range of practices found in health sciences (Bleakley, 2006; Dearing, 2009), as an ongoing debate, this thesis seeks to evaluate which measurement approach is of more practical significance in reflecting how LIFE is used by learners as opposed to pitting different theoretical and measurement positions.

3.5.2 Motivation within SRL-driven feedback scaffolding

Motivation is a process that sustains the directed drive and persistence of learning behaviour (Bergin, Ford, & Hess, 1993). Its strength and orientation are assumed to be indicative of “reason for action” (Ryan & Deci, 2000). Where there are manipulable elements of learning context constructed around the learning tasks, motivation has a strong impact on inter-learner and temporal learning experiences (Luckin, 2010, pg. 30-31). This is in line with a cognitive-motivational model of learning: students' learning strategy which determines the persistence and effort they put into learning is influenced by cognitive resources needed for learning (Valle et al., 2003). Scaffolding role in this context is for augmenting the learner's strategy by balancing the task challenge level with manipulable feedback support to maintain an optimal balance between the level of skill mastery and challenge within LIFE.

In this regard, it is arguable that motivation within such environments encapsulates within it, both quantitative attributes (time spent on platform/task, frequency and level of hints and feedback requested etc.) and qualitative attributes (perceptions of time-distortion, loss of self-consciousness, sense of control, self-efficacy beliefs etc.), especially where learners interpret their environment in such a way as to maintain a positive self-image in light of their own successes/failures in the learning process, a view that is borrowed from attribution theory (Weiner, 1974). Combining intrinsic motivation -which is motivation derived through the learning activity itself- with the extrinsic motivation of ability to adequately deliver new-born emergency care in the future is thus essential to develop motivating serious game (Drummond et al., 2017).

Within LIFE, the aim is to ensure that learners encounter tasks in a context of external affordances and constraints (e.g. presence of a goal and access to help, time available, etc.) predicated on self-assessed/prior subject-domain knowledge, which impacts motivation to engage with these tasks, learning strategies used, and the future consequences thereof (Winne & Hadwin, 1998). Based on these theoretical conceptualisations of learning environments, when the learning content made available to the learners is aligned to their learning goals, and instructional support is provided, this is posited to impact on the motivation and behaviour of the learner. This is illustrated in Figure 3.5.1.

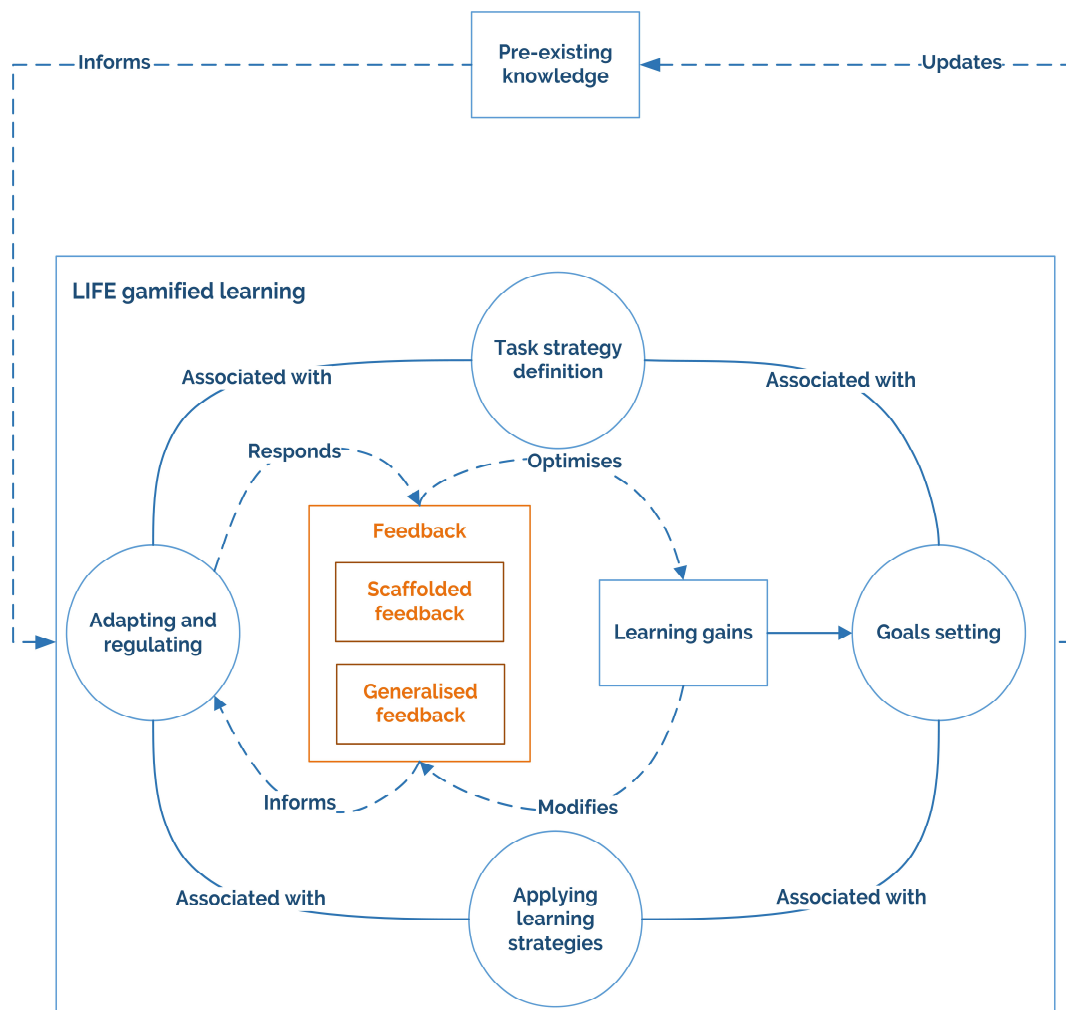


Figure 3.5.1: Preliminary conceptual model of adaptive self-regulated learning through LIFE platform

This research's theoretical position of scaffolding technique nature in improving upon learning specifies: (1) motivation to be a process that is qualitatively discernible throughout the learning process; (2) competence distinguishing performing from knowing to be central to learning – and this, I argue, is an irreducible minimum where the subject domain is of a critical nature as emergency care and therefore a source of extrinsic motivation; and (3) learning trajectories that cater to individual learner needs and supported by learning resources produce intrinsic motivational systems.

3.6 Summary of literature and theoretical review

From the detailed breakdown of the literature review and theoretical foundations for this thesis provided in chapter two and three, this thesis focuses on adaptive feedback for the following reasons:

1. SRL is a weakly recursively process that is facilitated by **feedback** at each stage (Winne & Hadwin, 1998) (Chapter 3.3.2). Adaptive digital learning environments support cognitive regulation of learning strategies by using continuous feedback loops to help learners regulate resources and emotions when learning (S. Lajoie et al., 2014). This regulation includes activities related to goal selection, planning, monitoring, control, and reflection: Feedback is the central mechanism in SRL (Butler & Winne, 1995; S. Lajoie et al., 2014).
2. In addition to regulation, cognition consists of declarative knowledge. cognition deals with the interplay between knowledge of one's abilities to perform tasks, the nature of the learning task, and the strategies one can employ to successfully perform the task (S. Lajoie et al., 2014). It is theorised to have a limited capacity which renders the learners prone to making errors in complex or time-limited learning tasks (e.g. delivery of emergency care in new-borns) (Winne H.P., 2011) (Chapter 3.3.1). In the presence of such performance errors (e.g. slipping or guessing) due to cognition's limited capacity (Brünken et al., 2010), to guide adaptive learning, a key objective becomes inference of the knowledge of the learning task being tutored.
3. Feedback is posited to enhance learning when "...it provides further information to correct or modify action through the construction and activation of more appropriate [action sequences]..." (Greenwood, 1995). The information provided ought to move the learner to a deeper understanding of the learning task (Rn, 2000). However, more feedback does not always equate to better learning: "...The amount of information given to the student must be what the student can use, rather than the amount the [tutor] may wish to give..." especially in light of the limited capacity of cognition (Brünken et al., 2010; Fournier, 1996; Lam et al., 2011; Rn, 2000). Feedback that is too elaborate is more likely to produce cognitive overload. On the other hand, if it has low specificity, it is more likely to be perceived by learners as useless (Brünken et al., 2010; Voerman et al., 2012). Effective feedback is posited to be specific but not too elaborate and presented in manageable units (Shute, 2008; Voerman et al., 2012).

Its timing, specificity, frequency, and type has varying effects in enhancing learning (Lam et al., 2011).

4. The Zone of Proximal Development (ZPD) is negotiated between the learner and a More Able Partner (MAP) such as a tutor (Luckin, 2010, pg. 27-29; Wood et al., 1992). It relies on the inferred knowledge mastery level to prescribe the learner with the appropriate amount and level of feedback, instruction or learning resource assistance (Kulik & Fletcher, 2016; Sapountzi A. et al., 2018). In the absence of flexibility in determining the instruction level or stratifying learning resource for the LIFE project (due to efforts to enforce standardised delivery of clinical guidelines training), feedback remains the most viable manipulable contingency for enhancing SRL learning using smartphones in LICs settings.
5. The fundamental principles of adaptive learning are complex but examinable through different lenses, e.g. through the lens of the tri-partite model of adaptive instruction (M. Vandewaetere et al., 2011) or the multi-dimensional model of personalisation (Holmes et al., 2018). A knowledge tracing approach (Sapountzi A. et al., 2018) that guides the delivery of tailored, differentiated feedback based on the individual learner's history of performance from previous learning opportunities arguably adheres to these principles. The work outlined here takes on a hybrid approach where both expert-based design of the optimal instructional sequence to guide learners in knowledge mastery (i.e. *designed adaptivity*), is combined with the use of algorithms to determine, for example, the extent of the learner's knowledge and the optimal feedback level to provide (i.e. *algorithmic adaptivity*) (Fontaine et al., 2019).

3.7 Justification of research

The review in Chapters [2](#) and [3](#) demonstrate that adaptive learning in a gamified context is analogous to the use of a More Able Partner (MAP) in a face-to-face learning environment. As a MAP, its focus is challenging the learner by gradually scaling back the assistance provided to learners as they progress through increasingly complex learning tasks, i.e. targeting the learner's zone of proximal development.

As such, it is linked to the learning task difficulty, and learner's performance and the type and level of support provided (Ma et al., 2014; M. Vandewaetere et al., 2011).

Despite the important role of smartphones in facilitating adaptive learning, there is still a lack of research investigating mobile-based adaptive-learning (Mousavinasab et al., 2018). Thus, there is not much evidence through empirical examples supporting how learning analytics such as knowledge tracing might be implemented in LICs context to support self-regulated learning at scale for healthcare providers on low-end smartphone devices.

Little evidence exists relating to the implementation of digital learning interventions that are relevant to the context of low-income settings, take into account health workers' initial and continuing clinical training needs, and adapt instructional support in light of skill mastery and performance as learners continue to develop knowledge (Bollinger et al., 2013; Brusilovsky & Peylo, 2003; Car et al., 2019; Edgcombe et al., 2016; S. V. Gentry et al., 2019; Greenhalgh, 2001; Ma, 2017). In non-clinical, high-resource contexts, adaptive instructional support has been shown to significantly outperform teacher-led large-group instruction, non-adaptive computer-based instruction, and paper-based instruction in producing learning gains (Fontaine et al., 2019; Ma, 2017; Ma et al., 2014; WHO, 2011). Within low-resource contexts, investigation of learner models (cognitive models that try to model observed student learning behaviours) needed to support such tailored instructional approaches in clinical training settings is underexplored and justified (Car et al., 2019; Ma, 2017). What remains unaddressed is whether for clinical training in LIC context, learning experiences and outcomes from within a gamified learning platform are comparatively better when using adaptive compared to standardised instructional support design approaches. *It does not appear that a gamified approach adopting an adaptive instructional support design has been empirically tested in a clinical context.*

Appreciating that reinforcement of knowledge and skill mastery is useful, interventions in emergency care training in low resource settings have tended towards face-to-face group training and, more recently, have used low-dose high-frequency (LDHF) in-person training in group settings (but did not

utilise technology) (Chaudhury et al., 2016; Willcox et al., 2017). Evidence to support the implementation of student-modelling approaches on digital platforms in clinical training in order to facilitate adaptive learning and improve learning outcomes is scarce (Fontaine et al., 2019), and virtually non-existent for emergency care training in low-income countries (Opiyo & English, 2015). Additionally, differences between learners' achievement goal orientations (such as skill mastery-intrinsic, mastery-extrinsic, performance-approach, performance-avoidance etc.) (Rawlings, Tapola, & Niemivirta, 2017) and how that is reflected in the uptake of smartphones-based learning approaches in low-income settings is largely unexplored. Inclusion of metrics that indicate the rate of progress in gaining experience or new skills (i.e. learning curves) and learning outcomes in digital-based clinical training interventions is rather sparse (Mieke Vandewaetere et al., 2013). Additionally, there is not much evidence of if and how the healthcare providers' length of learning trajectories used in knowledge tracing affects the ability to predict some aspects of their next interaction for a conjunctive learning space such as clinical training. There is also an insufficient empirical base that explores knowledge tracing prediction performance which explicitly addresses skill decay over time. The emphasis on skill decay over time rests in the risk implications for patient outcomes in clinical care provision.

Adaptive learning designed to manage and regulate learning experiences within a specific domain and a confined pedagogical approach posited to be *effective* in helping individuals gain new knowledge. This is shown to be due to its emphasis on learning strategies and self-regulation (S. Lajoie et al., 2014) where through guiding learner's personal reflection on own thinking, it encourages learner's control over own behaviour, strategic approach to learning tasks, and motivation to learn. Within self-regulation, a learner is offered suggestions on ways to scale their own zone of proximal development, but they have control on how to use the adaptive support provided (Luckin & Du Boulay, 1999). Though SRL driven interventions promote students' learning, they have differential effects based on the students' educational level (Panadero, 2017). And while SRL provides a quite specific picture of what its processes are, *how different learners' approach self-reflection in the presence of adaptive*

feedback, what the nature of the strategic learning adjustments they make look like, and whether it is associated with learning gains remains empirically unexplored. More so in low-income context and specifically within healthcare context settings, where the need for more better-trained health workers is urgent, how learners perceive, engage with and take control of adaptive learning on serious-gaming mobile applications remains unknown.

This research asserts that, if learning is socio-culturally situated and informed by context (Polly, Allman, Casto, & Norwood, 2017), then SRL research done in high-income countries (which are almost all) – including those especially looking into clinical context (van Houten-Schat et al., 2018)- would arguably be limited in reflecting learning behaviours from LICs. Previous studies in SRL corroborate this view (Barnard, Paton, & Lan, 2008; Zimmerman & Schunk, 2001, pg. 125) including recent evidence synthesis of global research on the use of digital learning platforms by healthcare providers (Car et al., 2019; S. V. Gentry et al., 2019).

This study seeks to contribute to the aforementioned evidence gaps and using the intervention LIFE, investigate the nature and relationship between self-regulated learning on an adaptive serious-gaming mobile platform, for clinical training in low-income contexts. Following on from the summary given in this chapter, and in line with the literature review, four overarching research questions are provided in the following section.

3.8 Research questions

The overarching aim of this thesis is to explore the potential of gamified smartphone-based learning platforms utilising adaptive instructional support strategies to provide directed self-regulated learning of emergency neonatal care delivery to health workers in low-income settings. Using LIFE which is designed to train health workers in low resource settings, the research questions I want to address are:

- 1. Within a gamified smartphone-based learning platform for clinical training, what is the effect of adaptive feedback on learning gains compared to standardised feedback?**

Particularly, how is this effect moderated by (a) the opportunity for knowledge mastery, and (b) the level of adaptive support provided?

- 2. How do healthcare providers' SRL strategies manifest and vary by individual learner characteristics? What influence do these strategies have on healthcare providers' learning gains from using LIFE?**

I hypothesise that there will be a significant enhancement of learning gains from engaged students who are reflective of their learning needs, set learning goals, evaluate the effectiveness of the learning strategies they use, and evaluate their own emerging understanding of the topic.

- 3. What are the perceptions of healthcare providers on how to enhance the design and implementation of more engaging learning experiences on smartphone-based gamified emergency care training? How do these healthcare providers perceive such contextualised learning experiences to influence their SRL efforts as they embark on bridging their knowledge gaps?** Based on the literature review, I hypothesise that adaptive self-regulated gamified learning would help students reflect about their learning using platforms such as LIFE, allowing for accelerated transfer to, and future learning of, topics related to the context of care functions in their setting.
- 4. Given a healthcare provider's sequence in the emergency care tasks that they have prescribed, which knowledge tracing approach is better at predicting the outcome of the next caregiving task? How does this prediction vary by the healthcare provider's spacing of their own learning, experience and clinical role?**

This thesis's contribution to the literature is three-fold: It makes significant methodological and theoretical contributions to three unexplored areas for research:

- 1) It explores the pattern, nature and experience of self-regulated learning using smartphone-based gamified platforms in clinical training;
- 2) It explores the concept of knowledge tracing to address the current lack of systematic understanding of what approaches might be most appropriate for implementing domain-specific adaptive learning on smartphone-based clinical training in LICs.
- 3) It evaluates whether adaptive learning design approaches are relatively effective in enhancing healthcare providers' learning gains, given their context of learning.

This thesis seeks to identify the theoretical connection between existing popular knowledge measurement theories used for providing adaptive learning on digital environments and how their practical implementation for clinical training [might better] integrate the crucial ingredient of pedagogy. The findings are of value to global health researchers, learning experience designers and implementors, and policymakers in LICs who are trying to understand, promote, and improve the capacity of healthcare providers to provide quality clinical care.

4 METHODOLOGY

4.1 Overall methodological design

In this chapter the overall design of the research study is presented. To answer the research questions outlined in Chapter [3.8](#), a mixed methods study-design spread out over three phases using methodological triangulation (Schneider D.K., 2012) was used. This research study used a concurrent mixed method design with time-lag between phases (Johnson & Onwuegbuzie, 2004) where data in the initial and final phases was collected through qualitative methods, sandwiching a phase of quantitative data collection. This approach allows for confirmation, cross-validation, and corroboration of findings between phases and between methods (Johnson & Onwuegbuzie, 2004; Schneider D.K., 2012).

One key proposition of this thesis is that the ways in which individual healthcare providers from LICs self-regulate their own learning on gamified smartphone devices may be enhanced by an adaptive feedback mechanism. The theoretical framework as outlined by Winnie & Hadwin's SRL model places feedback as a mediator between individual healthcare provider learning behaviours and learning outcomes over time. For the purposes of this thesis, SRL behaviours of the LIC healthcare providers needed therefore to be evaluated both qualitatively and quantitatively. Additionally, how accurately the feedback mechanism used in LIFE was able to trace the knowledge of healthcare providers, and its comparison to other knowledge tracing approaches was also necessary.

Phase one of this thesis was geared towards exploring the perceptions of healthcare providers on how to enhance the design and implementation of more engaging learning experiences on smartphone-based gamified emergency care training. During this phase, I employed the Nominal Group Technique (NGT) -a structured small-group consensus method explained in detail elsewhere (McMillan, King, & Tully, 2016) – to collect data. NGT was followed by semi-structured interviews which were carried out within a week to compliment it. The semi-structured interviews were done with a different set of study participants to validate the findings from the NGT phase. Phase one also included quantitative data

collection from LIFE student-step data, the purpose for which was to help support the implementation of the Bayesian Knowledge Tracing (BKT) adaptive mechanism (Corbett & Anderson, 1994) in the LIFE project to be used in phase two. This is explained in detail in chapters 3.4.1 and 5.2.

Phase two methods consisted of a randomised experiment, SRL latent profile analysis from self-administered questionnaire data, and opportunity-based knowledge tracing analysis through use of Learning Factor Analysis (LFA) approaches which have been highlighted before in [Chapters 3.4.2](#) (Cen et al., 2006). Phase two methods used a different set of study participants from phase one and were aimed at addressing research questions one (i.e. *Within a gamified smartphone-based learning platform for clinical training, what is the effect of adaptive feedback on learning gains compared to standardised feedback and how is this effect moderated by (a) the learning opportunity for knowledge mastery, and (b) the level of adaptive support provided?*) and research question two (i.e. *How do healthcare providers' SRL strategies manifest and vary by individual learner characteristics? What influence do these strategies have on healthcare providers' learning gains from using LIFE*). Phase III consisted of in-depth semi-structured open-ended interviews and evaluative analysis of different variants of knowledge tracing models on LIFE data. Phase III was conducted with the same participants from the experiment in phase II. Qualitative data from Phase III was used to *explore how do healthcare providers from LICs perceive such contextualised learning experiences to influence their SRL efforts as they embark on bridging their knowledge gaps* and contrasted with qualitative data from phase one to evaluate whether there was corroboration of findings between phases and participants. Additionally, phase three involved novel knowledge tracing approach whose accuracy performance was compared to phase two methodology in order to evaluate the most appropriate approach to knowledge tracing for healthcare using LIFE platform. These phases are illustrated in Figure 4.1.1.

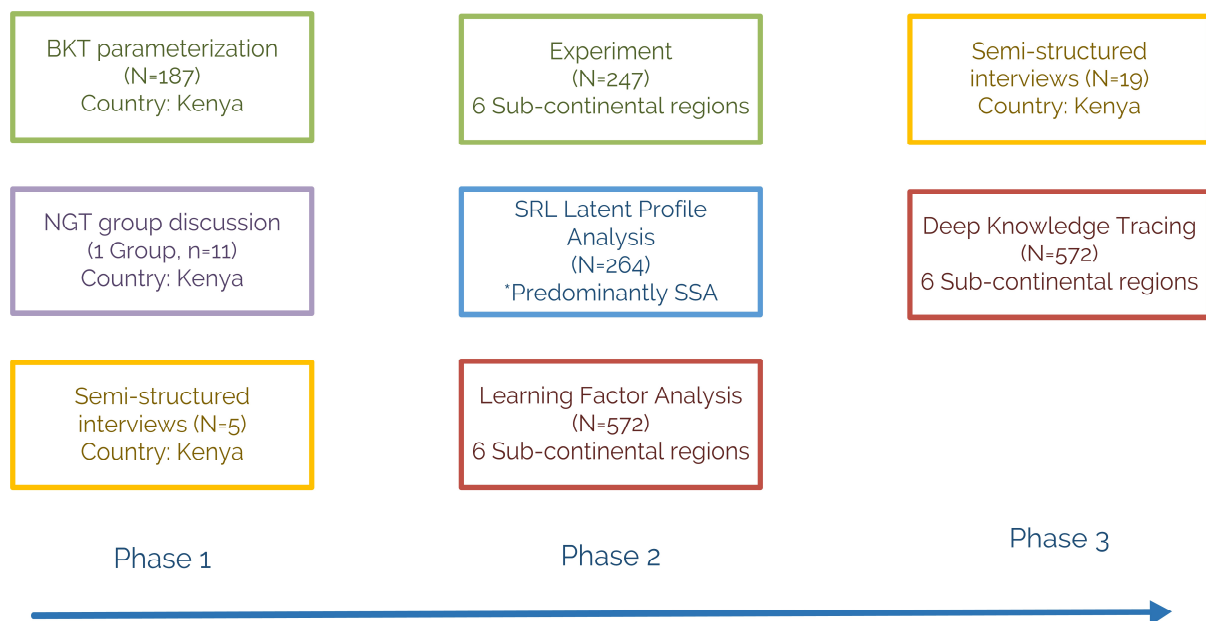


Figure 4.1.1: Overall study phases

The details and the rationale for the analytical approaches selected is discussed in detail together with the findings in Chapters 5-9, although some have already been addressed at length in the review of theoretical approaches this research takes (Chapters 3.3 – 3.6) and in the justification for research ([Chapter 3.7](#)).

In conducting this DPhil's research within the LIFE project, I introduced several research components (as illustrated in Figure 4.1.1) and technical components (e.g. integration of trace data collection using Google Firebase, BKT model fitting using C++) into the LIFE project to allow for data collection for the proposed research. During fieldwork, I designed the Nominal Group Technique study and materials but convened and facilitated it in collaboration with the LIFE team. I conducted data coding and initial thematic analysis with subsequent rounds of analysis peer-reviewed by the LIFE team. I also independently recruited participants for the qualitative semi-structured interviews, which I had independently prepared beforehand. I conducted the data coding and initial analysis of the qualitative data independently before the analysis being peer-reviewed by the LIFE team in subsequent rounds of interpretation of findings. This thesis contribution to the evolution of the LIFE project is illustrated by Figure 4.1.2.

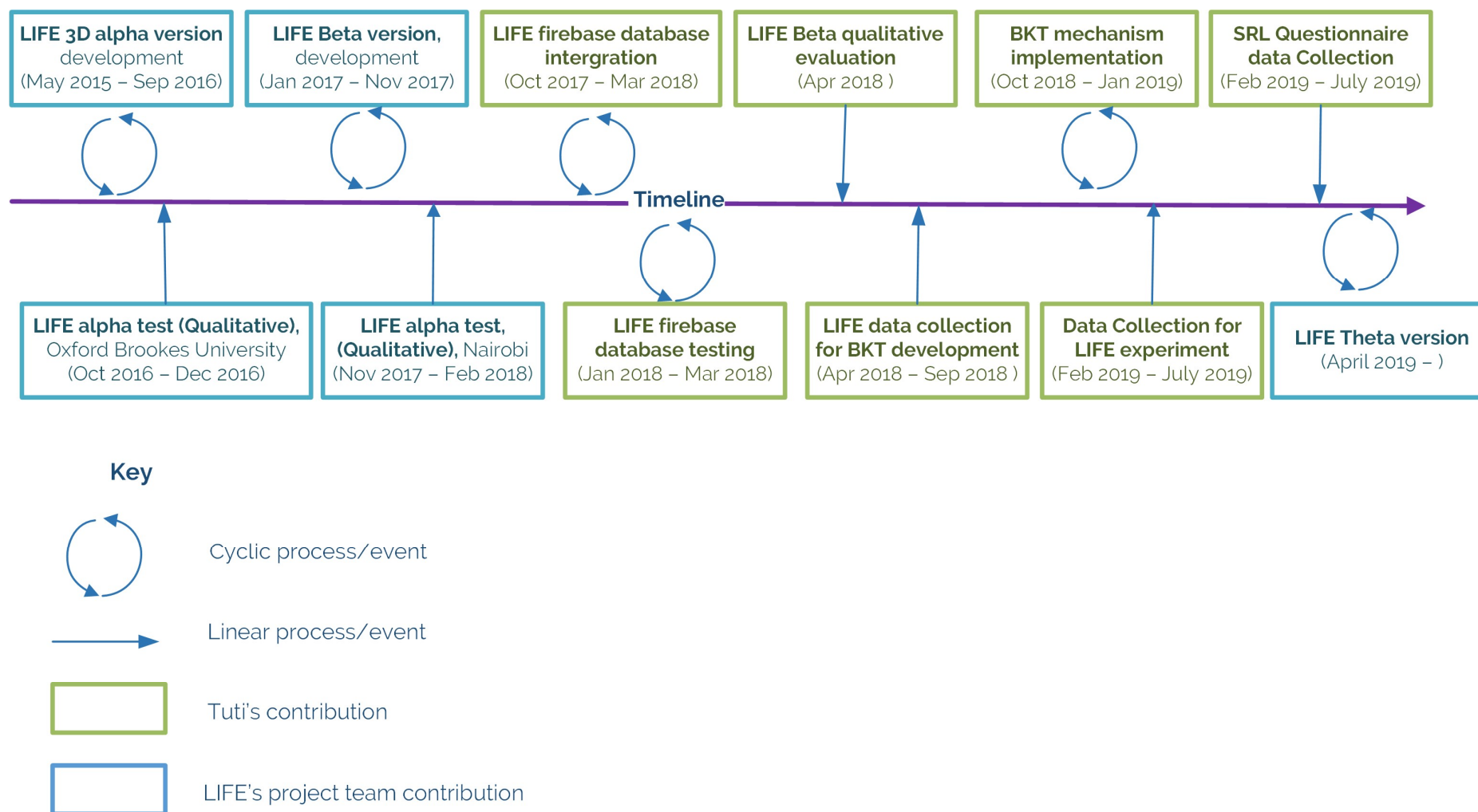


Figure 4.1.2: LIFE components distinguishing those implemented independently and specifically to allow for data collection for this DPhil project from the original platform.

4.2 Study participants and settings

The participants in all the study phases were healthcare providers who are in nursing, clinical or medical professional cadres that offer bedside patient care. Additionally, healthcare providers in practice and those in training were also eligible for inclusion in all the phases. Healthcare providers who had retired from clinical practice, or practiced in high-income settings, or participants who were not healthcare providers were excluded from the study. The determination of which country was a low-income country was made based on its classification by United Nations Office for the Co-ordination of Humanitarian Affairs (UN-OCHA).

4.3 Participant timeline

Enrolment of study participants for phase one (the pilot phase) was from April to October 2018. Enrolment of study participants for phase two and three began on February 1st, 2019 and continued up until July 31st, 2019. The rollout of the LIFE intervention was based on implementation study principles and outcomes (Peters, Adam, Alonge, Agyepong, & Tran, 2013) and informed by self-regulated self-directed learning (Winne H.P., 2011; Winne & Hadwin, 1998; Winne et al., 2006). It sought to understand and work within real world conditions, rather than trying to control for adoption, acceptability, coverage, and sustainability conditions or remove their influence on the study outcome (Peters et al., 2013). Subsequently, no training sessions were planned for study participants. Even though LIFE is designed for low-income contexts, there was no limit set by geographical coverage for self-directed healthcare providers who might be interested in undertaking this training; anyone could download the app. However, only healthcare providers from low-income countries were analysed.

4.4 Ethical considerations

All phases of the research highlighted here were approved by the Kenya Medical Research Institute's (KEMRI) Scientific and Ethical Review Committee (#3444) and the Central University Research and Ethics Committee of Oxford University (#ED-CIA-18-106). Informed consent was sought from participants either within the LIFE smartphone-application or in-person and will be addressed in detail

in the respective subsections of the methodology chapter of each of the phases where relevant. The informed consent process varied depending on the methodology and data collection procedures.

Overall, for the NGT data-collection, participation in this research phase was voluntary, with a representative of the participants consenting on behalf of all participants. All the in-depth semi-structured interviews required individual consent which happened before any data collection was conducted in efforts to ensure procedural ethics (Reid, Brown, Smith, Cope, & Jamieson, 2018). The conducted interviews were digitally recorded and transcribed by a transcription service and I checked them for accuracy. During the process of transcription all participants were anonymized by assigning interviewee codes.

Being conscious that the qualitative data collection relied heavily on the goodwill of healthcare providers with whom I had no relationships before, during and after the study, I believe that they consented to participate in this study based on exercising judgement about the validity of the qualitative study rather than on a pre-existing relationship with me, the interviewer. Additionally, the participants in all the qualitative data collection phases came from different cohorts at different timepoints in different locations: The possibility that they might have interacted with each other and thus influenced each other's response to the interview questions is minimal.

For the SRL survey tool, being that it was neither rolled-out in classroom set-up (which lends itself to frequent requests to include students as study participants), nor accessible to a huge number of learners such as in the case of Massive Online Open Courses (MOOCs), my expectation for high number of responses was minimal. Survey fatigue (Porter, Whitcomb, & Weitzer, 2004) -which from the pilot studies around LIFE, was raised as a barrier to participation in LIFE study- was also a concern in the SRL Questionnaire rollout. In efforts to ensure that the survey did not impinge on the use of LIFE by healthcare workers, it was embedded in LIFE as an optional component, with the informed consent process and data collection being separate and in-app as required by ethical approvals. This was done

to ensure that the participants had the power to opt in or out of various data collection procedures at different time-points.

The qualitative interviews and the SRL survey were designed to take as little time as possible away from the healthcare providers. To this end, in recognition of time constraints the participants faced, for the qualitative interviews, they were given the choice of deciding the venue of the interviews, and a suitable date and time. Also, the choice of the SRL tool was partly based on the length of the tool because longer survey tools tend to be off-putting to respondents and suffer from poorer response rates (Deutskens, De Ruyter, Wetzels, & Oosterveld, 2004).

4.5 Student-step data collected from LIFE application

Data collection was through the Android-based LIFE smartphone application, which would securely transmit a copy of anonymised student-step data to a Google Firebase distributed database. The unique identifier was linked to the mobile device that the participant used and was an encrypted calculated number that otherwise has no relevance outside the LIFE project. The survey data and the LIFE play data were kept in locations and used separate identifiers: this was done to maximise data security and ensure that the anonymity of participants was maintained. Student-step data collected included timestamp, whether the attempt was correct, the current attempt number, the time spent on learning task, whether feedback had been provided for each unique try per learning task, and the level of feedback provided. The learning trajectory data forms a complex hierarchically nested dataset with binary coded attempts to solve a task (0 = incorrect, 1 = correct) nested within tasks, nested within time-points, nested within learners.

This is illustrated in Figure 4.5.1 below with the referenced BKT mechanism, covered at length in Chapter 3.6, and will be explained later in Chapter 5.2.5. A more detailed breakdown of the tools and data collection will be provided in the methods section of each of chapters 5 to 9.

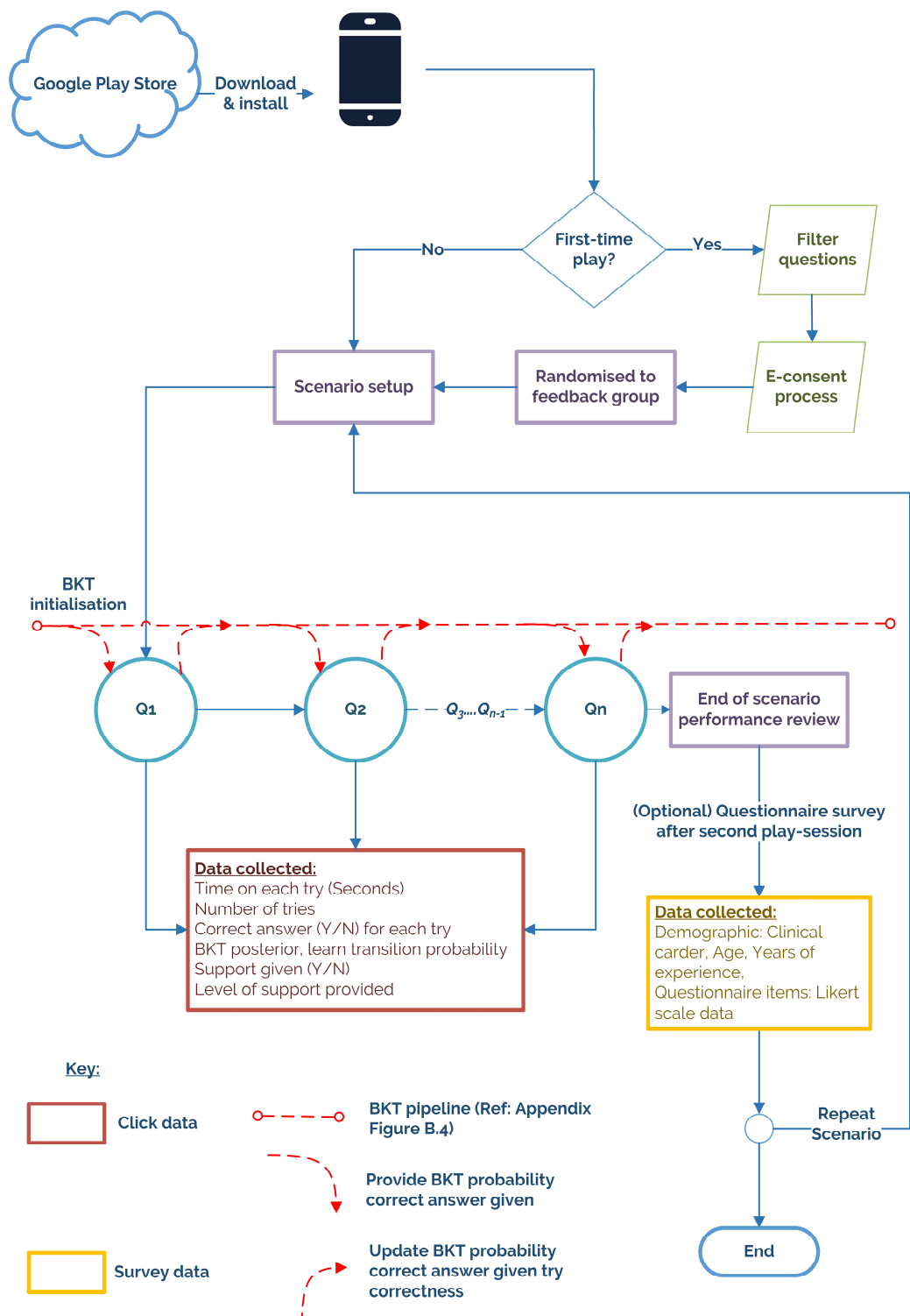


Figure 4.5.1: Data collection procedure within LIFE.

4.6 Summary of methodology

The qualitative and quantitative studies presented in this chapter, both the people-facing (those that assess SRL profiles, learning processes and learning gains) and technology-facing (those that evaluate adaptive knowledge tracing approaches on gamified platforms) serve to provide an implementation

science basis for adoption of smartphone-based adaptive learning approaches to gamified clinical training for LICs. In this chapter, I have outlined an overview of the overall research design and methodologies in a manner that is consistent with the theoretical underpinnings of this thesis ([Chapter 3](#)). I have also outlined the design of a concurrent mixed method design with time-lag between phases where data in the initial and final phases was collected through qualitative methods, sandwiching a phase of quantitative data collection. The qualitative phase of the was designed to explore the perceptions of healthcare providers on how to enhance the design and implementation of more engaging learning experiences on smartphone-based gamified learning. It also set to out to explore how healthcare providers perceived such contextualised learning experiences to influence their self-regulatory learning (SRL) efforts as they embark on improving the clinical care they provide.

Chapters 5-9 are the main results chapter of this document where the research findings are presented: They discuss the findings from the experimental study in Phase II ([Chapter 5](#)), which was aimed at addressing the first part of the first research question (*“Within a gamified smartphone-based learning platform for clinical training, what is the effect of adaptive feedback on learning gains compared to standardised feedback?”*).

Further findings from Phase II are presented in later sections: [Chapter 6](#) contains results from Latent Profile Analysis of self-regulated learning of healthcare providers from LICs using digital learning platforms. It is aimed at addressing the second research question (*“How do healthcare providers’ SRL strategies manifest and vary by individual learner characteristics? What influence do these strategies have on healthcare providers’ learning gains from using LIFE?”*). [Chapter 7](#) contains results from Learning Factor Analysis and is aimed at addressing the second part of the first research question (*“How is the effect of the adaptive feedback mechanism moderated by (a) the opportunity for knowledge mastery, and (b) the level of adaptive feedback support provided?”*). It includes the SRL profile indicator for healthcare providers which is derived from findings in [Chapter 6](#), as a key predictor of their learning curves from Learning Factor Analysis.

[Chapter 8](#) presents the findings related to the fourth research question: *Given a healthcare provider's sequence in the emergency care tasks that they have prescribed; which knowledge tracing approach is better at predicting the outcome of the next caregiving task? How does this prediction vary by the healthcare provider's spacing of their own learning, experience and clinical role?* To answer this question, Deep Knowledge Tracing model was employed to evaluate the hypothesised knowledge tracing approach in comparison to knowledge tracing approaches used in [Chapter 5](#) and [Chapter 7](#).

A key argument of this thesis is that the ways in which individual healthcare providers from LICs self-regulate their own learning on gamified smartphone devices *may be enhanced* by adaptive feedback mechanism. In efforts to support this argument, Chapters 5 aimed to tease out the effect of an adaptive feedback mechanisms on self-regulated learning gains. Chapters 6 and 7 aim at unearthing patterns of healthcare providers' SRL behaviours and their influence on learning gains. Using LIFE learning data, Chapters 7 in combination with chapter 8, also explored the nature and performance of the common knowledge tracing approaches which underpin adaptive mechanisms on digital learning environments. The details and the rationale for the analytical approaches selected are discussed in detail together with the findings in Chapters 5-9.

5 EFFECT OF ADAPTIVE FEEDBACK ON HEALTHCARE PROVIDERS SELF-REGULATED LEARNING GAINS: A RANDOMISED EXPERIMENT

5.1 Recap of Objectives

The primary objective of the phase II randomised experiment was to investigate whether adaptive feedback is more effective in advancing learning gains compared to standardised feedback for smartphone-based neonatal emergency care training. It aimed to address the first research question: *“Within a gamified smartphone-based learning platform for clinical training, what is the effect of adaptive feedback on learning gains compared to standardised feedback?”*. I hypothesised that healthcare providers randomised to receive adaptive feedback would have significantly higher learning gains than those randomised to receive standardised feedback. The secondary objective was to investigate how these learning gains, if any, are modified by healthcare providers’ characteristics and how they space their learning.

From the gaps identified in the literature review as outlined in Chapter 2.1.1 (i.e. mitigating against effects of lurking and confounding variables within gamified learning), this chapter aims to test a hypothetical causal chain where self-directed learning behaviours in gamified learning within LIFE affects the adaptive feedback mechanism, which in turn, affects healthcare provider’s learning gains. Chapter 5 (and 6) also emphasises the centrality of feedback in sustaining and enhancing self-regulated learning behaviours on smartphones devices. Finally, Chapter 5 seeks to generate evidence on the effect of such adaptive learning approaches for clinical training using gamified platforms in LICs, as deemed warranted by the findings of the recent systematic reviews on this subject (Car et al., 2019; Fontaine et al., 2019; Ma, 2017).

5.2 Methodology

The methodology outlined in this section (Chapter 5.2) utilises an experimental approach, and its structured adheres to the CONSORT guidelines for reporting randomised experiments (Moher, Schulz, Altman, & Group, 2001).

5.2.1 Study design

The study was a parallel-group double-blinded, randomised experimental design with an allocation ratio of 1:1. The participants were randomised to receive the intervention or to be in the control group when they launched the LIFE app for the first time on their personal smartphone devices.

5.2.2 Eligibility criteria

The participants were healthcare providers who were undergoing training (e.g. students) or actively providing nursing, clinical or medical professional cadres that offer bedside patient care. Healthcare providers who had retired from clinical practice, who practiced in high-income settings, or participants who were not healthcare providers were excluded from the analysis.

5.2.3 Study setting and recruitment

This study was confined to participants from low-income countries (as defined by UN-OCHA) which stand to benefit from training in the management of paediatric emergencies. Distribution of the intervention was through the Google Play Store with recruitment efforts using three main avenues:

- A. Remotely, through raising awareness of the Android Google application on social media platforms and online health networks forums every couple of weeks during the study duration to promote voluntary, self-enrolment in use of the training tool; The first avenue involved using platforms that are focused on health capacity building in LICs, in line with LIFE's focus.
- B. Employing snowballing sampling strategy by using ETAT+ trainers, previous study participants, and lecturers of medical sciences at local Kenyan universities to suggest (and encourage) other health workers to self-recruit into the study; and
- C. Recruiting participants using purposive selection strategy where every month, a unique eligible participant set was convened to ensure the recruitment targets were diverse and met.

Snowballing sampling is a type of convenience sampling where a group of people (in this case, already recruited participants) recommend potential participants for the study (Etikan, Alkassim, & Abubakar, 2016). The cycle of recommendation of participants for inclusion into the study continued during the

study duration, as one of the mechanisms by which participants are recruited. This also served as an indicator of LIFE intervention scaling. Avenues two and three relied on the following sites:

1. Kenyatta National Hospital - Neonatal Nurse Training Unit
2. College of Health Sciences, University of Nairobi
3. Kenya Medical Training College (Nairobi)
4. Gertrude's Garden Children's Hospital Nurse Training School (Nairobi)

5.2.4 The intervention

The intervention in this experimental study was the adaptive immediate feedback that is provided while learning through LIFE. The content to be learnt was based on a neonatal resuscitation guideline course that is already offered in nine low-income countries (Ayieko et al., 2011; Irimu et al., 2008; Tuyisenge et al., 2014). The smartphone app was publicly available on the Google Play Store where it was downloadable and installable to compatible android-based smartphones. All study participants received a link to the LIFE application hosted on Google Play Store or directions on how to locate it on Google Play Store. The LIFE application was designed to target Android's SDK 19 as at February 2019, which was the minimum version of Android supported then and targeted 100% of Android devices (Unreal Engine, 2018). The smartphone application had already undergone alpha testing on a pilot cohort of healthcare providers' smartphones from Kenya since October 2017 to ensure stability and reliability of its functioning on different mobile phones.

The number of standardised feedback levels was determined by global health academics and expert medical trainers involved in ETAT+ training in SSA in consultations with the relevant medical professional bodies accrediting continuous professional development. The adaptive immediate feedback provided to experimental group participants was designed to arouse meaningful reflective learning from continuous interaction between the learners and the smartphone-based training (Dewey, 1998; Kiili et al., 2012). As didactic-procedural form of feedback (Holmes et al., 2015) aligned to my theoretical framing, it was designed to force the healthcare provider to contemplate over the incorrect care provision choices they provided in their failed attempt (e.g. "**Some** of the selected

actions are not **appropriate** at this **stage**” focuses on the number of wrong choices, and their placement within the clinical care-giving pathway and is meant to force reflection as to which stage they are most appropriate).

The feedback provided to the experimental arm after each incorrect attempt at a learning task had three cascading levels of detail based on predicted probability that the learner’s next attempt was going to be correct. The mechanism for the delivery of the cascaded feedback was only available to participants in the experimental arm. Using the *Holmes et al* (Holmes et al., 2018) multidimensional model of personalisation, in this experiment, the adaptive mechanism targets the personalisation of how feedback is to be presented and when it is to be introduced. The wording of the feedback provided for Level 1 in Table 4.6.1 (which was only available to participants in the experimental arm) was also dependent on the number of incorrect choices the learner had selected and the actual incorrect choices themselves. This is illustrated in Table 5.2.1 and by Figure 5.2.1.

Table 5.2.1: Feedback content provided in LIFE smartphone application for Neonatal Resuscitation Scenario A Training for the individual learning tasks			
Feedback	Level 0 (Minimal Feedback)	Level 1 (Reflective feedback)	Level 2 (Detailed feedback: Explanatory)
Quiz			
Quiz 1	No, that is not correct. Try Again.	[Both/One/None] of these pieces of information are incorrect in directly affecting the recommended sequence of actions to resuscitate a new-born baby.	Incorrect. The two key questions are: • Is there any meconium in the liquor? • Is there anyone else who can assist me? Both these pieces of information directly affect the recommended sequence of actions to resuscitate a new-born baby. The other questions certainly affect the health of mother and / or baby and will be useful to know at some stage, but they are not vital for the first few minutes of neonatal resuscitation.
Q2	No, that is not correct. Try Again.	[X] of these pieces were incorrect. Of the selected, [Y] are not necessary	This selection was incorrect. In addition to the resuscitaire (a flat, warm surface) you need the following: • warm towels • a bag-valve device • small facemasks (size 0 or 1)

			• a suction machine • a wide-bore sucker • a clock • stethoscope These are essential for basic resuscitation of the newborn.
Quiz 3	These actions are not appropriate at this stage. Try again.	[Both/One] of the selected choices are not appropriate	Incorrect. The correct answers are: • Receive the baby in a warm dry towel • Place baby under a warmer on a flat surface.
Quiz 4	These actions are not appropriate at this stage. Try again.	[Both/One] of these actions are not appropriate at this stage. Try again.	Incorrect. The correct questions are: • Dry the baby all over and stimulate by rubbing its back for 20 seconds • Wrap the baby in a warm dry towel
Quiz 5	These actions are not appropriate at this stage. Try again.	[All/Some/One] of the selected actions are not appropriate at this stage	Incorrect. The correct questions are: • Is the baby crying? • Is the baby moving? • Is the baby pink rather than pale?
Quiz 6	Try again		Correct Answer: • Shout for help
Quiz 7	No, that is not correct. Try again. What will you do next?	Here's a hint: neither [user answer] nor Y is correct. Try again.	Incorrect. The most important thing to do once you have decided the baby needs resuscitation is to put the baby's head in the neutral position.
Quiz 8	That is not the best head position for the airway. Try again		The correct head position is shown here. Try again.
Quiz 9	No, that is not correct. Try again.	Here's a hint: neither [user answer] nor Y is correct. Try again.	Incorrect. It is important now the airway is open to recheck for breathing. Try again.
Quiz 10	No, that is not correct. Try again.	Here's a hint: neither [user answer] nor Y is correct. Try again.	If the baby is not breathing, the priority is to ventilate using a bag valve mask. Try again.
Quiz 11	That is incorrect. This mask is not the right size for this baby. Try again		This is incorrect. The correct facemask is shown here. The facemask should fit over the nose and mouth of the neonate and make a good seal with the face and chin, without covering all or part of the eyes. The shape is less important than the correct size and either circular or triangular masks are fine. Generally, a size 0 or size 1 mask will be the best fit for a newborn. It is important not only to have the correct size but also to position it well, so that air can be pushed into the lungs through the open airway and not leak out

			around the sides between the mask and the face.
Quiz 12	That is not correct. Try again.	[All/Some/One] of these options are not appropriate. Try again.	That is not correct. The priority is to inflate the lungs with air. To do this you need to maintain an open airway using the neutral head position, give 30 breaths per minute and see the chest rise and fall when you squeeze the bag. Other actions will delay this life-saving manoeuvre.
Quiz 13	No, that is not correct. Try again.	Here's a hint: neither [user answer] nor Y is correct. Try again.	That is not correct. You should now check for a pulse to identify whether the baby has an adequate cardiac output. This is not done earlier because the most common problem a newborn has is failure to breathe.
Note: No Reflective feedback for was provided for Quiz 6, 8, and 11 because of ongoing changes to these learning tasks during the study period.			

The control group study participants received standardised non-adaptive immediate feedback after each incorrect attempt at a learning task, with the feedback on the first incorrect attempt asking the learner to retry (Level 0, Table 5.2.1) and the feedback on the second attempt giving a detailed explanation of the correct choices to select (Level 2, Table 5.2.1). Unlike in the experiment group, participants in the control group did not receive any reflective feedback (Level 1, Table 5.2.1) nor was the decision on the feedback provision probabilistic. The mechanism by which feedback was provided is detailed in the next section and illustrated by Figure 5.2.1 whose referenced equations draws from Chapter 3.6.1 and Appendix C.1.

The LIFE application was both the learning and the measurement tool: The learning task is synonymous with the quiz as learning is designed to be embedded in the formative evaluation. At the end of a successful completion of a learning session, the platform provided a performance score based on whether each learner's response to the learning tasks was correct on first attempt. For this study, a learning session was conceptualised as every unique initiation (i.e. iteration) of the neonatal

resuscitation learning scenario training round on the LIFE smartphone application (illustrated in Scenario A in Figure 1.3.2).

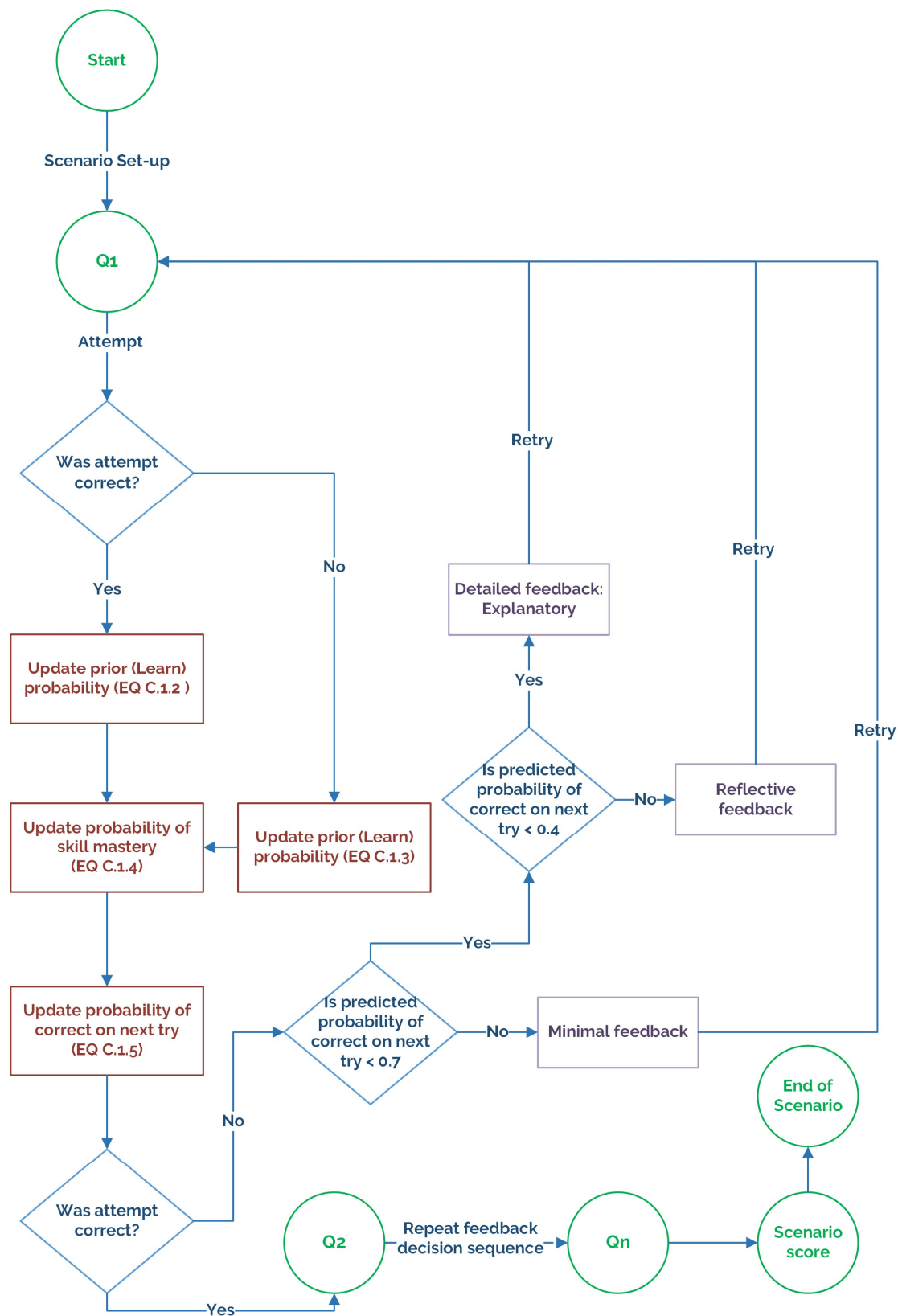


Figure 5.2.1: The flow of feedback provision.

5.2.5 Implementation of BKT mechanism for adaptive feedback

BKT was designed as the MAP mechanism to implement the feedback outlined in Table 5.2.1. The participants, learning data and procedure for fitting the standard BKT model used is outlined in this section. This pre-study phase was necessary for setting up the BKT mechanism for the experiment and stems from a pre-study data collection phase illustrated by Figure 4.1.1.

5.2.5.1 Pilot study design, population and settings

The pilot study for setting up the BKT mechanism for the experiment was a retrospective observational study of 187 healthcare providers from both public and private hospitals in LICs, who represented clinical cadres such as nurses, clinical officers and medical doctors, with varying experience from students to consultants. Participants were enrolled into the study through a combination of snowballing and convenience sampling strategy from April 2018 - October 2018. Recruitment occurred through use of peer referrals among clinicians, publicised through regional clinical meetings and clinical conferences and medical training institutions. The eligibility criteria for inclusion were that the participants had to be healthcare providers from LICs either in training for, and/or actively providing bedside clinical care.

5.2.5.2 Study tool and variables

The learning scenario used in this LIFE version for the pilot study provided simulation training on the contextualised management of new-born resuscitation through a series of sixteen learning interactions that elicit responses from learners in the form of multiple-choice answers or performing interactive tasks. At the end of a successful completion of simulation tasks, the application provided performance score based on the learner's first attempt at each learning interaction. Data collection was through the Android-based LIFE smartphone application, which would securely transmit a copy of anonymised student-step data to a Google Firebase distributed database. The outcome of interest was specified as getting each answer correct on the first try in the current learning session. Standardised feedback was provided immediately after all incorrect attempts. If a learner had gotten an attempt incorrect the first time, the feedback message would just indicate that their choice was incorrect, and that they needed to retry. For all subsequent incorrect attempts after the first one, the learner would get a detailed feedback message explaining what the correct options were and why

with an option for more information windows. At this stage of piloting LIFE, demographic data from study participants was not yet covered by the ethical approvals for the LIFE project. A summary of these variables is illustrated in Figure 5.2.1.

Table 5.2.2: Summary statistics of pilot data from LIFE game play

Indicator	Complete*		Incomplete**		P-Value [‡]
	Mean	SD	Mean	SD	
Time spent on each question (in seconds)	12.78	9.19	14.57	10.96	0.228
Number of feedback messages provided for failed attempts per question i.e. feedback	0.26	0.44	0.32	0.46	0.369
Cumulative tries on a question across sessions (referred to as Opportunities from here onwards)	2.18	2.72	2.62	3.4	0.328
Average performance (%) ***	55.66	28.08	49.02	31.42	0.132
Note: *Learners who completed at least one session: N=77, **Learners who did not complete at least one learning session: N=110. ***Average performance based on number of quizzes attempted. [‡] From evaluating if there is a difference in the mean of the values between 'Complete' and 'Incomplete' groups					

5.2.5.3 Statistical analysis

Data manipulation and statistical analyses for the data from the pilot phase were performed using R and C++ software (R Core Team, 2013). For the purposes of analysis, healthcare providers in the pilot phase with missing data (i.e. incomplete learning session on LIFE) were also included in the analysis. I fitted the standard BKT model using the procedure detailed in this GitHub Repository (M. Yudelson, 2016) using Gradient Descent optimisation algorithm for minimising the loss function. Although Individualised BKT is possible, its outputs do not provide skill-specific parameters for guessing, slipping and learning (Figure 3.6.1) rather they provide student-specific parameters. This would have made the individualised BKT parameters from the pre-phase data inappropriate for use in the experimental study.

5.2.5.4 BKT model output and parameters

The fitted standard BKT model had a predictive accuracy of 0.6431 (Log likelihood= 3955.25, AIC = 8006.50, BIC=8329.73, RMSE=0.4708). From the fitted model, skill-specific parameters generated - which are explained in detail in [Chapter 3.4.1](#)- are listed in Table 5.2.3 and used to populate BKT parameters in equations in C.6 in [Appendix C.2](#).

Quiz	Prior (a)	Transition (b)	Slip (c)	Guess (d)
1	0.5172	0.2526	0.2051	0.1254
2	0.6113	0.2092	0.2778	0.0982
3	0.1675	0.4096	0.2251	0.2949
4	0.5199	0.3569	0.2296	0.1931
5	0.5647	0.266	0.2426	0.1317
7	0.4591	0.419	0.1426	0.2393
9	0.5152	0.4655	0.1149	0.268
10	0.5283	0.3712	0.2374	0.1954
12	0.4586	0.3819	0.0982	0.2041
13	0.3441	0.3594	0.1003	0.2731
Note: Quiz 6, 8, and 11 were either introduced after the pilot or constantly undergoing changes (i.e. not stable), and therefore omitted from the study.				

Using Table 5.2.3 parameters for calculations as outlined in [Appendix C.2](#), learner trajectories would end up being differentiated by the outcome of each subsequent learning opportunity (i.e. Previous cumulative tries on a question across all sessions) on each skill. This is illustrated in Figure 5.2.2, for a single learning "...concept, principle, fact, skill, schema, production rule, misconception..." (Kurt VanLehn et al., 2007). Y-Axis represents the predicted probability that the individual healthcare provider has learnt the KC given the performance at the last opportunity.

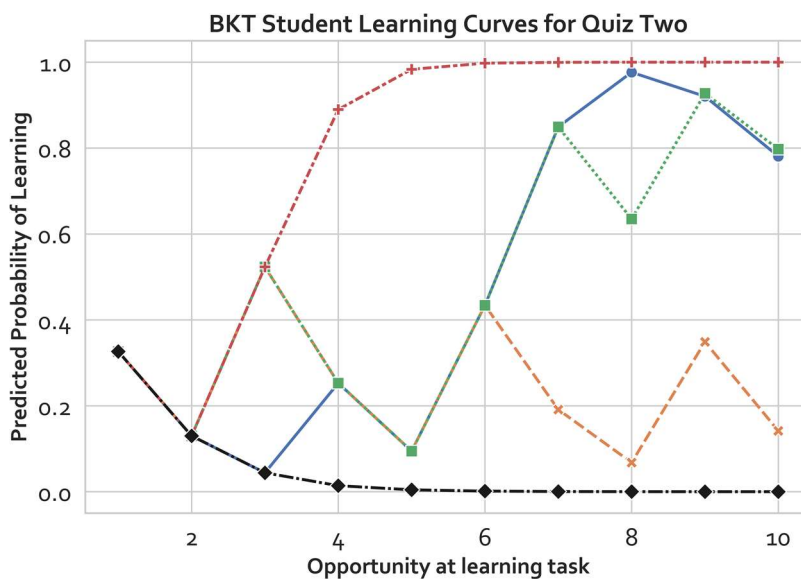


Figure 5.2.2: Example of BKT learning trajectories from a single Knowledge Component (KC) for five different healthcare providers.

In LIFE's case, the ZPD boundaries were defined by expertly-set thresholds using the prediction probability of skill mastery from the inferred latent knowledge state. These thresholds, while seemingly arbitrary, provide a decision criterion by which to determine feedback level to provide when supporting the differentiated learning trajectories. Probability of getting the next attempt correct (Appendix C, equation C.5) was used as the decision criteria for deciding level of feedback to provide and followed the mechanism illustrated by Figure 5.2.1. Probability thresholds of **<0.4**, **0.4 to 0.7** and **>0.7** were used for detailed, reflective and minimal feedback respectively for the experiment arm (Table 5.2.1). From previous studies (Chounta et al., 2017), uncertainty about student performance is high when a student model predicts the learner's ability to make a successful attempt with a probability around 0.5. This probability threshold region, where the prediction error is considered high, has been referred to by one study as the "Grey Area" and conceptualised to approximate the student's ZPD. The reasoning for selecting these thresholds as analogous to targeting ZPD mechanism is that:

"...The space 'Above the Grey Area' [> 0.7 in this case] denotes the region where the student is predicted to answer correctly (the fitted probability is considerably higher than the cut-off threshold) and consequently may indicate the area above the ZPD; that is, the area in which the student is able to answer a question without any assistance. Accordingly, the space 'Below the Grey Area' [< 0.4 in this case] denotes the area where the student is predicted to answer incorrectly (the fitted probability is considerably lower than the cut-off threshold) and consequently may indicate the area below the ZPD; that is, the area in which the student is not able to carry out the task either with or without assistance..." (Chounta et al., 2017)

Admittedly, the symmetry and size of "the Grey Area" could change depending on the classification threshold and the learning objectives. However, with no prior baseline to work with for LIFE learning data, thresholds similar to Chounta et al were imitated for simplicity. Mapping out this latent space is burgeoning area of interest that warrants further research in the future.

From Holmes et al (Holmes et al., 2018) multidimensional model of personalisation, in this phase II experiment, the adaptive mechanism targeted the personalisation of **how** feedback was to be presented and **when** it is to be introduced. This methodology for evaluating adaptive feedback mimics

a common approach in experimental design known as A/B testing which is used to optimise components such as feedback or user interface on digital platforms (Kaufmann, Cappé, & Garivier, 2014; Kohavi & Longbotham, 2017). In particular, the experiment described here emulates A/B tests' commonly used Hypothesis Experiment Data-Driven Development (HYPEX) model, whose description and use is detailed elsewhere (Fabijan et al., 2018). Such an experimentation approach can be used for digital health interventions where (1) there is need for data-driven decision making on inclusion of features, (2) the data is trustworthy, but (3) assessing the added value of features on the digital platform is difficult (Kohavi & Longbotham, 2017).

5.2.6 Outcomes

The primary endpoint for both arms of the experiment was the completion of two learning sessions using LIFE, the first session being treated as pre-test and the second session being treated as post-test. Both scores were converted into percentages. From the pre-post scores, the study's main comparative outcome was the learning effect size (g), (Scott B. Morris, 2008) with the formula for its calculation provided below:

$$d = C_{pp} * \left(\frac{(M_{post,T} - M_{pre,T}) - (M_{post,C} - M_{pre,C})}{SD_{pre+post}} \right) \quad (5.2.1)$$

where the pooled standard deviation is defined as

$$SD_{pre+post} = \sqrt{\frac{(n_T - 1)SD_{pre,T}^2 + (n_C - 1)SD_{pre,C}^2 + (n_T - 1)SD_{post,T}^2 + (n_C - 1)SD_{post,C}^2}{2(n_T + n_C - 2)}} \quad (5.2.2)$$

and the bias adjustment is provided by the formula:

$$C_{pp} = 1 - \frac{3}{4(2n_T + 2n_C - 4) - 1} \quad (5.2.3)$$

and

Sample size of treatment group: n_T

Pre-test mean for treatment group: $M_{pre,T}$

Post-test mean for treatment group: $M_{post,T}$

Standard deviation of means for pre-test in treatment group: $SD_{pre,T}$

Standard deviation of means for post-test in treatment group: $SD_{post,T}$

Sample size of control group: n_C

Pre-test mean for control group: $M_{pre,C}$

Post-test mean for control group: $M_{post,C}$

Standard deviation of means for pre-test in control group: $SD_{pre,C}$

Standard deviation of means for post-test in control group: $SD_{post,C}$

This effect size, also referred to as *Morris G* (Scott B. Morris, 2008), represents the standardised mean difference between the relative change in the scores in the study arms. Based on education literature (Ellis P.D., 2009; Glass, Smith, & McGaw, 1981; Ma et al., 2014), effect sizes of approximately ~0.2, ~0.5, and ~0.8 are considered as small, moderate and large effect sizes respectively. These thresholds represent the magnitude of the effect and reflect my assumption that a statistically significant result is not necessarily important or meaningful. For example, for an effect size of around 0.2, the difference between the study groups is trivial even if it is statistically significant (Ellis P.D., 2009; Ian Walker, 2008).

In addition to randomisation -which eliminates or at least dramatically reduces biases influencing this study's primary outcome- the method of calculation of this experiment's primary outcome: (1) is robust in managing pre-existing knowledge differences among learners, (2) allows for the estimation of the intervention effect even when experiment and control groups are non-equivalent and (3) considers the variances of both pre-test and post-test scores. This contrasts with other forms of effect-size calculation such as Hedges G, and Becker D, which only use pre-test pooled variances (Scott B. Morris, 2008). In this model, the pre-test and post-test variances were assumed to be homogeneous. The secondary outcome considered was the individualised learning gain of the study participants, defined as the relative change in performance score of healthcare providers divided by the maximum score they could have improved upon (Marx & Cummings, 2007). This was calculated from the performance scores from learning sessions following each other, with the performance from the first

session treated as pre-test score and the one from the last session treated as a post-test score. The normalised score ranges from -1 to +1. The formula for this calculation is given below:

$$Learning\ gain\ (c) = \begin{cases} \frac{post - pre}{100 - pre} & post > pre \\ drop & post = pre = 100 = 0 \\ 0 & post = pre \\ \frac{post - pre}{pre} & post < pre \end{cases} \quad (5.2.4)$$

Equation 5.2.4 helps to minimise the low test-score bias in calculation of individual health providers knowledge gain and makes interpretation of learning scores is straightforward by ensuring the learning gain neither falls within a non-symmetric range of scores nor includes infinity values (Marx & Cummings, 2007). When the normalised learning gains are positive (i.e. > 0), the learners are making performance gains from what they learnt as a ratio of the maximum amount they could have learned.

The key difference between the primary and the secondary outcome is the unit of analysis: The primary outcome of the experiment is an effect size whose unit of analysis is the group level, while the secondary outcome is an effect size whose unit of analysis is the individual. While both are standardised mean differences, the former being baselined at group and the latter being baselined at the individual level, the latter effect size also adjusts for individual healthcare providers characteristics and self-directed spacing of learning.

5.2.7 Participant timeline

Enrolment of study participants began on February 1st, 2019 and continued up to July 31st, 2019. This study's rollout of LIFE's intervention was based on implementation study principles and outcomes (Peters et al., 2013) and informed by self-regulated self-directed learning (Winne H.P., 2011; Winne & Hadwin, 1998; Winne et al., 2006). It sought to understand and work within real world conditions, rather than trying to control for adoption, acceptability, coverage and sustainability conditions or remove their influence on the study outcome (Peters et al., 2013). Subsequently, no training sessions

were planned for study participants. Even though LIFE is designed for low-income contexts, this study set no limit by geographical coverage for self-directed healthcare providers who might be interested in undertaking this training; anyone could download the app. However, I only analysed healthcare providers from low-income countries. Participants without any geographic location data (due to refusing to grant the LIFE application the required Android permissions) were assumed to be from developing countries given that the recruitment efforts were directed towards professional groups in these countries.

5.2.8 Sample size calculation

Similar interventions in other subject domains have been found to have a mean effect size of between 0.22 (95% CI: 0.16 – 0.27) to 0.70 (95% -0.08-1.49) (Fontaine et al., 2019; Ma et al., 2014). Drawing from these studies, I use the smallest effect-size of the two systematic review studies to ensure a well powered experiment. To detect the effect size of 0.22 with a two-sided 5% significance level and a power of 80%, a sample size of 83 participants per group who reach the primary endpoint of the study was necessary. A sample size calculation for a one-way analysis of variance, together with one-sample and paired-sample t-test analysis using the same power calculation parameters produced the same required sample size of 166 participants.

Based on the alpha tests of the LIFE smartphone app, I assumed a 50% drop-out rate of study participants - with drop-out defined as the incomplete or single use of LIFE smartphone application- and planned to recruit at least 332 participants to account for this drop-out rate. To encourage repeated usage of LIFE, all participants received email reminders from the time they were enrolled on the study once every two weeks and this was suspended after they had received three reminders. From interviews of the study participants in the alpha phase of the LIFE application development, those who were from lower clinical cadres tended to characterise the cost of phone data charge necessary for downloading the smartphone application as too high. To mitigate against burdening participants with this extra personal cost due to participating in the study, they could request for reimbursement of costs within a few weeks of using the LIFE smartphone-app.

Demographic data were collected within the application in its initial use by study participants at the end of the first learning session if they consented. This was because demographic data were deemed more sensitive than trace play data, and therefore required additional consent as per the ethical approval. This meant that for learners who dropped out before completing the first session, or chose not to fill-in that data, no demographic data was collected. From these data, statistical analyses were conducted to evaluate whether there was any systematic bias in attrition of study participants. Study variables used in this analysis were study groups, exposure to previous neonatal training, clinical cadre, age, and level of experience. This was because level of expertise given the differences in training pathways for clinical cadre and length of practice might produce differences in effectiveness of learning (Wijk, Ponzer, Heikkilä, Kihlström, & Nordquist, 2019). Additionally, age was included to evaluate whether it was associated with the pattern and effectiveness of smartphone-based learning given its novel nature requiring digital acuity (James D Pickering, 2017). I judged sex not to be theoretically influenced by the socio-cognitive framing of this experiment's research questions.

5.2.9 Randomisation

For allocation of the participants, an in-app algorithm randomly generated a value between zero to three when the smartphone application was launched for the first time. If the value was either zero or one, it was recoded to zero, otherwise recoded as one. This procedure is illustrated by Figure E.1 in Appendix E.2. For development reasons, this algorithm was implemented using a randomisation routine provided by the game engine which I assumed to be reliable (QuantStart, 2019). The algorithm determined whether a participant was allocated to the control or experiment group. It also blinded both the study participants and intervention staff to the allocation of participants to groups during the experiment, but not at the analysis stage. Sequence generation for random allocation was a computerised procedure pegged on a single instance (i.e. smartphone application installation instance) that mimicked a coin-tossing procedure. Therefore, using permuted blocks of random sizes to assign participants to either control or experimental group was not possible and thus not implemented.

5.2.10 Statistical methods and analysis

For the primary endpoint, I used Morris G effect size to analyse the differences between study arms of relative change in scores within the arm. This was assessed after the performance from the second round of completing training through the mobile application had been recorded. Secondary analysis was conducted using regression analysis, with the dependent variable being the normalised learning gain and the independent variables being healthcare providers' demographic characteristics and the game play characteristics (e.g. spacing of repeat learning session, amount of time spent on learning task, previous exposure to neonatal training etc.), to evaluate their effect on learning gains. All the analysis was conducted using the *Statsmodel* package for Python 3.4 (Seabold & Perktold, 2010). The progression of the analysis of the effect of adaptive feedback versus standard feedback size was from group level (using the primary outcome, *Morris g*), to individual level (secondary outcome, *Individualised normalised gain*) to the learning item level (which is addressed by a Learning Factor Analysis modelling approach highlighted later in [Chapter 4.8](#) and [Chapter 7](#)).

The primary learning outcome used in this study could not be computed for study participants whose drop-out was characterised by lack of at least two complete learning sessions. Without a post-baseline assessment, 'intention-to-treat' analysis could not be done unless I imputed outcomes – which tends to produce biased estimates (Abraha et al., 2015). I therefore did not conduct an intention-to-treat analysis. However, drop-out numbers are reported in relation to those who reached the study's primary endpoint, with their implications discussed considering self-regulated, self-directed learning (Winne H.P., 2011; Winne & Hadwin, 1998; Winne et al., 2006).

Qualitative interviews were conducted in parallel to the experiment from a small sample (N=19) of the healthcare providers who participated in the study, regardless of whether they reached the primary endpoint or not. These interviews were used to explore healthcare providers perceptions of self-regulated learning that affected the contextual use of the smartphone-based learning platform. They provided a context for interpreting the observed learning outcomes from this study and are addressed in [Chapter 9](#) in this thesis.

5.2.11 Data Management

The primary data collected from the study participants' Android smartphone application is held on their devices with a back-up copy synchronised to *Google Firebase* a secure distributed online database server, after being transmitted in an encrypted format. Qualitative data collected after the experiment were stored on encrypted password-protected USB devices, and transferred to secure password protected servers in Kenya and Oxford. De-identified data were shared with study staff based at the research institutions listed on the ethical approval forms.

5.2.12 Informed Consent

Individual participant consent was elicited from within the mobile application before collection of any demographic data in addition to using explicit Android permission requests for collecting trace anonymised learning data from the application. This approach of in-app informed consent to informed consent is not uncommon in medical research; it has been described in detail in a systematic review (Moore et al., 2017), as well as specifically for mobile application based research (Grady et al., 2017). The informed consent process also included information on the credentials and affiliations of the researchers, and developers of the LIFE platform. Participants had no way of knowing whether they were receiving the “intervention of interest”.

5.3 Results

5.3.1 Baseline characteristics of the participants

At the end of the experiment's data collection period in phase II, 897 participants had been enrolled, of whom 572 were eligible for inclusion into the study and 247 (43.18%) ended up reaching the study's primary endpoint (Figure 5.3.1) as expected during the study design phase (Tuti et al., 2019)¹. SRL questionnaire data is reported in [Chapter 6](#).

¹ The protocol and analysis results were published in Journal of Medical Internet Research (JMIR) with the code for analysis found in the following GitHub repository: <https://github.com/AfrikanaCollective/Statistical-Programming-Projects/tree/master/Experiment>

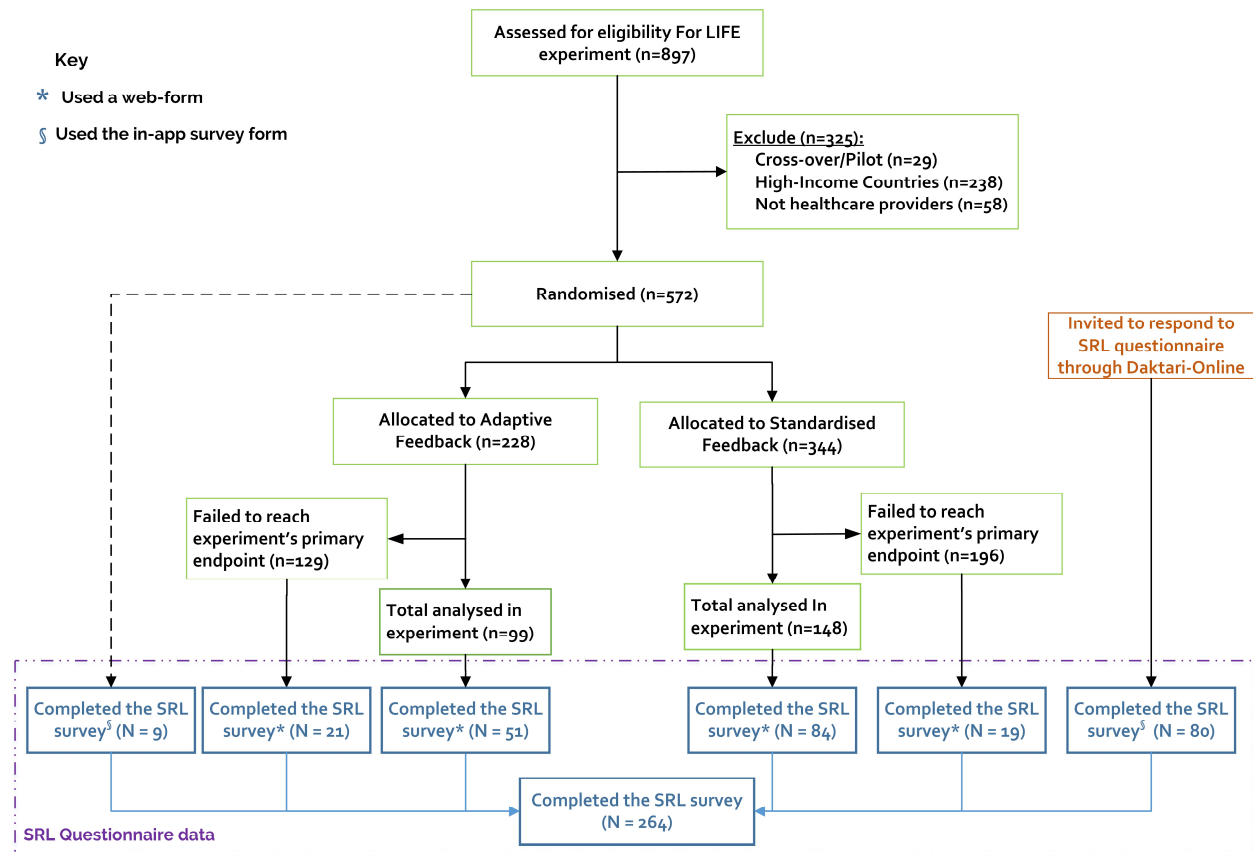


Figure 5.3.1 Experiment's enrolment flow-chart.

It was challenging to enforce the allocation ratio because the algorithm used for randomisation was not centralised and therefore unable to implement a blocking mechanism in participant to group allocation. Randomisation was happening on the individual healthcare provider's personal smartphone independent of how other participants were allocated and is illustrated in [Appendix D.2](#). In addition to lack of an enforceable blocking mechanism, the software library used for the random generator function (C++ rand function) is most likely to have caused an unequal allocation ratio since it has varied efficiency across multiple smartphone devices which remain largely unknown (Lavavej, 2013; QuantStart, 2019). It is highly likely that this was what contributed to the observed disparity in allocation ratio. However, this did not render the allocation of study participants to the study arms non-random, as detailed in Table 5.3.1.

Table 5.3.1: Tests of whether differences in allocation rate between study arms resulted in non-equivalent LIFE app use for all study participants included						
Indicator	All eligible participants (N=572)			Participants who reached endpoint (N=247)		
	Control Arm (N=344)	Experiment Arm (N=228)	P-value*	Control Arm (N=148)	Experiment Arm (N=99)	P-value*
	Mean (SD)	Mean (SD)		Mean (SD)	Mean (SD)	
Number of sessions started	2.57 (2.54)	2.75 (2.51)	0.407	4.22 (3.03)	4.61 (2.87)	0.306
Number of sessions completed	2.31 (1.66)	2.54 (1.83)	0.208	3.17 (1.66)	3.51 (1.85)	0.145
Time taken per session (in seconds)	205.83 (141.98)	189.89 (124.64)	0.157	198.54 (91.13)	201.33 (96.54)	0.821
Questions attempted per session	7.31 (3.33)	6.9 (3.48)	0.164	9.15 (1.28)	9.05 (1.24)	0.544
Time per learning task (in seconds)	14.79 (14.82)	15.44 (17.11)	0.006	14.15 (14.44)	15.53 (17.76)	<0.001
Feedback messages per learning session	4.99 (6.707)	4.83 (6.30)	0.637	4.27 (6.55)	4.12 (5.30)	0.699
Performance**	61.94 (24.79)	62.15 (21.03)	0.922	71.35 (15.9)	68.94 (14.86)	0.228
Note: *From difference in means t-test between the two study arms. **Calculated as the percentage of quiz correct on first try out of all learning tasks attempted.						

Of the enrolled participants, 175 (30.59%) provided their background information, and of these 135 reached the experiment's primary endpoint (Figure 5.3.1). From a difference in means test and difference in proportions test, there was no significant difference between the experiment and control group of both those who dropped out and those who reached the study's endpoint by intervention use characteristics (Table 5.3.1) or demographic characteristics (Table 5.3.2). However, healthcare providers of clinical officer cadre, and those who had specialised training who dropped out were significantly different from those who reached study's primary endpoint (Table 5.3.3). The attrition bias is substantively significant only if the differing characteristic is correlated with the experiment's outcome measures (Dumville, Torgerson, & Hewitt, 2006). This will be revisited in the secondary analysis reporting.

Table 5.3.2: Baseline characteristics of study participants where demographic data is available						
	All Healthcare providers who provided background information (n=175)			Healthcare providers who reached the study's primary endpoint (n=135)		
Indicator	Control (n=103)	Experiment (n=72)	P-value*	Control (n=84)	Experiment (n=51)	P-value*
	Mean (SD)			Mean (SD)		
Age (years)	31.2 (8.26)	29.4 (6.99)	0.122	30.81 (8.42)	29.18 (7.24)	0.235
Experience (years)	7.21 (7.05)	6.44 (7.81)	0.516	6.83 (6.66)	6.46 (8.33)	0.793
Sessions Started	3.92 (3.02)	4.22 (3.37)	0.547	4.32 (3.00)	5.35 (3.39)	0.077
	N (%)		P-value [‡]	N (%)		P-value [‡]
<i>Clinical Cadre</i>						
Doctor	37 (35.92%)	30 (41.67%)	0.333	31 (36.9%)	22 (43.14%)	0.236
Clinical Officer	17 (16.50%)	10 (13.89%)	0.263	14 (16.67%)	5 (9.8%)	0.867
Nurse	40 (38.83%)	28 (38.89%)	0.817	31 (36.9%)	21 (41.18%)	0.31
Other	9 (8.74%)	4 (5.56%)	0.529	8 (9.52%)	3 (5.88%)	0.773
<i>Clinical Training Level (whether completed general/speciality training and professionally registered)</i>						
Specialised	23 (22.33%)	11 (15.28%)	0.417	19 (22.62%)	6 (11.76%)	0.942
General Officer	38 (36.89%)	29 (40.28%)	0.723	29 (34.52%)	21 (41.18%)	0.219
Intern	9 (8.74%)	7 (9.72%)	0.529	8 (9.52%)	6 (11.76%)	0.339
Student	33 (32.04%)	25 (34.72%)	0.315	28 (33.33%)	18 (35.29%)	0.408
Done ETAT+?	60 (58.25%)	47 (65.23%)	0.283	49 (58.33%)	33 (64.71%)	0.231
Note: *P-value from difference in means between study arms within progression category. [‡] P-value from a difference in proportions test between study arms within progression category.						

Table 5.3.3: Tests of whether dropout rate is different between study arms using baseline characteristics of study participants where demographic data is available							
Indicator	Healthcare providers who did not reach study’s primary endpoint (n=40)			Healthcare providers who reached the study's primary endpoint (n=135)			Attrition Bias Check
	Control (n=19)	Experiment (n=21)	P-value*	Control (n=84)	Experiment (n=51)	P-value*	
	Mean (SD)			Mean (SD)			P-value**
Age (years)	32.95 (8.26)	29.95 (6.5)	0.188	30.81 (8.42)	29.18 (7.24)	0.235	0.401
Experience (years)	8.84 (7.05)	6.4 (6.59)	0.326	6.83 (6.66)	6.46 (8.33)	0.793	0.471
Sessions Started	2.16 (2.46)	1.48 (0.6)	0.252	4.32 (3.00)	5.35 (3.39)	0.077	N/A
	N (%)		P-value‡	N (%)		P-value‡	P-value††
Clinical Cadre							
Doctor	6 (31.58%)	8 (38.1%)	0.333	31 (36.9%)	22 (43.14%)	0.236	0.148
Clinical Officer	3 (15.79%)	5 (23.81%)	0.263	14 (16.67%)	5 (9.8%)	0.867	0.038
Nurse	9 (47.37%)	7 (33.33%)	0.817	31 (36.9%)	21 (41.18%)	0.31	0.405
Other	1 (5.26%)	1 (4.76%)	0.529	8 (9.52%)	3 (5.88%)	0.773	0.261

Table 5.3.3: Tests of whether dropout rate is different between study arms using baseline characteristics of study participants where demographic data is available							
Indicator	Healthcare providers who did not reach study's primary endpoint (n=40)			Healthcare providers who reached the study's primary endpoint (n=135)			Attrition Bias Check
	Control (n=19)	Experiment (n=21)	P-value*	Control (n=84)	Experiment (n=51)	P-value*	
Clinical Training Level (whether completed general/speciality training and professionally registered)							
Specialised	4 (21.05%)	5 (23.81%)	0.417	19 (22.62%)	6 (11.76%)	0.942	0.041
General Officer	9 (47.37%)	8 (38.1%)	0.723	29 (34.52%)	21 (41.18%)	0.219	0.358
Intern	1 (5.26%)	1 (4.76%)	0.529	8 (10.71%)	6 (11.76%)	0.339	0.394
Student	5 (26.32%)	7 (33.33%)	0.315	28 (33.33%)	18 (35.29%)	0.408	0.116
Done ETAT+?	11 (57.89%)	14 (66.67%)	0.283	49 (58.33%)	33 (64.71%)	0.231	0.082
Note: *P-value from difference in means between study arms within dropout category. **P-value from test in difference in means between due to dropout. † P-value from a difference in proportions test between study arms within progress category. ††P-value from test of difference in dropout rate between the intervention groups.							

The effects of attrition bias appear to be minimal as reported in Table 5.3.1 – 5.3.3. There is good reason to infer that the randomisation generally achieved its purposes of balancing participants from different backgrounds between the intervention arms while also mitigating against selection bias despite the unexpected allocation ratio.

5.3.2 Effect of adaptive feedback on learning gains across study arms

From the performance of the learners, there was a substantive improvement of about 30% above pre-test scores in both the control and experiment arms (Table 5.3.4). Because the overall learning change from LIFE use was large, this may obscure a direct effect from the adaptive feedback. From Levene's test, the assumption that the variance of the performance scores was homogenous held true (Schultz, 1985).

Table 5.3.4: Performance of learners within study arms								
Study Group	Pre-test			Post-test			LIFE smartphone app effect*	
	N	Mean (%)	SD	N	Mean (%)	SD	Correlation**	Effect Size
Control Group	148	56.08	23.17	148	85.81	15.55	0.226	1.031 (0.789-1.274)
Experiment Group	99	50.40	21.42	99	81.82	17.69	0.335	1.272 (0.966-1.577)

Table 5.3.4: Performance of learners within study arms								
Study Group	Pre-test			Post-test			LIFE smartphone app effect*	
	N	Mean (%)	SD	N	Mean (%)	SD	Correlation**	Effect Size
Note: *Effect size based on intra-individual changes in test scores accounting for correlation between the scores in a single-group pre-test post-test design [48]. **Pearson's R correlation between the scores within the study arm.								

The main outcome of interest, which was effect of adaptive feedback on learning gain, was calculated after the second round of play and its calculation is explained in Chapter 5.2.6. From the analysis if the primary outcome, at group level, adaptive feedback had an intervention effect of 0.09 (95% CI: -0.31 – 0.46, p-value = .474) which is both statistically and substantively insignificant. It is noteworthy that the calculation for the effect size computed here does not treat the scores as repeated measures but as independent data. To minimise this bias, a correction adjustment factor of 0.998, explained in [Chapter 4.6.7](#) was applied to the effect size. The adjustment factor suggests that the degree of bias in effect size calculation was minimal given it is a very small correction adjustment. The number of feedback messages provided during the experiment period is reported in Table 5.3.5 illustrating that the control group received no reflective feedback.

Table 5.3.5: Number of feedback messages provided				
<i>Feedback messages for learners who were enrolled into the study (N= 572)</i>				
Feedback	Minimal, N (%)	Reflective, N (%)	Detailed, N (%)	All
Control	2067(27.96)	0(0)	2348(31.76)	4415(59.71)
Experiment	269(3.64)	1097(14.84)	1613(21.82)	2979(40.29)
All	2336(31.59)	1097(14.84)	3961(53.57)	7394
<i>Feedback messages for learners who reached study's endpoint (N= 247) *</i>				
Control	899(29.4)	0(0)	955(31.23)	1854(60.63)
Experiment	48(1.57)	543(17.76)	613(20.05)	1204(39.37)
All	947(30.97)	543(17.76)	1568(51.28)	3058
Note: *Counts are only for the first two complete learning sessions				

5.3.3 Effect of adaptive feedback on individual learning gains

The primary outcome analysis results are likely to suffer from two forms of detection bias: subject-treatment interaction bias and differential time-effect bias (Scott B Morris & DeShon, 2002). For example, in this experiment, because it assumed that the intervention effect is homogenous across varied usage and study population, given Table E.1 in Appendix E, it is highly probable that the unequal

time-difference between the pre-post testing of outcomes in the treatment and control groups may affect the study outcome (i.e. maturation and testing effect). The extent to which the learning intervention affects healthcare providers differentially affects the variability of the outcome measure over time (Scott B Morris & DeShon, 2002). To mitigate against these biases and explore how LIFE use and learner background characteristics might affect learning gains, I conducted secondary analysis using normalised knowledge gain at individual level adjusted for time difference between learning sessions.

When considering differences in LIFE use at individual healthcare provider level, learners who had a space of a week or more between subsequent use of LIFE had significantly lower normalised gains by -0.395 (95% CI: -0.557 to -0.232, p-value < .001) than those who had spaced their learning to an hour or less. Any variation in spacing the self-directed use of LIFE which was a less than or equal to a week did not produce significant changes to the normalised gains. The adaptive feedback mechanism had a significant effect on healthcare providers' normalised learning gains of 0.523 (95% CI: 0.345 to 0.702, p-value < .001). Longer time spent on learning tasks, and a higher ratio of feedback hints provided per attempt at a learning task were significantly associated with lower normalised student learning gains. Healthcare providers' previous face-to-face training in ETAT+ content had no significant effect on their learning gain. This is illustrated in Table 5.3.6, model A.

	Model A: All learners (N = 247)			Model B: Learners with demographic information (N=135)		
	β (SE)	P-value	95% CI	β (SE)	P-value	95% CI
Intercept	0.79(0.042)	<0.001	0.707-0.872	0.851(0.088)	<0.001	0.677-1.026
<i>Ref: Spacing <=1 Hour</i>						
Spacing <= 1 Day	0.027(0.054)	0.609	-0.078-0.133	-0.045(0.072)	0.534	-0.188-0.098
Spacing <= 1 Week	-0.142(0.078)	0.069	-0.294-0.011	-0.28(0.097)	0.005	-0.472--0.088
Spacing <= 1 Month	-0.395(0.082)	<0.001	-0.557--0.232	-0.445(0.129)	<0.001	-0.7--0.19
<i>Ref: Group=Control</i>						
Group = Experiment**	0.523(0.091)	<0.001	0.345-0.702	0.644(0.15)	<0.001	0.347-0.941
Time spent on learning task	-0.09(0.023)	<0.001	-0.135--0.046	-0.036(0.038)	0.345	-0.11-0.039
Help Ratio*	-0.826(0.133)	<0.001	-1.087--0.565	-1.116(0.219)	<0.001	-1.549--0.683

Ref: Done ETAT before = No						
Done ETAT before =Yes	-0.013(0.04)	0.748	-0.092-0.066	-0.092(0.056)	0.107	-0.204-0.02
Ref: Cadre = Nurse						
Cadre = Clinical Officer				0.01(0.085)	0.904	-0.158-0.179
Cadre = Doctor				0.14(0.063)	0.027	0.016-0.265
Cadre =Other				0.124(0.107)	0.247	-0.088-0.336
Ref: Level=Student						
Level = Intern				0.007(0.096)	0.945	-0.184-0.197
Level = General Officer				-0.001(0.07)	0.986	-0.139-0.137
Level = Specialised				-0.045(0.085)	0.595	-0.213-0.123
Age				0.022(0.027)	0.419	-0.031-0.075
Experience				-0.033(0.031)	0.295	-0.094-0.029
Adjusted R ² , Homoscedastic test p-value***	0.344, 0.018			0.407, 0.611		
Note:						
* The number of hints given as a ratio of number of tries a learner had in the second single learning session.						
** The adjusted effect size from the adaptive feedback experiment						
*** Breusch-Pagan test for homoscedasticity, Heteroskedasticity is indicated if p-value < 0.05						

When considering demographic and background characteristics of the healthcare providers, their clinical cadre and level of practice/training had no significant effect on their learning gains except for doctors: their learning gains were significantly higher by 0.14 (95% CI: 0.016 - 0.265, p-value =.027) (Model B, Table 5.3.6). Adjusting for healthcare providers' background characteristics significantly increased the effect of adaptive feedback on individual student normalised learning gains to 0.644 (95% CI: 0.347 to 0.941, p-value: <.001). It also improved the proportion of variance in student normalised gain that was explained by the independent variables by 18.31 % (Model A, Table 5.3.6).

Overall, the independent variables had low multicollinearity in both variants of the regression model in Table 5.3.6, as illustrated in Table 5.3.7. The low values in Table 5.3.7 indicated that I could reliably assess the contribution of predictors to the models. Both models explained between 34.4% - 40.7% of the variance in the normalised learning gains of the healthcare providers using LIFE intervention to train on neonatal emergency care. However, only Model B in Table 5.3.6 satisfied all the statistical modelling assumptions for linear regressions. Also, from Table 5.3.6, the attrition bias was substantively not significant because the differing characteristics ("Clinical Officer" role and

“Specialised Training” level) were not significantly correlated with the learning gain outcome measures (Dumville et al., 2006).

Table 5.3.7: Multicollinearity tests for the learning gain outcome				
	Time	Help Ratio	Age	Experience
Variance Inflation Factor (VIF)	1.041328	1.018684	1.025009	1.046533
Note: With variable correlations (VIF) of less than 1.05, the regression results from secondary analyses (Table 5.3.1) are considered very reliable.				

A fraction of the learners had more than two rounds of play as illustrated in Figure 5.3.2 and 5.3.3 and Table 5.3.8. However, with each round of play the numbers of healthcare providers dropped by around 40-60%. Fitting a longitudinal model using the individual learning gain outcome was inadvisable because the variances of more than one linear combinations of time-effects were close to zero (i.e. Singular) indicating that the model would be overfitting (Ben Bolker, 2019). Mixed-ANOVA and repeated measures ANOVA longitudinal models (also known as analysis of profile models) were deemed inappropriate because they make strong (unrealistic) assumptions that the variances and correlations at each occasion (iteration) are equal i.e. sphericity assumption (Fitzmaurice & Ravichandran, 2008). Also, they are not well suited for mistimed measurements and are more appealing where observations are relatively larger compared to measurement occasions (iterations). Additionally, to produce reliable estimates, they require balanced data over time and complete case analysis, which was not possible given the pattern of the numbers of learner over time (Table 5.3.8).

Table 5.3.8: Learners in study arms over each iteration								
Iteration \ Group	1	2	3	4	5	6	7+	All
Control	242 (25.16%)	148 (15.38%)	75 (7.8%)	40 (4.16%)	27 (2.81%)	12 (1.25%)	17 (1.77%)	561 (58.32%)
Experiment	158 (16.42%)	99 (10.29%)	59 (6.13%)	38 (3.95%)	22 (2.29%)	13 (1.35%)	12 (1.25%)	401 (41.68%)
All	400 (41.58%)	247 (25.68%)	134 (13.93%)	78 (8.11%)	49 (5.09%)	25 (2.6%)	29 (3.01%)	962 (100%)

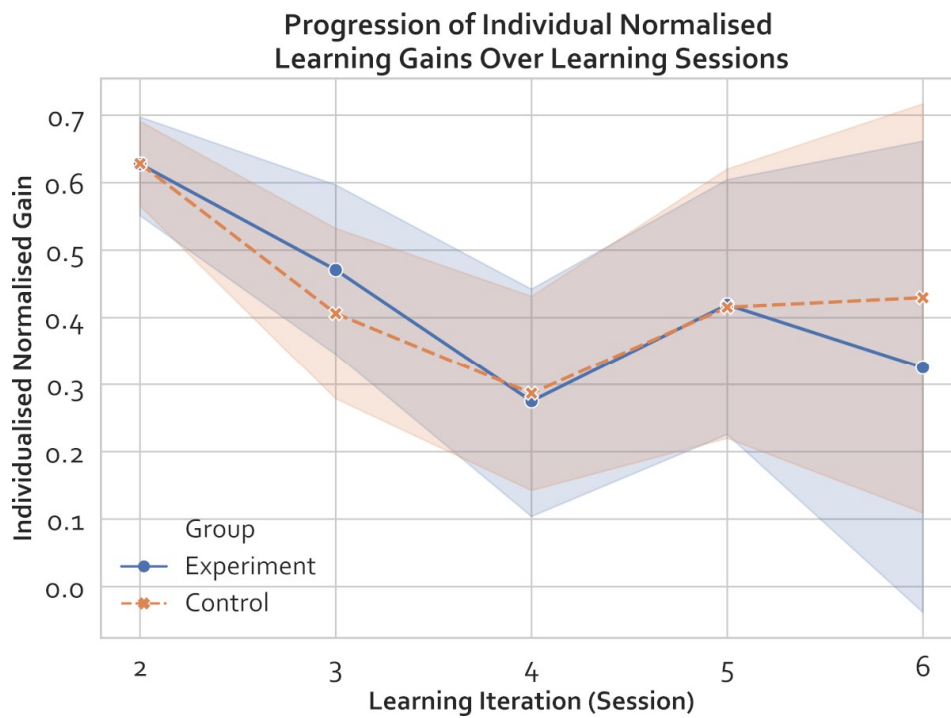


Figure 5.3.2: Change of student learning gains over learning sessions by study arm

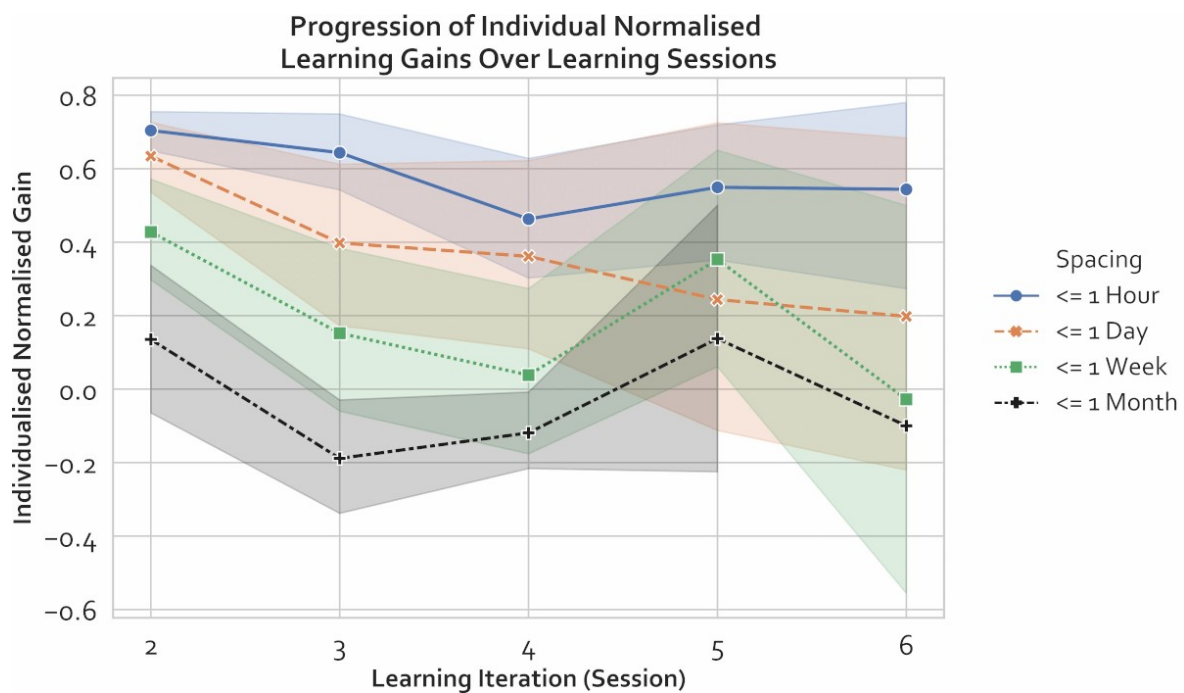


Figure 5.3.3: Change of student learning gains over successive learning sessions by spacing options

To overcome the challenges of using analysis of profiles models, linear mixed effects models were used with raw performance scores and time between measurements. It differs from the typical analysis of profiles longitudinal models by accommodating unbalanced data between the groups and modelling the covariance from repeated measures in a parsimonious way with a relatively small data

sample (Fitzmaurice & Ravichandran, 2008). Table 5.3.9 provided summary statistics from longitudinal analysis of the repeated use of LIFE application where the evaluation is at the individual learning session.

Table 5.3.9: Learning analysis of learning gain using LIFE			
Predictors	Estimates (θ)	95% CI	p-value
(Intercept)	6.18	4.77 – 7.59	<0.001
Spacing Gap [‡]	0.10	0.05 – 0.16	<0.001
Iteration	0.78	0.43 – 1.12	0.003
Experiment (Ref: Control)	-0.46	-0.94 – 0.01	0.056
Spacing Gap*Iteration	-0.06	-0.09 – -0.04	<0.001
Iteration*Group: Experiment	0.07	-0.10 – 0.24	0.444
Random Effects			
Intra-Class Correlation (ICC)	0.41		
N_{User}	400		
N_{Iteration}	7		
Observations	962		
Marginal R² / Conditional R²	0.151 / 0.503		
Note: [‡] Spacing gap is provided as the natural log of minutes between the measurement occasions. The outcome is the performance score as a percentage.			

Healthcare providers had significant learning gains with each repeated use of LIFE. However, longer time spacing between learning sessions significantly reduced the learning gain, and the experiment did not significantly influence the learning gains over LIFE repeated use. The variance explained by fixed factors i.e. Experimental arm and learning spacing options is 15.1%. When accounting for how learners varied over the iterations, this variance increased by 3.3 times to 50.3%. The intraclass correlation (ICC), which is the total variance in the score accounted for by clustering of learners across the iterations was 41%.

A plausible practical interpretation of the Table 5.3.9 findings censored at the seventh iteration is that, overall, there is some evidence that repeating the entire learning session had an increasing effect in bridging the gap in the learning gain across all learning spacing options and experimental arms. The

experiment did not have any additive effect in the overall learning gains of healthcare providers when evaluated at the level of repeated learning sessions. The lack of significant effect of adaptive feedback from findings in Table 5.3.9 is not unexpected given the findings of no effect at learning session and group level in Chapter 5.3.2 In light of the findings in Table 5.3.6, the unit of analysis might impact whether an effect is observed given that the adaptive mechanism was designed to function at the level of the student-step, as opposed to the entire learning session (for example, where adaptive learning pathways with differentiated content are used). Chapters [9](#) explores this notion further by offering a detailed analysis of adaptive mechanism at the level of the individual student-step.

5.4 Summary of findings

The effect of adaptive feedback on healthcare providers' learning gain was found to be $g = 0.09$ (95% CI: -0.31 – 0.46, p -value = .474). In exploratory analysis, using normalised learning gains, when subject-treatment interaction and differential time effect was controlled for, this effect increased significantly to 0.644 (95% CI: 0.35 - 0.94, p -value <.001) with immediate repetition -which is a moderate learning effect- but reduced significantly by 0.28 within a week. With each additional log minute difference between the current and the previous learning session, the performance score of healthcare providers was significantly reduced by 0.06 (-0.09 – -0.04, p -value < .001). The overall learning change from LIFE use in both arms was large and may have obscured a direct effect of feedback.

6 SELF-REGULATED LEARNING BEHAVIOURS OF HEALTHCARE PROVIDERS FROM LOW-INCOME COUNTRIES:

6.1 Recap of Objectives

In this chapter, the analysis is aimed at answering research question two (i.e. *How do healthcare providers' SRL strategies manifest and vary by individual learner characteristics? What influence do these strategies have on healthcare providers' learning gains from using LIFE*). Despite digital learning platforms offering more valuable opportunities for self-directed learning in clinical training (M. K. Lee, 2015), and shown to produce a learning effect from self-directed learning that is equivalent to blended and face-to-face learning for training healthcare providers (George et al., 2019), a lot of learners grapple with self-directed learning on digital environments that offer low instructional support (S. P. Lajoie & Azevedo, 2006; J. Wong et al., 2019). Additionally, for a typical healthcare provider -who is the target learner in this context- digital learning is only a small slice of their day: they are already overworked, underpaid and under-resourced (Sandra Barteit et al., 2019; Couper et al., 2018).

Therefore, any learning intervention should reflect their situated learning needs and address ways for enhancing engagement, inquiry and understanding. Such digital learning interventions also require highly motivated self-directed learning from students to allow them to realise their objectives (Zheng, Rosson, Shih, & Carroll, 2015). In the LICs context, this creates a challenge in providing support to healthcare providers to help them realise their learning objectives on digital learning environments because in general, we do not know what is going on from the learner's perspective. There is a scarcity of evidence from LICs on digital learning platforms that account for the contextual situated learning needs of learners (such as individual healthcare providers) as the learners continue to gain knowledge through them (Bollinger et al., 2013; Car et al., 2019; Edgcombe et al., 2016).

The goal of this chapter is to explore (1) How healthcare providers' SRL behaviours on digital learning platforms manifest and vary by healthcare provider characteristics, and the influence these strategies may have on learning outcomes. In answering this research question, the broader contributions of the findings to the field are:

- ❖ Generating understanding on SRL's behavioural expressions in digital learning platforms for a diverse healthcare provider population in low-income settings, which until now have remained mostly under-researched.
- ❖ Illuminating on considerations of how the links between learning gains and SRL strategies may inform the targeted implementation of digital learning interventions for healthcare providers, such as adaptive instructional scaffolding.

This chapter also generates the foundation upon which the adaptive feedback mechanism can be evaluated conditional on the healthcare providers' SRL profile in the next chapter 7.

6.2 Methodology

6.2.1 Study context, participants, and procedure

In this context, there is not an active registry of healthcare providers from which I could derive a generalisable sample partly due to the lack of up-to-date administrative records of active healthcare providers (North, Shung-King, & Coetzee, 2019). But even if that were to be resolved, fundamentally, getting such a sample would still be challenging due to the scarcity of professionally approved and curated healthcare-based digital learning platforms relevant to clinical practice guidelines from these regions. The research study used two digital learning platforms, a popular regional online-learning computer-based platform (i.e. "Daktari-Online") and LIFE. Both platforms are recognised by relevant clinical professional bodies and are currently being used by healthcare providers from East and Central Africa and South-East Asia regions. This was done to try to attain a context representative sample while reflecting the two most common self-directed digital learning platforms for clinical training in this context: mobile- and computer-based learning modes. The learning content are not massive open online courses; they are controlled by relevant healthcare regulatory bodies with their access targeting clinicians only. The use of these two exemplar platforms was to ensure (1) digital platforms used for learning are varied (mobile- vs. computer-based), (2) other low-income contexts apart from East Africa are represented, and (3) multiple content modalities are represented (i.e. gamified versus non-gamified).

Daktari Online (Daktari Media Africa LLC, 2019) is a platform that seeks to be a one stop shop providing clinical training as continuous professional development for the healthcare provider fraternity in East Africa. It is an online clinical training platform whose objective is to provide an avenue through which the community of healthcare professionals (HCPs) can learn and earn Continuous Professional Development (CPD) points. The award of these CPD points is regulated by clinical professional practice bodies such as Kenya Medical Practitioners and Dentists Board. It utilises a user-subscription Pay-As-You-Go model and is marketed through these health-professional bodies' websites, medical conferences and medical training institutions as a recognised issuer of CPD points. Unlike LIFE, its modular content is varied and not limited to a single clinical speciality or patient population and its reach is limited to East Africa.

Recruitment of survey respondents from Daktari Online was through using a mailing list to invite healthcare providers who had used the platform to learn in the last one year to participate in the study. This was circulated after receiving the necessary ethical approval from the national medical research institutions. Participation was optional and data collected after the web-based informed consent process. Recruitment of survey respondents from LIFE participants occurred at the end of using the LIFE smartphone app. The survey was publicised in parallel to the experiment recruitment (Chapter 5.2.3) and to the same target population. All participants included had engaged in online learning before participating in the survey.

The final study sample was 264, of whom 184 (69.70%) were recruited through LIFE, and 80 (30.30%) were recruited through Daktari Online. This consisted of participants from both public and private hospitals in low income countries such those in Sub Saharan Africa and South East Asia; in clinical cadres such as nurses, clinical officers and medical doctors; and with experience levels varying from students to consultants. Data from these participants would aid in exploring SRL in LICs contexts for continuous professional development across clinical domains, and healthcare experience levels. This sample is not generalisable to healthcare providers from this region. But it is, by clinical cadre and

practice level, probably the most diverse exposition of SRL in healthcare providers from LICs providing insights to SRL's transferability to this context.

6.2.2 Study instruments, variables, and data management

Data was collected using an adapted validated Online SRL Questionnaire tool (OSLQ) which demonstrated internal consistency of scale for all SRL sub-components, as illustrated by previous studies (Kizilcec, Pérez-Sanagustín, & Maldonado, 2017; Tavakol & Dennick, 2011). This survey tool was also preferred due to its inclusion of SRL measures that are based on established instruments and have been shown to have high reliability (Barnard, Lan, To, Paton, & Lai, 2009; Kizilcec et al., 2017; Littlejohn & Milligan, 2015). The exact tool used in this study can be found in Appendix F.1. De-identified and anonymised data was stored on a privately secured Google Firebase database. Additional fields collected from healthcare providers included clinical cadre, clinical practice level, age and experience in years. Survey respondents using LIFE platform had additional learning data that included performance scores as a percentage.

6.2.3 Statistical analysis

Initially, the planned analysis was Confirmatory Factor Analysis (CFA) of SRL, fitted in similar fashion to original study using the same instrument (Barnard et al., 2009) and illustrated by Figure F.1 in Appendix F. This was done to assess whether the SRL measures for healthcare providers from LICs were consistent with the hypothesised conception of SRL from previous studies (Kizilcec et al., 2017; Tavakol & Dennick, 2011) as informed by theory (Winne H.P., 2011; Winne & Hadwin, 1998; Winne et al., 2006).

To cater for the possibility of participant responses to instrument items being skewed, Robust Maximum Likelihood estimator (MLR) was used to minimise inflation of estimates and underestimation of parameters' standard errors. For CFA model fit evaluation, the chi-square test (χ^2) -which measures the difference between the observed covariance matrix and that predicted by the model- was not used because it is very sensitive to sample size and might therefore be misleading. Instead, Comparative Fit Index (CFI) ≥ 0.95 , Tucker-Lewis index (TLI) ≥ 0.95 , Root Mean Square Error of Approximation (RMSEA) $< .06$ and Standardized Root Mean Residual (SRMR) $\leq .08$ for acceptance

based on continuous measures (Hu & Bentler, 1999). These alternatives are robust to, and adjust for, the sample size apart from minimising the negative bias from a chi-square test (Bentler, 1990; Gatignon, 2010).

If the CFA model performance thresholds were not met, I would shift to using Latent Profile Analysis (LPA) (B. Muthén, 2004; B. O. Muthén, 2001; Pastor, Barron, Miller, & Davis, 2007) to explore if there are healthcare providers clusters of latent SRL profiles that are heterogenous with respect to one another. LPA is a technique that identifies groups of healthcare providers in the survey sample who had similar response patterns to the items in the SRL instrument. Within the identified latent profiles, healthcare providers belonging to a specific profile would have a higher probability of sharing similar responses to the survey instrument compared to healthcare providers in other profiles. The procedure for fitting such models is explained in detail elsewhere (B. Muthén, 2004; Stanley, Kellermanns, & Zellweger, 2017).

CFA model fitting that assumes healthcare providers vary quantitatively along the SRL latent dimension and that the qualitative variation in SRL is captured by having different factors for different dimensions of the variation (i.e. measurement invariance). LPA does not assume measurement invariance (it may hold, but it is not forced), and allows healthcare providers in different SRL profiles to vary quantitatively and qualitatively from one another within these sub-scales (Olivera-Aguilar & Rikoon, 2018). Because of the absence of the measurement invariance assumption (which produced the poor fit statistics from CFA), LPA can still make use of these sub-scales, hence the switch over to LPA where CFA fit statistics are undesirable.

Both the CFA and LPA models would be fitted using maximum likelihood estimator approach in *Mplus* 7.4 to estimate model parameters (Pastor et al., 2007). For LPA, the initial and final iterations would be 4000 and 40 respectively, and they were settled upon in efforts to minimise structural under-identification. To determine the final number of latent profiles to use, Lo-Mendel-Rubin(LMR) test cut-off p-value < 0.05 would be used, combined with cut-off > 0.8 for normalised measure of aggregated classification uncertainty (i.e. Entropy) (Nylund, Asparouhov, & Muthén, 2007). Higher values of

entropy higher values indicate more precise assignment of individuals to latent profiles. To ensure replicability of findings from the LPA model, it was run using 20 randomly drawn samples of the participants to test whether I could replicate the decision for the number of profiles selected. Logistic regression would be conducted by using the profile membership as a dependent variable and demographic characteristics as possible predictor variables. This would be done to explore how demographic characteristics were associated with the identified SRL profiles.

6.2.4 Post-hoc analysis

ANOVA was used to explore if there were statistically significant differences in the SRL strategies across the identified SRL profiles. Where a violation of the normal distribution assumption in the responses to the instrument's variables existed, Kruskal–Wallis non-parametric test (Spurrier, 2003) would be used instead. Post-hoc pairwise t-tests/Whitney-Mann-Wilcoxon tests (Fay & Proschan, 2010) were conducted to drill down to where possible differences in the SRL profiles might lie. This is because ANOVA/Kruskal-Wallis test would only give an indication of whether there is enough evidence to suggest that in a particular SRL strategy, one or more SRL profiles might be significantly different from another SRL profile but no indication of which profile those may be or by how much (López, Valenzuela, Nussbaum, & Tsai, 2015).

6.2.5 Summary

The use of LPA was, admittedly, chosen in retrospect to early findings from CFA analysis proposed in the methods section of this chapter. This was primarily because the CFA thresholds necessary for reliable inference were not met for the CFA analysis to continue. Essentially, it was included as alternative exploratory analysis because the SRL measures for healthcare providers from LICs were not consistent with the hypothesised conception of SRL from previous studies (Kizilcec et al., 2017; Tavakol & Dennick, 2011) as informed by theory (Winne H.P., 2011; Winne & Hadwin, 1998; Winne et al., 2006). This will become self-evident in the findings in the next section of [Chapter 6](#).

6.3 Results

6.3.1 Baseline characteristics of the participants

The data reported was collected between 7th February 2019 and 31st July 2019. The final study sample was 264 of whom, 184 (69.70%) had been recruited through LIFE, and 80 (30.30%) recruited through *Daktari Online*. The breakdown of demographic characteristics of the participants is provided in Table 6.3.1 and 6.3.2². Majority of respondents were medical officers -equivalent to junior doctors in UK- (N=100, 37.9%), followed by nurses (N=72, 27.3%). They were predominantly between 25 and 44 years of age (N=186, 71.5%) and practicing health workers not undergoing any training (N=162, 61.4%).

Table 6.3.1: SRL survey participants by platform			
Platform	Daktari Online	LIFE	All
	N (%)	N (%)	N (%)
Clinical Cadre			
Consultant (<i>Doctors whose practice requires training in a speciality area</i>)	26(9.85%)	10(3.79%)	36(13.64%)
Medical Officer (<i>Doctors equivalent to "Junior Doctor" or GP cadres in the UK</i>)	49(18.56%)	51(19.32%)	100(37.88%)
Clinical Officer (<i>A cadre commonly found in Sub-Saharan Africa</i>) *	4(1.52%)	27(10.23%)	31(11.74%)
Nurse	0(0%)	72(27.27%)	72(27.27%)
Other	1(0.38%)	24(9.09%)	25(9.47%)
Clinical Level			
Active Practice (Not undergoing any training)	50(18.94%)	112(42.42%)	162(61.36%)
Active Practice (Currently undergoing training)	24(9.09%)	12(4.55%)	36(13.64%)
Student	1(0.38%)	55(20.83%)	56(21.21%)
Other	5(1.89%)	5(1.89%)	10(3.79%)
Age (in years)			
18-24	1(0.38%)	43(16.29%)	44(16.67%)
25-34	34(12.88%)	86(32.58%)	120(45.45%)
35-44	32(12.12%)	34(12.88%)	66(25%)
45-54	6(2.27%)	16(6.06%)	22(8.33%)
55-64	5(1.89%)	1(0.38%)	6(2.27%)
65+	2(0.76%)	0(0%)	2(0.76%)
Missing	0(0%)	4(1.52%)	4(1.52%)
Experience (in years) **	10.98 (10.05)	6.92 (7.47)	8.15 (8.52)

² The statistical analysis code used for Chapter 6 can be found in the following GitHub Repository:

<https://github.com/AfrikanaCollective/Statistical-Programming-Projects/tree/master/Structural%20Equation%20Modelling%20et%20al>

Table 6.3.1: SRL survey participants by platform			
Platform	Daktari Online	LIFE	All
All	80(30.3%)	184(69.7%)	264
Note: *Wilson, A., Lissauer, D., Thangaratnam, S., Khan, K. S., MacArthur, C., & Coomarasamy, A. (2011). A comparison of clinical officers with medical doctors on outcomes of caesarean section in the developing world: meta-analysis of controlled studies. <i>Bmj</i> , 342, d2600. **Experience in years provided as Mean (Standard Deviation).			

Table 6.3.2: Baseline characteristics of survey respondents								
	Age Bracket	18-24	25-34	35-44	45-54	55-64	65+	All
Cadre	Clinical Level							
Consultant	Active Practice (Currently undergoing training)	0	1	2	1	0	0	4(1.5%)
	Active Practice (Not undergoing any training)	0	2	21	6	2	1	32(12.1%)
Medical Officer	Active Practice (Currently undergoing training)	0	14	8	0	0	0	22(8.3%)
	Active Practice (Not undergoing any training)	0	32	10	2	4	0	48(18.2%)
	Student	15	10	0	0	0	0	25(9.5%)
	Other	0	2	1	1	0	1	5(1.9%)
Clinical Officer	Active Practice (Currently undergoing training)	2	4	1	0	0	0	7(2.7%)
	Active Practice (Not undergoing any training)	0	12	4	1	0	0	17(6.4%)
	Student	2	4	0	0	0	0	6(2.3%)
Nurse	Active Practice (Currently undergoing training)	0	3	0	0	0	0	3(1.1%)
	Active Practice (Not undergoing any training)	2	24	15	11	0	0	52(19.7%)
	Student	13	3	1	0	0	0	17(6.4%)
Other	Active Practice (Not undergoing any training)	3	7	3	0	0	0	13(4.9%)
	Student	6	2	0	0	0	0	8(3.0%)
	Other	1	0	0	0	0	0	1(0.4%)
All		44(16.7%)	120(45.5%)	66(25%)	22(8.3%)	6(2.3%)	2(0.8%)	260
Note: Four respondents (1.52%) had missing age data.								

There were no survey responses from “Daktari-Online” users who were nurses and only one who was a student (Table 6.3.1). Clinical officers were more likely to be users of LIFE with Daktari-Online having on average, three more years of experience compared to LIFE users (Table 6.3.1).

6.3.2 Summary of OSLQ tool responses

Overall, the responses to the SRL questionnaire tool had high internal consistency, with a Cronbach alpha value of 0.986. This is better than compared to the validating study’s 0.92. This internal reliability was also confirmed by an omega coefficient of 0.903, which is more robust to the instrument’s length than the Cronbach alpha, and a more appropriate measure of scale reliability (Dunn, Baguley, & Brunsden, 2014). This is illustrated in Table 6.3.3.

Table 6.3.3: Descriptive Statistics of SRL Questionnaire Tool Items						
Item	Item Description	Mean	SD	Skew	Kurtosis	Shapiro-Wilk Test p-value*
Goal Setting, Cronbach Alpha: 0.841, Omega Coefficient: 0.895 RMSEA = 0.000, SRMR = 0.000; CFI = 1.000, TLI = 1.000						
I1	I had personal goals for the [platform] training scenario	3.93	1.34	-1.08	-0.03	< 0.001
I2	I had realistic performance targets for the [platform] training scenario	3.79	1.32	-0.87	-0.38	< 0.001
I3	Goals helped me manage my learning through [platform]	4	1.27	-1.2	0.41	< 0.001
Strategic Learning, Cronbach Alpha: 0.944, Omega Coefficient: 0.758 RMSEA = 0.176, SRMR = 0.049; CFI = 0.910, TLI = 0.731						
I4	When planning my learning, I adapted strategies that have worked before	3.82	1.32	-0.91	-0.33	< 0.001
I5	I considered alternative options <i>not provided</i> which might be better solutions	3.22	1.44	-0.21	-1.26	< 0.001
I6	I asked myself questions on learning content before I began learning using [platform].	3.62	1.45	-0.66	-0.94	< 0.001

Table 6.3.3: Descriptive Statistics of SRL Questionnaire Tool Items

Item	Item Description	Mean	SD	Skew	Kurtosis	Shapiro-Wilk Test p-value*
I7	I pre-planned my study time to better accomplish my learning tasks.	3.31	1.5	-0.34	-1.33	< 0.001
Task strategies, Cronbach Alpha: 0.991, Omega Coefficient: 0.684 RMSEA = 0.073, SRMR = 0.022; CFI = 0.981, TLI = 0.944						
I8	I asked myself how what I'm learning is related to what I already know	4.03	1.26	-1.31	0.66	< 0.001
I9	I changed strategies when I did not meet my personal learning targets	3.59	1.35	-0.69	-0.7	< 0.001
I10	I had difficulty understanding new information on the [platform]'s course(s)	2.27	1.41	0.74	-0.81	< 0.001
I11	While learning through the [platform], I made notes to help me understand	3.27	1.56	-0.3	-1.44	< 0.001
Elaboration (Includes Item I8), Cronbach Alpha: 0.95, Omega Coefficient: 0.72 RMSEA = 0.000, SRMR = 0.000; CFI = 1.000, TLI =1.000						
I12	I tried to match the [platform] content with other sources of information	3.73	1.42	-0.82	-0.71	< 0.001
I13	Previous work experience was useful when learning through the [platform]	4.24	1.14	-1.66	1.98	< 0.001
Self-Evaluation (Includes Item I5), Cronbach Alpha: 0.986, Omega Coefficient: 0.702 RMSEA = 0.000, SRMR = 0.000; CFI = 1.000, TLI =1.000						
I14	The score at the end reflected how well I have learned	3.96	1.25	-1.17	0.37	< 0.001
I15	I have a good sense of what I have learned through the [platform]	4.23	1.1	-1.63	2.09	< 0.001
Help Seeking, Cronbach Alpha: 0.979, Omega Coefficient: 0.746 RMSEA = 0.056, SRMR = 0.03; CFI = 0.984, TLI =0.968						
I16	If I am having trouble learning, I prefer to figure it out by myself	3.13	1.5	-0.23	-1.36	< 0.001

Table 6.3.3: Descriptive Statistics of SRL Questionnaire Tool Items						
Item	Item Description	Mean	SD	Skew	Kurtosis	Shapiro-Wilk Test p-value*
I17	If I did not understand a task, I would prefer help to be provided before any initial attempt at it.	3.63	1.39	-0.62	-0.92	< 0.001
I18	I had access to peers for help when needed	2.69	1.56	0.24	-1.48	< 0.001
I19	When I needed help, I asked my colleagues for more information	2.77	1.62	0.15	-1.61	< 0.001
I20	The help I asked from my colleagues was different from what the [platform] provided me with.	2.23	1.45	0.72	-0.94	< 0.001
Note: [platform] is a placeholder for the learning platform that was being referred to by the survey; Questionnaire Cronbach Alpha: 0.986, Omega Coefficient: 0.903; Item scales were: “<i>Very true for me</i>”: (5), “<i>True for me</i>”: (4), “<i>Quite true for me</i>”: (3), “<i>Sometimes true for me</i>”: (2), “<i>Not at all true for me</i>”: (1). *Quantifies how likely the instrument’s item was drawn from a normal distribution. A p-value < 0.05 indicates that the item was not normally distributed.						

The convergent validity of the scale could also be inferred from the moderate to strong correlations for almost all the sub-scales except for help-seeking sub-scale as illustrated in Figure 6.3.1. Figure 6.3.1 item descriptions are provided in Table 6.3.3, and its cell values represent pairwise correlation, with only statistically significant correlations shaded.

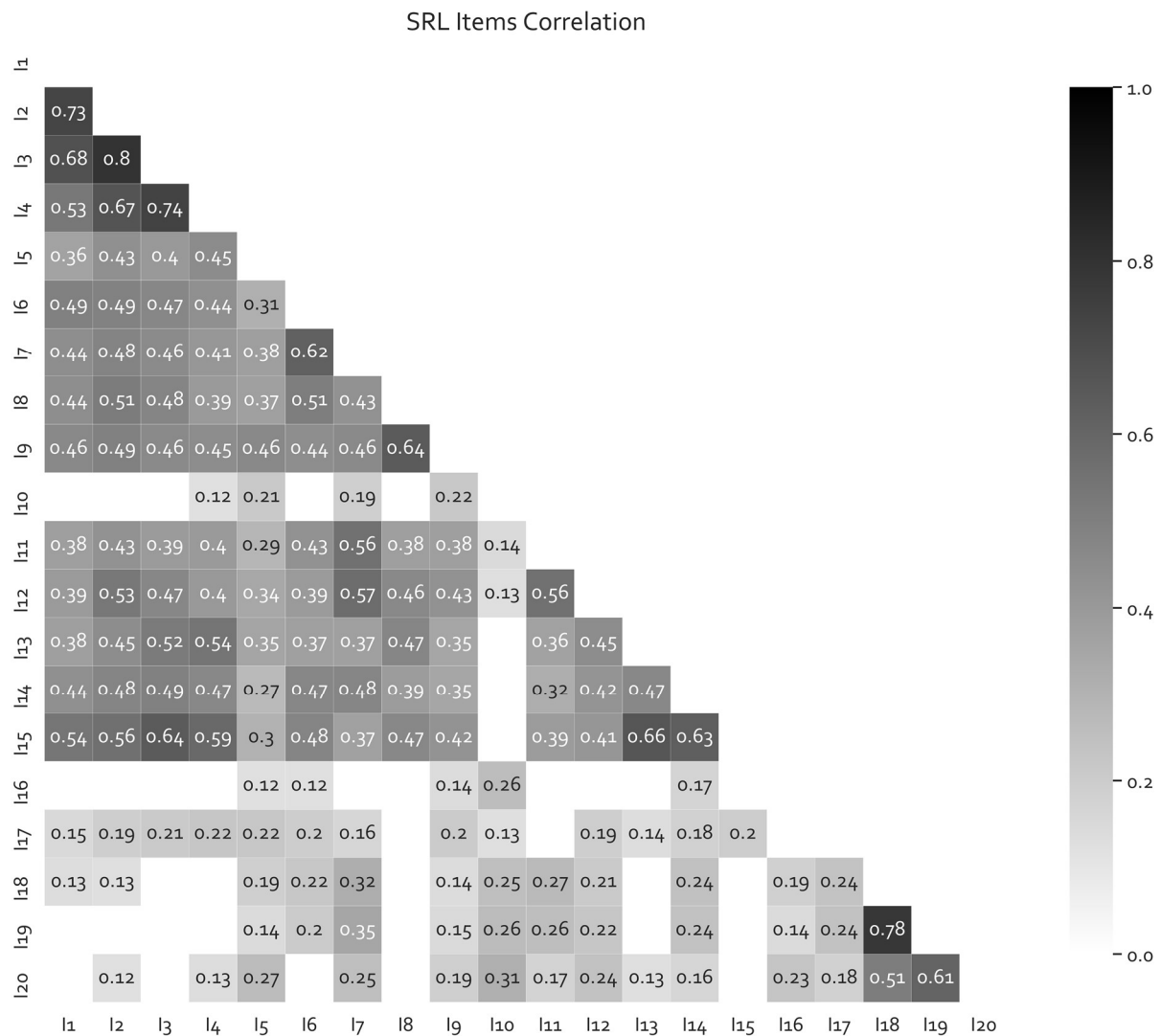


Figure 6.3.1: SRL Item pairwise correlation representation.

The overall CFA model performance in the form of SRL using the original digital self-regulated learning scale was (RMSEA = 0.078, SRMR = 0.079; CFI = 0.867, TLI = 0.845). Based on the preferred thresholds outlined in Chapter 6.2.3, this was not satisfactory performance and was poorer than the original study deriving the tool (RMSEA = 0.04, CFI = 0.96, TLI = 0.95) but better than the validating study's results (RMSEA = 0.081, CFI = 0.666) which modified the tool slightly (Jansen, Van Leeuwen, Janssen, Kester, & Kalz, 2017). The poor CFA model fit statistics are despite this instrument having better composite and sub-scale Cronbach alpha scores than those of the validating studies (Barnard et al., 2009; Jansen et al., 2017; Kizilcec et al., 2017). Further examination of the SRL components revealed strategic learning and task-strategy sub-scales to have particularly unsatisfactory RMSEA and TLI scores (Table 6.3.3).

The SRL measures for healthcare providers from LICs were not consistent with the hypothesised conception of SRL from previous studies (Kizilcec et al., 2017; Tavakol & Dennick, 2011) as informed by theory (Winne H.P., 2011; Winne & Hadwin, 1998; Winne et al., 2006). This is indicative that important opportunities to conceptualise/reconceptualize SRL in this population and context exist which 9 begins to address. The rest of the analyses reported from here to the end of in Chapter 6 focuses on latent profile analysis (LPA).

6.3.3 Latent Profile Analysis (LPA) Results

Table 6.3.3 illustrates results from fitting LPA models which were implemented to identify the SRL profiles of healthcare providers from LICs. These model fit statistics are results from comparing a model with K profiles to a model with (K-1) profiles. The test statistic from Lo-Mendel-Rubin (LMR-LRT) indicates whether the model with K profiles is significantly better than the model with K-1 profiles. If it is lower than 0.05, the model with K profiles is preferred, otherwise, the model with K-1 profiles is preferred.

Table 6.3.4: Latent profile analysis results for Self-Regulated Learning							
Number of Profiles	Log Likelihood	Parameters	AIC	BIC	ABIC	LMR-LRT p-value	Entropy
2	-8509.11	62	17142.22	17363.93	17167.36	0.004	0.990
3	-8257.47	84	16682.94	16983.32	16716.99	0.013	0.957
4	-8122.58	103	16451.16	16819.48	16492.92	0.149	0.940
5	-8045.83	128	16347.67	16805.39	16399.56	0.493	0.960
AIC – Akaike Information Criterion, BIC – Bayesian Information Criterion, ABIC - Sample Adjusted Bayesian Information Criterion, LMR – Lo-Mendell-Rubin test							

There was a consistent decrease in model deviance (log likelihood) with the increasing number of profiles up until five profiles. Even though AIC, BIC and ABIC model estimates indicate superior model fit with four profiles, LMR's p-values point to three profiles being optimum. Despite LMR p-value indicated a three-profile model as the optimum cut-off, further examination of the three- versus four-profile model revealed distinctive substantive differences in favour of the four-profiles model (Figure 6.3.2, Appendix F: Figure F.2). This was further supported by a high level of profile distinctness

(Entropy) and relatively better model fit statistics (Table 6.3.4). Based on the high entropy, lower BIC, and ABIC, and substantive difference in patterns in four-profile model, I therefore settled on using four SRL profiles for subsequent analyses. In comparison to previous studies which have used this survey instrument, a four-profile solution is not unusual, with an entropy of 0.94 being higher than that from findings of those previous studies (Barnard-Brak, Paton, & Lan, 2010).

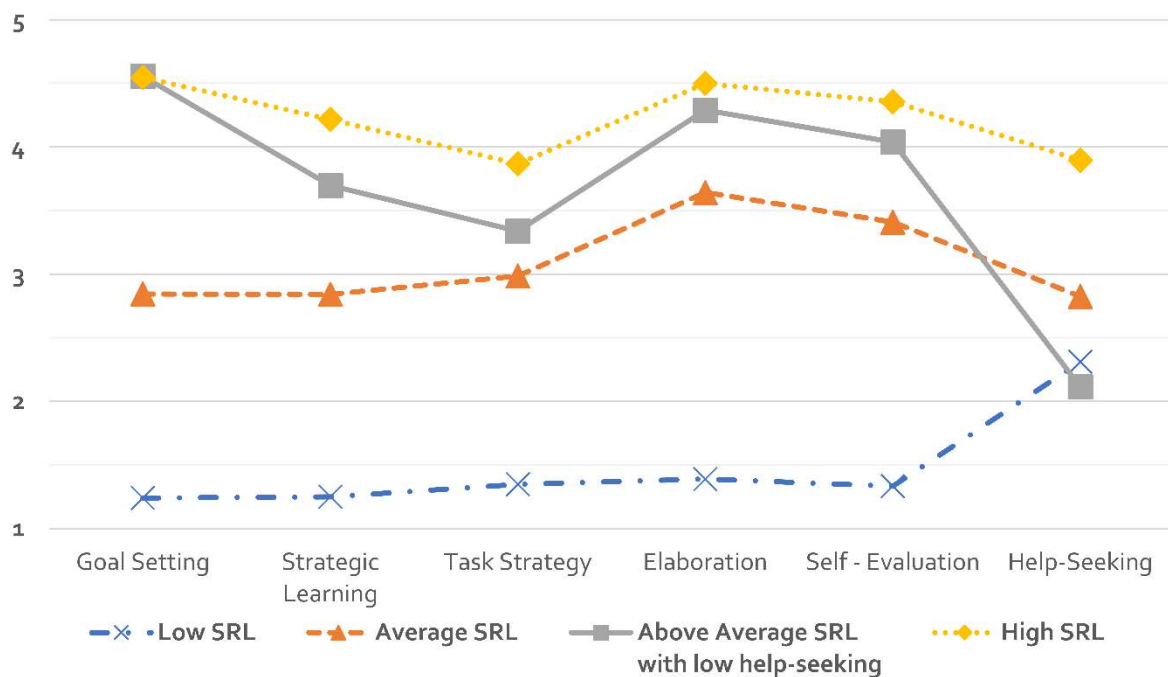


Figure 6.3.2: Self-Regulated Learning profiles of healthcare providers in LICs based expected item means from OSLQ tool

From the four-profile LFA model, Figure 6.3.2 illustrates the estimated sub-scale means of with items grouped by SRL skill components (i.e. sub-scales). The four identified profiles could be characterised as *Low SRL* (N=28, 10.61%), *Average SRL* (N=61, 23.11%), *Above average SRL with low help-seeking* (N=108, 40.91%), and *High SRL* (N=67, 25.38%) profiles. In general, it is encouraging that for the healthcare providers in this sample, it is only a minority (10.6%) that demonstrated low self-regulation when it comes to learning using digital devices.

Healthcare providers in the Low SRL profile using Daktari-Online were on average, less likely to have set goals, had low strategic learning, elaboration and self-evaluation SRL skills, and were more likely to seek help when they embarked on learning compared to those in the same profile using LIFE (Figure 6.3.3). In general, healthcare providers using Daktari-Online were, on average, more likely to be in

the *Low* and *Average* SRL profiles. Those using LIFE were, on average, more likely to be in the *Above Average* and *High* SRL profiles (Table F.1 Appendix F).

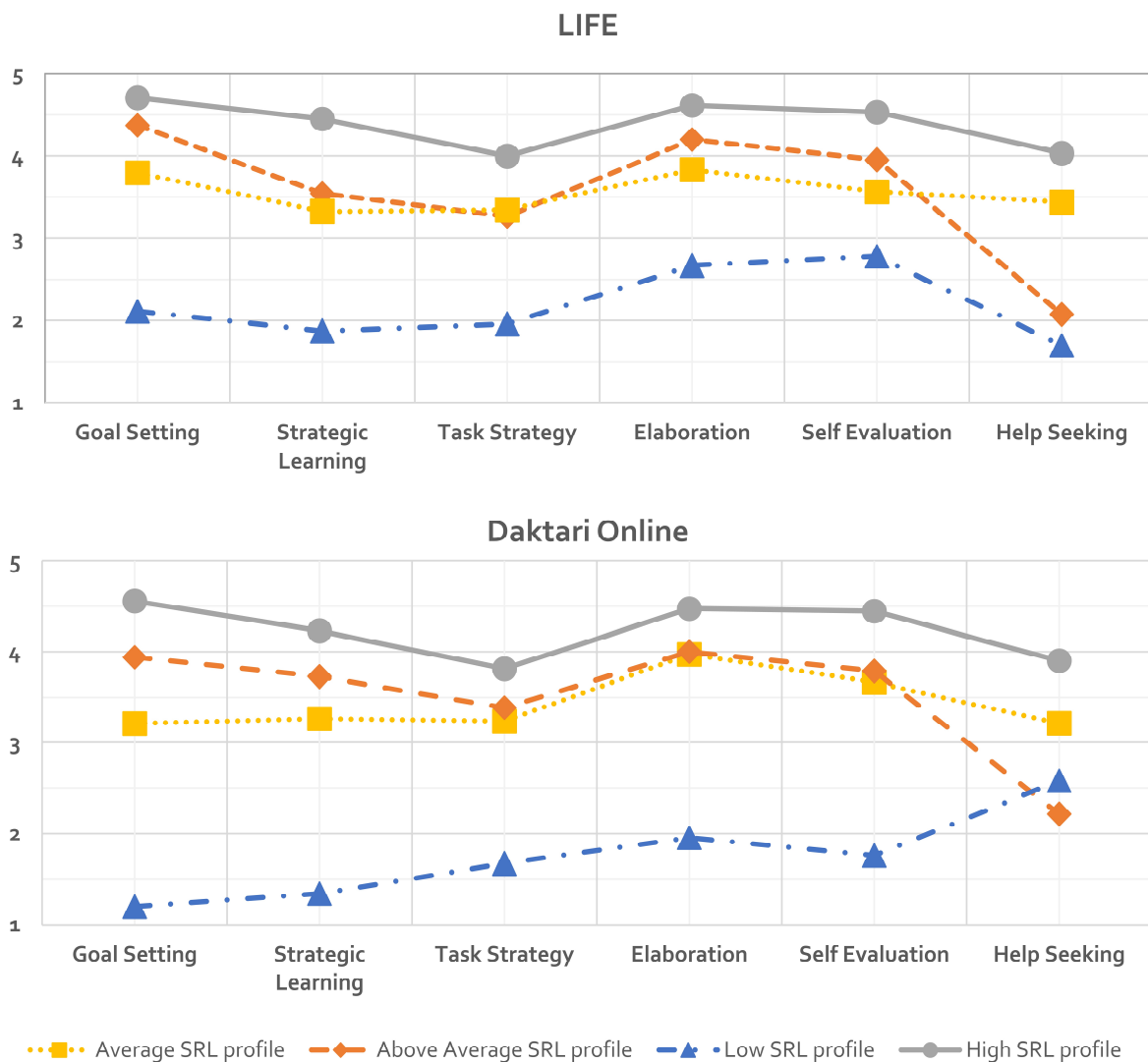


Figure 6.3.3: SRL skills by learning platform

Due to a lot of missingness in the demographic sub-categories of data (Table 6.3.1, Table 6.3.2), it would be inadvisable to conduct a multinomial regression since the model ended up being rank-deficient. A workaround that was used was multiple logistic regressions with binary outcome of either being in the specific LPA profile or not. The results of these logistic regressions are detailed in Table 6.3.5. When the odds of profile membership are broken by background characteristics, high self-regulated learners significantly differed from the other profiles based on experience. In this profile, with each increase in a year of experience, health providers had 43% increase in the odds of being in

high SRL profile (Table 6.3.5). Counter-intuitively, consultants -who are the highest skilled healthcare providers- had three times the odds of being in low SRL profile compared to other clinical speciality levels. Use of computer-based online learning platforms (i.e. *Daktari-Online*) was significantly associated with the odds of being in the average or above-average SRL profile. Healthcare providers using computer-based learning platforms were 13 times more likely to be in the average SRL profile compared to those using the smartphone learning platform. They also had a 36% decrease in the odds of being in the average SRL profile with a unit increase in their years of experience (Table 6.3.5). Healthcare providers in the above-average SRL profile with low help-seeking were 89% less likely to be users of the *Daktari-Online* platform. They were also 70% and 65% less likely to be consultants and clinical officers (Table 6.3.5). Due to the few numbers in the low self-regulated learning profile, some estimates in that column are considered highly unreliable and findings for this profile should be interpreted cautiously. It also exemplifies the rank-deficient challenge for conducting a multinomial logistic regression.

Consideration was given to whether there were significant differences in the means of SRL skills based on profile group membership using the post-hoc Whitney-Mann-Wilcoxon pairwise tests. The results revealed significant differences in all the multiple comparisons of all the SRL skills between the four SRL profiles except for *Task Strategy* Skill between the “Average” and “Above-Average” SRL profiles (p-value=0.878). SRL strategies for *Help-Seeking* between the “Low” and “Above-Average” SRL profile were borderline significant (p-value=0.049). Non-parametric procedure was used because of the violation of the assumption of normal distribution of all the SRL instrument items (Table 6.3.3). Based on Kruskal-Wallis test statistic, there is a significant difference in the SRL strategies of healthcare providers in low-income contexts from the four-profiles membership model. This is despite not using the three SRL profile as indicted by the LMR p-value cut-off threshold (Table 6.3.4). This finding, together with the pairwise non-parametric post-hoc tests, suggests overall significant differences across all profiles in almost all SRL subscales. These results are illustrated in Table 6.3.6.

Table 6.3.5: Logistic regression results predicting profile membership

Profiles	Low (N = 28)			Average (N=61)			Above Average with low help seeking (N=108)			High (N=67)		
Predictors	OR	95% CI	p-value	OR	95% CI	p-value	OR	95% CI	p-value	OR	95% CI	p-value
(Intercept)	0.06	0.02 – 0.21	<0.001	0.07	0.03 – 0.19	<0.001	2.24	1.17 – 4.30	0.015	0.27	0.13 – 0.54	<0.001
<i>Clinical Cadre, Ref: Medical Officer</i>												
Consultant	3.2	1.08 – 9.53	0.036	1.26	0.47 – 3.42	0.644	0.3	0.10 – 0.91	0.033	0.94	0.31 – 2.84	0.912
Clinical Officer	0.28	0.03 – 2.41	0.248	2.18	0.65 – 7.26	0.206	0.35	0.14 – 0.91	0.032	2.53	0.98 – 6.52	0.055
Nurse	0*	0.00 – Inf*	0.989	1.92	0.62 – 5.92	0.258	0.56	0.26 – 1.18	0.128	1.64	0.72 – 3.72	0.236
Other	0*	0.00 – Inf*	0.993	4.93	1.46 – 16.68	0.01	0.37	0.13 – 1.04	0.059	1.38	0.45 – 4.25	0.57
<i>Level, Ref: Active Practice (Not training)</i>												
Active Practice (Training)	2.25	0.76 – 6.66	0.142	0.49	0.19 – 1.29	0.149	0.75	0.28 – 1.97	0.557	1.78	0.67 – 4.71	0.249
Student	1.12	0.17 – 7.24	0.901	1.15	0.45 – 2.91	0.774	0.8	0.39 – 1.63	0.544	1.27	0.58 – 2.77	0.544
Other	1.37	0.13 – 14.97	0.796	1.26	0.29 – 5.42	0.756	0.48	0.09 – 2.69	0.406	1.76	0.37 – 8.51	0.479
Experience	0.84	0.52 – 1.35	0.47	0.64	0.43 – 0.96	0.031	1.1	0.79 – 1.52	0.581	1.43	1.04 – 1.96	0.026
<i>Platform, Ref: LIFE</i>												
Daktari-Online	3.34	1.00 – 11.18	0.051	13.06	4.50 – 37.91	<0.001	0.11	0.05 – 0.27	<0.001	0.41	0.16 – 1.07	0.068
Nagelkerke's R²			0.349			0.225			0.230			0.095

Note: *Numbers too few to compute reliable estimates

Responses to help seeking sub-scale were all low in all groups apart from high SRL profile, with task strategy also having poor performance. These two SRL sub-scales had very high variance as indicated by their standard deviation (Table 6.3.6). Considering SRL theory, this might be indicative that interventions to improve persistence and effort-regulation of healthcare providers in LICs when using digital learning platforms might maximise their learning outcomes. I posit that the improvement of persistence and effort-regulation in digital learning might also be influenced by what help-seeking support is provided to the healthcare providers when learning and how they would prefer it to be packaged.

Table 6.3.6: Self-regulated learning skills by profiles					
	Low SRL profile	Average SRL profile	Above Average SRL with low help-seeking	High SRL Profile	Kruskal - Wallis Test*
SRL Skill	<i>Mean (SD)</i>	<i>Mean (SD)</i>	<i>Mean (SD)</i>	<i>Mean (SD)</i>	<i>p-value</i>
Goal Setting	1.29 (0.67)	3.09 (1.07)	4.00 (1.02)	4.51 (0.64)	< 0.001
Strategic Learning	1.36 (0.77)	3.03 (1.09)	3.45 (1.30)	4.33 (0.91)	< 0.001
Task Strategy**	1.59 (1.14)	3.03 (1.19)	3.22 (1.42)	3.97 (1.24)	< 0.001
Elaboration	1.83 (1.33)	3.43 (1.17)	3.88 (1.15)	4.44 (0.79)	< 0.001
Self-Evaluation	1.73 (1.18)	3.22 (1.14)	3.70 (1.16)	4.39 (0.87)	< 0.001
Help-Seeking	2.16 (1.42)	3.09 (1.16)	2.24 (1.38)	4.00 (1.20)	< 0.001
Note: *Kruskal-Wallis Test performed row-wise i.e. by SRL skill. **Task Strategy skill for the “Average” SRL profile was not statistically different from “Above-Average” SRL profile. All the rest pairwise evaluations were significantly different from each other.					

6.3.4 Influence of SRL strategies on healthcare providers’ learning gains.

For the second research question in this chapter, learning performance data in the form of normalised learning gains (Marx & Cummings, 2007) from healthcare providers who used LIFE were used. The records of healthcare providers performance from *Daktari-Online* were not in a form capable of generating learning gain scores and neither was it covered by the ethical approvals for this thesis. It was therefore not requested. The calculation of learning gains used here whose scale lies between -1 and 1, is described in detail in [Chapter 5.2.6](#) and elsewhere (Marx & Cummings, 2007). When the normalised learning gain is positive, the healthcare providers were making performance gains in what they learnt when it is evaluated as a ratio of the amount they could have learned. From the analysis

of learning gains from LIFE users, learners in the *above-average* and *high* SRL profiles had substantively more repeated plays with improved performance than healthcare providers in the other SRL profiles (Figure 6.3.4).

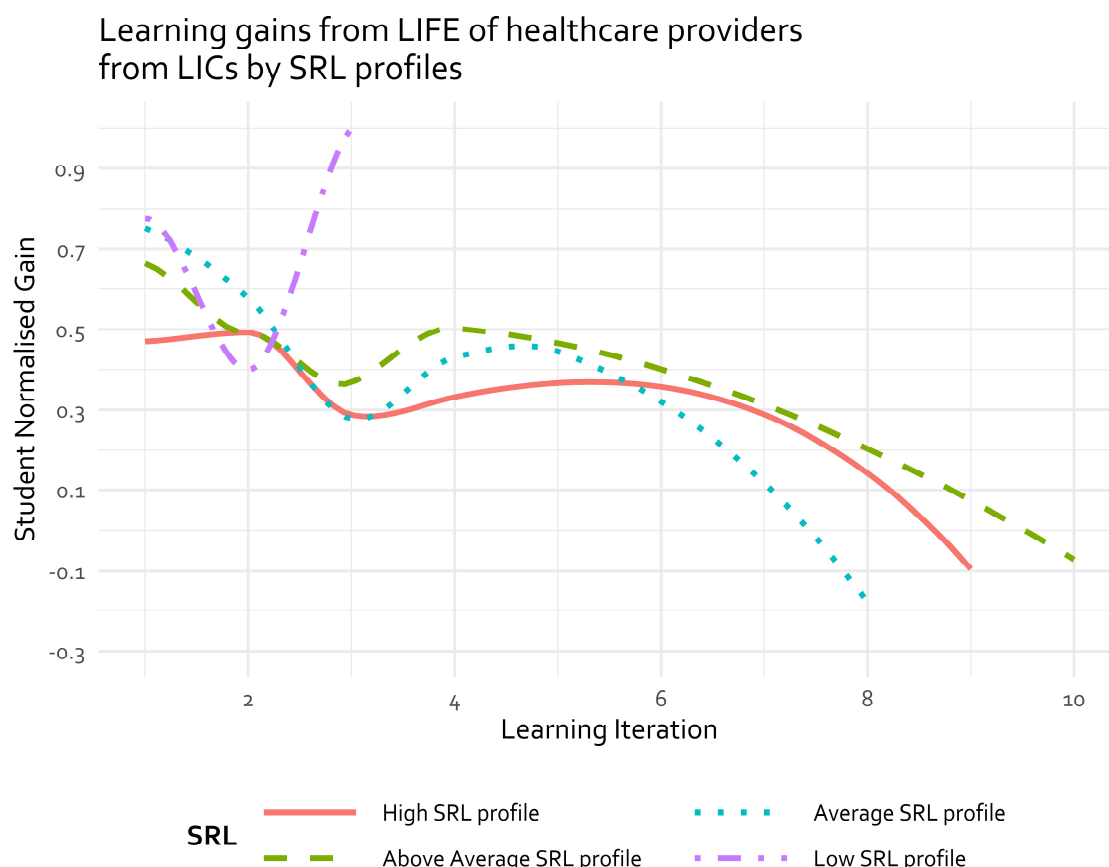


Figure 6.3.4: Learning gains of healthcare providers by SRL profiles

With each repeated learning session, feedback on individual learning tasks were adapted for some of the learners who were struggling to make gains (Chapter 5). All included SRL profiles had positive learning gains with each reinforcement learning cycle, but healthcare providers in low SRL profiles - who were more likely to be consultants (Table 6.3.5)- had the least number of repeated learning sessions but achieved maximum normalised learning gains by their third learning session. Consistent with low task strategy SRL skill, learning persistence of the healthcare providers in all profiles apart from *low* SRL was not associated with maximised learning gains. The average SRL profile learners ceased learning when their performance stagnated or deteriorated (Figure 6.3.4).

Actively practicing healthcare providers who not undergoing any formal training appeared to be the only category of learners with moderate to maximum learning gains from the learning content in LIFE, while the other categories' learning gains would typically be erratic and tail off towards negative gains (Figure 6.3.5).

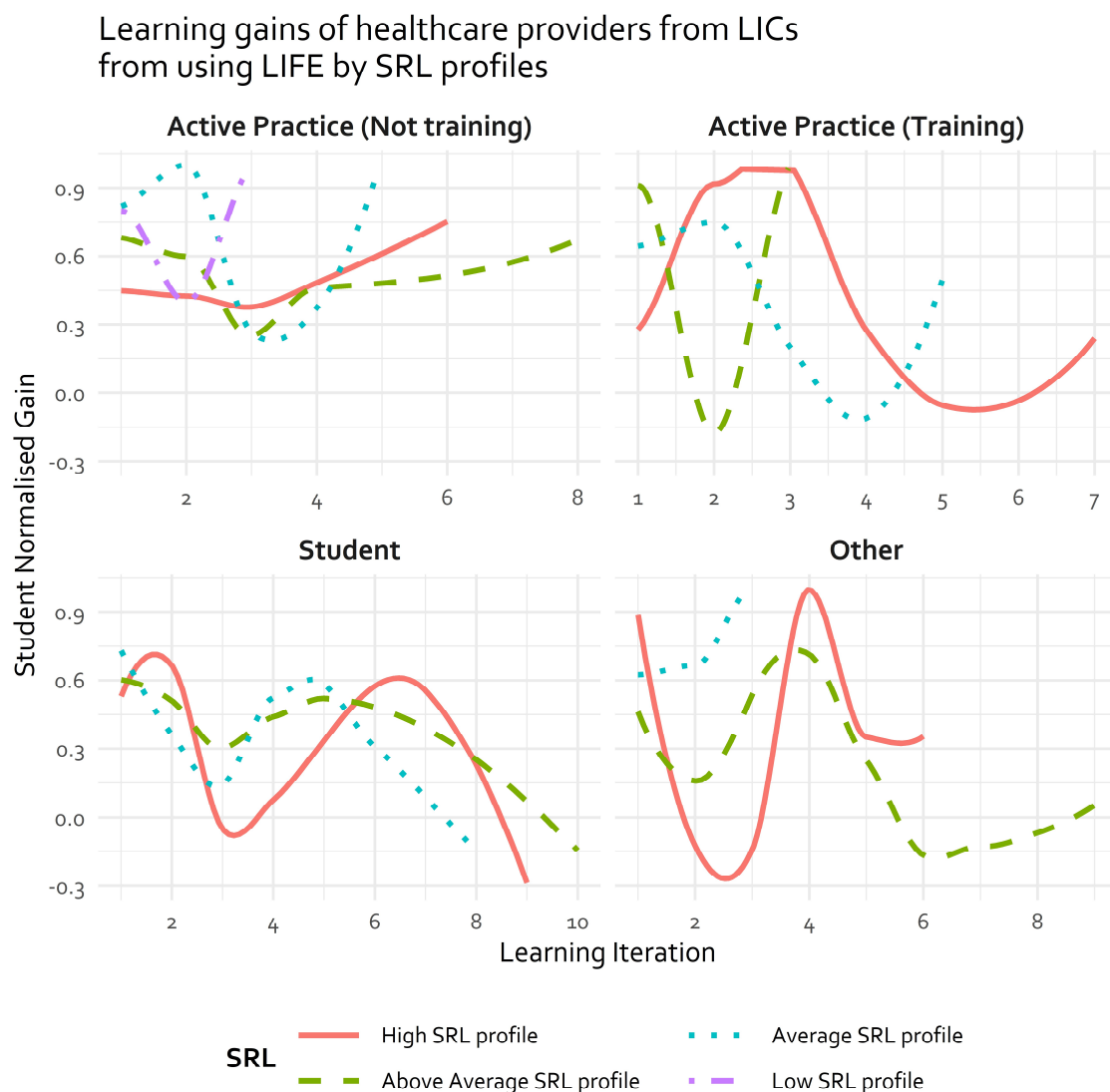


Figure 6.3.5: SRL profiles by clinician training status

Given the varied but low task strategies and help-seeking behaviours in the SRL profiles (Table 6.3.6), it is conceivable that the lack of maximising on the possible learning gains from using the LIFE platform in all SRL profiles apart from low SRL profile is attributable to a low persistence and effort-regulation with limited help provided/sought. In the low SRL profile, which has higher odds of having consultants

than any other cadre (Table 6.3.5), and demonstrated maximisation of learning gains with minimal reinforcing of the learning content, and earliest drop-out (Figure 6.3.4), this might also be indicative that the learning content in general might not be as challenging for learners in this profile after several practice sessions.

6.4 Summary of findings

The analysis in Chapter 6 was intended to generate insights about Self-Regulated Learning (SRL) by healthcare providers in low-income countries on digital learning environments, whether they be computer- or smartphone-based. In summary, for the first research question in this section, using Latent Profile Analysis (LPA) the findings demonstrated that there are four SRL profiles in this population of learners: *high*, *above-average with low help-seeking*, *average*, and *low* SRL profiles. Healthcare providers in the *high* SRL profile scored relatively highly in each SRL sub-scale (skills) with the opposite being true for those in *low* SRL profile. Healthcare providers in average SRL profile had average and relatively monotonic response to all SRL sub-scales while those in the above-average SRL profile having low response to help-seeking behaviours and above-average to high response in the other SRL sub-scales.

Consultants had three times the odds of being in the low SRL group, and online computer-based learners had 13 times the odds of being the average SRL profile. The above-average with low help-seeking SRL learners were 89% less likely to have used *Daktari-Online* and at least 65% less likely to be clinical officers or consultants. A unit increase in the years of experience in providing healthcare significantly increased the odds of healthcare providers being in the high SRL profile by 1.5 times. Only 10.1% of healthcare providers who took part in the study would be regarded as having low self-regulated learning. In addressing the second research objective for this section, when normalised learning gain was considered for learners where that information was available (i.e. LIFE users), compared to healthcare providers in the low SRL profile (who were mostly consultants), the other three SRL profiles had lower learning gains which were never maximised even with more repeated learning cycles. This might be because the LIFE learning content is typically not easily and affordably

accessible to lower clinical cadres; they are likely to have had limited previous training exposure to it. Even within self-directed learning, it might require more instructional support to ensure maximum learning gains are realised as illustrated in Figure 6.3.4. Additional qualitative work was conducted to understand how the context of clinical practice and learning platform use informed healthcare providers' perception of their own SRL. This is outlined in [Chapter 9](#).

7 SELF-REGULATED LEARNING PROCESS OF HEALTHCARE PROVIDERS FROM LICS

7.1 Recap and Objectives

Learning interventions that account for individual health workers' needs to deliver differentiated instructional support as they continue to develop knowledge through it (i.e., adaptive learning) are virtually non-existent in Low Income Countries (Bollinger et al., 2013; Edgcombe et al., 2016; Greenhalgh, 2001). Such learning adaptations are common in the Intelligent Tutoring Systems (ITSs) literature, where while not necessarily cognisant of contexts like SSA or clinical training, learner interactions within the digital learning platform tend to be tracked as a sequence of student-driven steps (Kurt Vanlehn, 2006). That is, when a student attempts a learning task (step), the ITS records whether it was successful, whether any system-initiated assistance was provided, and may provide instructional support based on the learner's performance. These kinds of data points are what are used for student learning needs' modelling and subsequently adapting content (Chi et al., 2011). The learning tasks that produce these data points represent unique Knowledge Components (KC) which, reflect learning "...concepts, principle, fact, skill, schema, production rule, misconception..." (Kurt VanLehn et al., 2007). Common student modelling approaches for KCs are Learning Factor Models and include, Additive Factor Models (AFMs) (Cen et al., 2008), and Performance Factor Models (PFMs) (Pavlik Jr et al., 2009), with detailed explanations of these provided in [Chapter 3.4.2](#).

The aim of this phase two study was to analyse learning data derived from clinical training on smartphone devices in LICs in order to (1) Evaluate from the most common learning factor models (i.e. AFM and PFM), which is more suited to represent student learning in this context by assessing their predictive ability on learners' content mastery; and (2) Evaluate how the effect of adaptive feedback learning gains within a serious-gaming platform was moderated by practice opportunity, learner's own self-regulation and the level of adaptive support provided using the model identified in objective one. This chapter is aimed at addressing the second part of research question one (i.e. *How is the effect of*

adaptive feedback identified in Chapter 5 moderated by (a) the learning opportunity for knowledge mastery, and (b) the level of adaptive support provided?)

The decision to use these models instead of BKT to explore how self-directed learning opportunity is impacted by learning gains and adaptive feedback provided was four-fold: (1) The analysis in this chapter was done to allow for more robust learning curve estimation from practice opportunities that controlled for the potential lurking and confounding effects of learner-specific proficiency learning item-difficulty, and self-regulated learning behaviours, which up to this point have remained unattended to (2) it would also elucidate the effect of adaptive feedback mechanism at the individual learning task level (3) it would allow for the comparative evaluation of knowledge tracing approaches' assumptions validity and predictive accuracy on LIFE data, an evidence gap raised in Chapter 3.4, and finally (4) it was done to overcome some of the standard BKT challenges explained in detail to in [Chapter 3.4.1](#).

7.2 Methodology

7.2.1 Study design, setting and participants

This study was an observational study of healthcare providers from both public and private hospitals, in clinical cadres such as nurses, clinical officers and medical doctors, with experience levels varying from students to consultants. The participants were from LICs in six sub-continental regions (Appendix E, Table E.1). They were the same learners enrolled into to the experimental study detailed in Chapter 5.2.3 with the same eligibility criteria i.e. participants had to be either in training for, or active in, bedside clinical care in LICs.

7.2.2 Study Variables

Data collection strategies were the same as those in the experimental evaluation of adaptive feedback (Chapters 5.2.11 – 5.2.12) and illustrated in Figure 4.5.1. For the purposes of the proposed analysis, the outcome of interest (i.e. independent variable) was specified as getting each answer correct on the first try in current learning session, and, is binary. The dependent variables of interest were: (a) Time spent on learning task, (b) Cumulative count of all previous attempts per learning tasks across all sessions (i.e. opportunities), including their breakdown into previous failures and successes, (c)

Whether feedback had been provided for each unique try per learning task, (d) Level of feedback provided including the adaptive feedback evaluated (Chapters 5.2.4 – 5.2.5) and, (e) the spacing between the learning opportunities. Additionally, learner's self-regulated learning profiles generated from analysis outlined in Chapter 6 were included as a dependent variable where healthcare providers had data on that variable.

7.2.3 Statistical Methods, Missing Data, And Sensitivity Analyses

Data manipulation and statistical analyses were performed using R software's *glmer* package (Bates, Mächler, Bolker, & Walker, 2014; R Core Team, 2013). Variables of interest are reported using their mean value and standard deviation. In the first phase of the analyses, to address its first objective, PFMs and AFMs were used to construct the learning models from LIFE data. To compare the ability of these models to make accurate predictions of knowledge mastery by healthcare providers while minimising selection bias, the classification accuracy and area under curve (AUC) for the Receiver Operating Curve (ROC) model performance metrics were used. The classification accuracy metric represents the true proportion of correctly predicted learner attempts by the models where, in this case, the cut-off probability threshold of 0.5 was used. Prediction probability > 0.5, was labelled as learner being predicted to get the next attempt correct on first try in the current learning session, with the converse being true. This decision threshold is a commonly used cut-off for binary outcomes of interest. The AUCROC metric here represents how capable the models are in distinguishing by discriminating, the outcome of a learner's attempt by tracking the trade-off between the model's sensitivity (true positive rate) and its false positive rate (i.e. 1 - specificity) (Narkhede, 2018).

These metrics were computed from training the two modelling approaches on a fraction of the data (i.e. training set) and testing their predictive performance on remaining fraction (i.e. the test dataset). Starting with 10% of dataset as the training dataset and the remaining 90% as the testing dataset, 10-fold cross-validation (Kohavi, 1995) was applied to the training dataset, from which the classification accuracy was calculated. Using the trained model, the AUC score was computed from fitting the model on the testing dataset. This procedure was repeated with a larger training dataset, incremented by

10% in each iteration until 50% of the original dataset had been used as the training dataset. The data used as the training was randomly selected from the full dataset, with data from a single learner constrained to be in either the training or the test dataset, but not both. This was done to evaluate these models' predictive ability on the performance of 'unseen' learners while minimising overfitting. The model with the better performance would be used for subsequent analyses. More explanation of the cross-validation procedure is provided in [Appendix G](#).

In this analysis, users with incomplete learning sessions were included. Evaluation of whether learning data from incomplete learning sessions was significantly different from the complete learning sessions was done. This evaluation also included assessing whether the variables themselves were normally distributed, then performing unpaired two-samples t-test if they were and Mann-Whitney-Wilcoxon test if they were not (Gibbons & Chakraborti, 1991; Wilcoxon, 1990). This was done to provide a context and strong basis for inference based on learning session completion rate.

7.2.4 Summary

The analysis outlined in the methods section of this chapter was meant to assess how learning curves of healthcare providers using LIFE evolve over practice opportunities and the role that varied levels of feedback, spaced repetition and their SRL profiles played in this evolution. This analysis also sought to evaluate how well better modelling approaches that address the shortcomings of BKT model perform on LIFE learning data. In chapter 9, qualitative approaches adopted ensured the views of the healthcare providers about such learning tools and approaches are echoed in this thesis and inform the understanding and interpretation of findings of chapters 5-7.

7.3 Results

7.3.1 Summary of LIFE play game statistics

The data reported were observed between 1st February 2019 and 31st July 2019. These are the same healthcare providers as those reported in Chapter 5. Of the 572 healthcare providers who downloaded and started playing the LIFE game in this period, 400 (69.93%) had at least one complete learning session. Table 7.3.1 provides a summary of variables of interest from all healthcare providers who

used LIFE for learning, broken down into those who had a complete learning session and those who did not complete a learning session³. Demographic data within the LIFE application was collected at the end of a learning session, had an opt-in approach with additional informed consent requested, and was optional. Not all participants provided this information. These demographic data, together with the accompanying survey data were used to generate Self-Regulated Learning (SRL) profiles of the healthcare providers have been described in detail in Chapters [5.1](#) and [6](#).

Table 7.3.1: Summary statistics of data from LIFE game play sessions				
Indicator	Complete (N=946)	Incomplete (N=563)	P-value*	All (N=1509)
	<i>Mean (SD)</i>	<i>Mean (SD)</i>		<i>Mean (SD)</i>
Average performance (%) **	72.46 (25.67)	41.99 (38.33)	<0.001	61.09 (34.32)
Time spent on each question before an attempt	14.33 (6.92)	16.87 (11.85)	0.012	15.27 (9.15)
Previous cumulative tries on a question across all sessions i.e. Opportunities	3.07 (3.08)	3.83 (3.64)	0.3412	3.46 (4.09)
Average count of feedback messages provided per learning session	5.36 (6.5)	4.24 (6.57)	<0.01	4.94 (6.55)
Feedback***	Complete, N (%)	Incomplete, N (%)		All, N (%)
<i>None</i>	9743 (86.34)	1541 (13.66)		11284 (60.21)
<i>Essential</i>	1726 (73.51)	622 (26.49)		2348 (12.53)
<i>Reflective</i>	859 (76.49)	264 (23.51)		1123 (5.99)
<i>Detailed</i>	2632 (66.05)	1353 (33.95)		3985 (21.26)
Learning Spacing	Complete, N (%)	Incomplete, N (%)		All, N (%)
<i>None</i>	316 (55.24)	256 (44.76)		572 (37.91)
<i><= 1 Hour</i>	364 (70)	156 (30)		520 (34.46)
<i><= 1 Day</i>	131 (71.58)	52 (28.42)		183 (12.13)
<i><= 1 Week</i>	83 (69.17)	37 (30.83)		120 (7.95)
<i><= 1 Month</i>	47 (63.51)	27 (36.49)		74 (4.9)
<i>> 1 Month</i>	19 (47.5)	21 (52.5)		40 (2.65)
Note: *From a Mann-Whitney-Wilcoxon test checking whether the true difference in mean ranks between the groups is equal to zero. **Average performance based on number of learning tasks attempted and not the total number of tasks a learner should attempt. ***Counts are across all sessions and can be multiple per session				

³ The statistical analysis code used for Chapter 6 can be found in the following GitHub Repository: <https://github.com/AfrikanaCollective/Statistical-Programming-Projects/tree/master/Knowledge%20Tracing/LFA>

The nature of incomplete learning sessions was such that they had significantly lower performance, had significantly less feedback messages per learning session, spent significantly more time on the learning tasks even though the number of opportunities at learning task was not significantly different from complete learning sessions. The fewer feedback messages on more time on task for the incomplete learning sessions might be indicative of a relatively more apprehensive learning behaviour using LIFE. This was explored qualitatively in the detailed analysis outlined in [Chapter 4.9](#) and [9](#).

7.3.2 Model Performance: AFM versus PFM

This section is devoted to the evaluation of the accuracy of AFM and PFM models in representing student learning from LIC context by assessing their predictive ability on learners' content mastery. The AFM model had a mean classification accuracy of 0.689 with a minimum and a maximum of 0.674 and 0.696 respectively over the different training-validation dataset ratios. Its mean AUCROC score was 0.748 with a minimum and maximum of 0.737 and 0.764 respectively. On the other hand, the PFM model had a mean classification accuracy of 0.767 with a minimum and maximum of 0.751 and 0.776 respectively. The PFM's AUCROC score was better, with a mean of 0.85 and a minimum and maximum of 0.841 and 0.856 respectively. On average, the PFM model had 11.3% higher classification accuracy and 13.6% higher prediction accuracy than the AFM model. In both instances, PFM outperformed AFM in accurately predicting and distinguishing the performance from 'unseen' healthcare providers (i.e. the hold-out test dataset) on LIFE knowledge components.

The accuracy and AUC scores indicate that the AFMs are more likely to be overfitting the learning data from LIFE than the PFM model, given the relatively weaker performance of AFM in prediction of learning data from 'unseen' students on 'seen' steps. Another plausible explanation for the difference in performance might be that the disaggregation of learning opportunities into previous failures and successes is more informative of LICs healthcare providers' learning behaviours than just keeping a tally of the count of previous opportunities. Overall, it appears that PFM model has best model performance, minimises bias of overfitting compared to the AFM model, and would explain learner behaviour more accurately. This is supported by p-value of <0.001 from DeLong's test comparing the

AUCROC curves from AFM and PFM models for all training-test dataset thresholds. The DeLong test evaluated whether the AUC score from the PFM model was significantly higher than that of the AFM model (Robin et al., 2011).

In summary, when considering how best to evaluate how the effect of feedback on LICs healthcare providers' learning gains is moderated by previous opportunity for knowledge mastery and level of adaptive support provided, from the findings in this section, using the PFM model is the preferred approach to answer that question.

7.3.3 Moderation effect of level of feedback given previous learning opportunity

Overall, consistent with the findings from Chapter 5.3, being in the study arm that was exposed to adaptive feedback mechanism did not have any significant effect on the odds of getting the next attempt at a learning task correct on the first try (Table 7.3.2). However, in all levels of feedback provided, the odds of getting the next try correct on first try was significantly higher (OR: 5.58, 95% CI: 4.29–7.26, p-value: <0.001) where the learner received reflective feedback compared to receiving detailed feedback (Table 7.3.2, model A). The experimental group was perfectly mediated by the reflective feedback level.

Table 7.3.2: Results from PFM model of probability of getting the attempt correct on first try based on learning opportunity						
Model	Model A: All learners			Model B: Learners with demographic information available		
Predictors	Odds Ratios	95% CI	p-value	Odds Ratios	95% CI	p-value
(Intercept)	0.84	0.72 – 0.99	0.036	1.64	1.30 – 2.08	<0.001
In the Experiment Group (Ref: No)						
Yes	1.07	0.90 – 1.28	0.44	0.91	0.72 – 1.15	0.446
Has undertaken ETAT+ course before (Ref: No)						
Yes	0.99	0.83 – 0.17	0.907	1.09	0.86 – 1.38	0.488
Opportunity*Feedback Level (Ref: Detailed)						
None	0.56	0.49 – 0.65	<0.001	1.17	0.98 – 1.41	0.09
Essential	1.36	1.15 – 1.60	<0.001	0.86	0.69 – 1.08	0.192
Reflective	5.58	4.29 – 7.26	<0.001	4.13	2.99 – 5.70	<0.001

Table 7.3.2: Results from PFM model of probability of getting the attempt correct on first try based on learning opportunity						
Model	Model A: All learners			Model B: Learners with demographic information available		
Opportunity* Task Time	0.9	0.85 – 0.95	<0.001	1	0.94 – 1.07	0.904
Opportunity* Learning Spacing (<i>Ref: None</i>)						
<= 1 Hour	1.09	0.94 – 1.27	0.244	0.8	0.65 – 0.99	0.039
<= 1 Day	0.61	0.51 – 0.75	<0.001	0.42	0.33 – 0.53	<0.001
<= 1 Week	0.46	0.38 – 0.56	<0.001	0.32	0.24 – 0.41	<0.001
<= 1 Month	0.2	0.16 – 0.26	<0.001	0.2	0.15 – 0.27	<0.001
> 1 Month	0.24	0.16 – 0.34	<0.001	0.36	0.24 – 0.53	<0.001
Opportunity * SRL Profile (<i>Ref: High SRL</i>)						
Low SRL				1.99	1.16 – 3.42	0.013
Average SRL				1.7	1.39 – 2.08	<0.001
Above Average SRL				1.07	0.93 – 1.24	0.337
Task Time *SRL Profile (<i>Ref: High SRL</i>)						
Low SRL				0.91	0.56 – 1.47	0.705
Average SRL				1.05	0.90 – 1.22	0.535
Above Average SRL				0.98	0.89 – 1.08	0.689
Random Effects						
Intraclass Correlation Coefficient (ICC)	0.1879			0.1219		
N	572 Student			164 Student		
Student-Step Observations	18120			8342		
Marginal R ² / Conditional R ²	0.534 / 0.614			0.529 / 0.581		
AIC/BIC/Log Likelihood/Deviance	16320.1/16640.1/-8119.1/16238.1			8235.2/8565.5/-4070.6/8141.2		

After controlling for the moderating effect of the healthcare provider's SRL profile on both time on task, and count of previous opportunity, the reflective feedback was the only level of feedback associated with a significantly higher odds of getting the next attempt at a learning task correct after considering previous opportunities (Table 7.3.2). When considering how the LIFE learning curve is associated with healthcare providers spacing of their own learning, the odds ratio of a healthcare provider correctly recalling the right task drops by between 20% within the hour to 70% within a week of going through the learning session (Table 7.3.2). This finding supports the emphasis for Low-Dose-High-Frequency (LDHF) model of learning for clinical training, especially in a space where the learners'

exposure to such clinical scenarios might be rare. From these findings, smartphone based LDHF training might have higher impact on knowledge mastery if emphasis for spaced repetition is on weekly basis, in line with findings from Chapter 5.3. The reflective feedback (including the adaptive mechanism used to provide it), had the most consistent, positive effect on the learning rate of the healthcare providers and minimised wheel spinning with each learning opportunity in all varied spacing of self-regulated repetition (Figure 7.3.1).

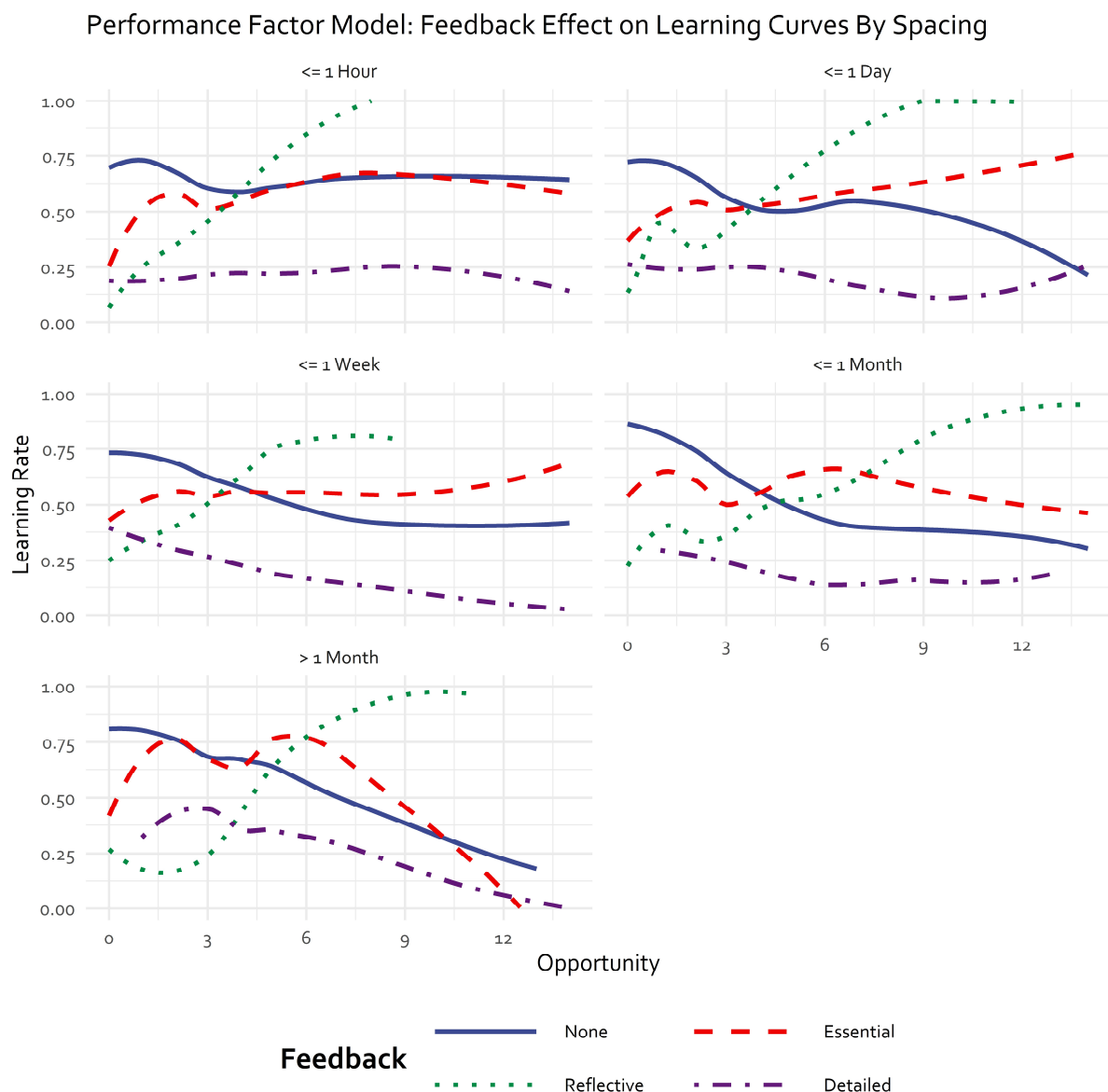


Figure 7.3.1: PFM Learning curves of feedback effect by self-regulated spacing of learning repetition

From the adaptive feedback experiment in Chapter 5, a key difference between the experimental groups was that only those in the experiment arm received reflective feedback (Table 5.3.5). Previous research posits that a good learning curve would display a logarithmic pattern indicating that students increased their accuracy with practice and thus that learning appears to be well-supported (Peddycord-Liu et al., 2018). The learning curve from reflective feedback is most reflective of a positive effect on learning rate. Healthcare providers in the *High* and *Above-Average* SRL profiles did not have significant differences in their learning rate given learning opportunities (Model B, Table 7.3.2).

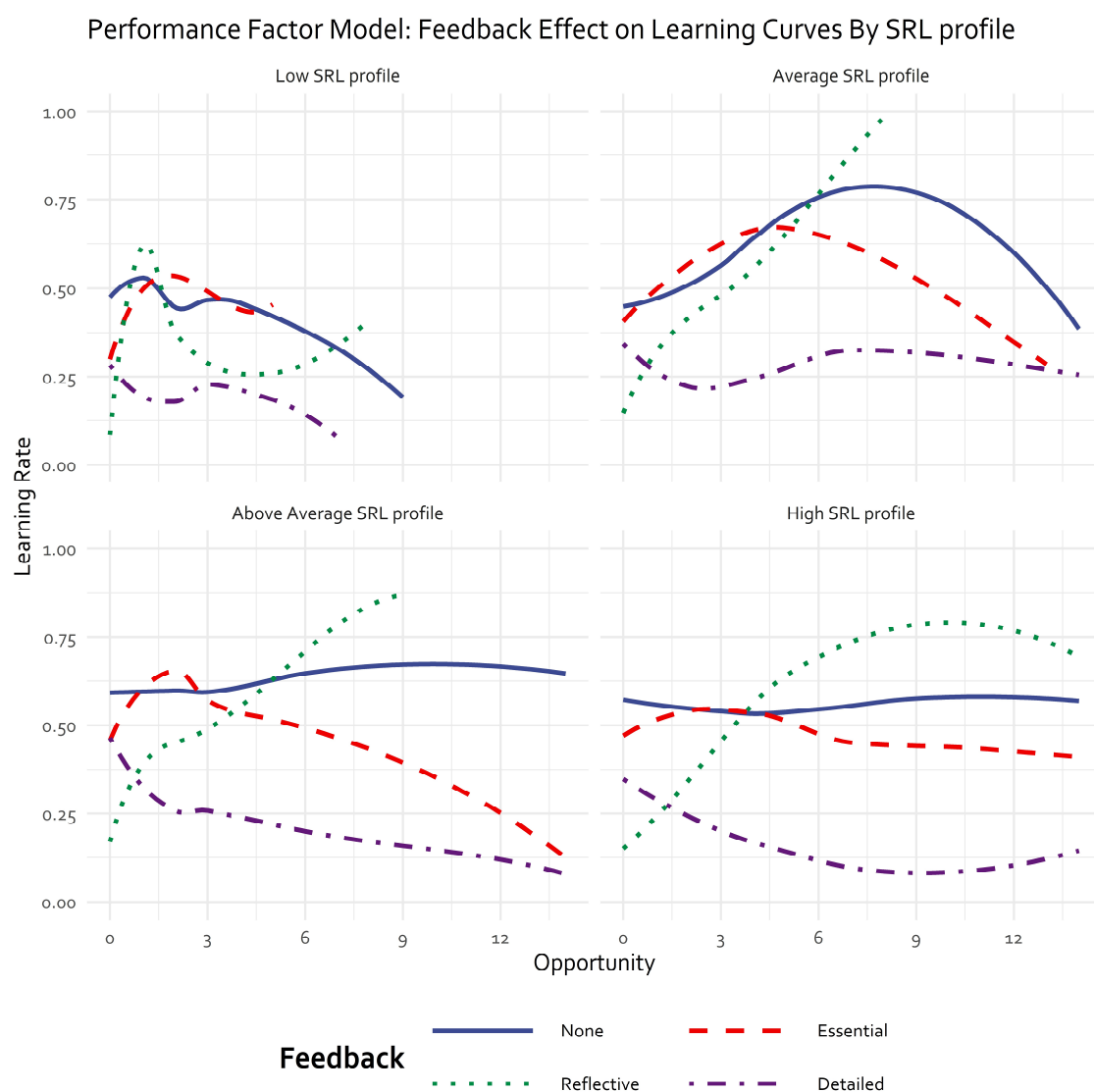


Figure 7.3.2: PFM Learning curves by SRL profile

However, the opposite is true for those with *Low* or *Average* self-regulation, who had higher learning rates given learning opportunities compared to those in *High* SRL profile. Reflective feedback (and by

extension the mechanism that generated it) produced the fastest increase in the healthcare provider's learning rate over increasing learning opportunities in all SRL learning profiles except the low SRL profile (Figure 7.3.2). The poor and erratic learning rates from healthcare providers in low SRL profiles are consistent with the findings in Chapter 6.3.4. Consultants/specialists -who are the highest skilled healthcare providers- and years of experience were significantly associated with the odds of being in the *Low* SRL profile despite demonstrating higher learning gains in fewer repeated learning cycles (Table 7.3.2, Figure 6.3.4 in Chapter 6.3.4). The difference in the learning curves between Figure 6.3.4 and Figure 7.3.2 is due to two factors: (1) the amount of student-step learning data used for Figure 7.3.2 is more and includes incomplete sessions, and (2) the unit of learning analysis has changed from a learning session (Figure 6.3.4) to student-step data i.e. learning task (Figure 7.3.2). Considering the SRL findings reported in Chapter 6.3.4 and 6.4, It is plausible that adapting the complexity of the learning content might be one way to keep healthcare providers in the low SRL profile engaged for longer.

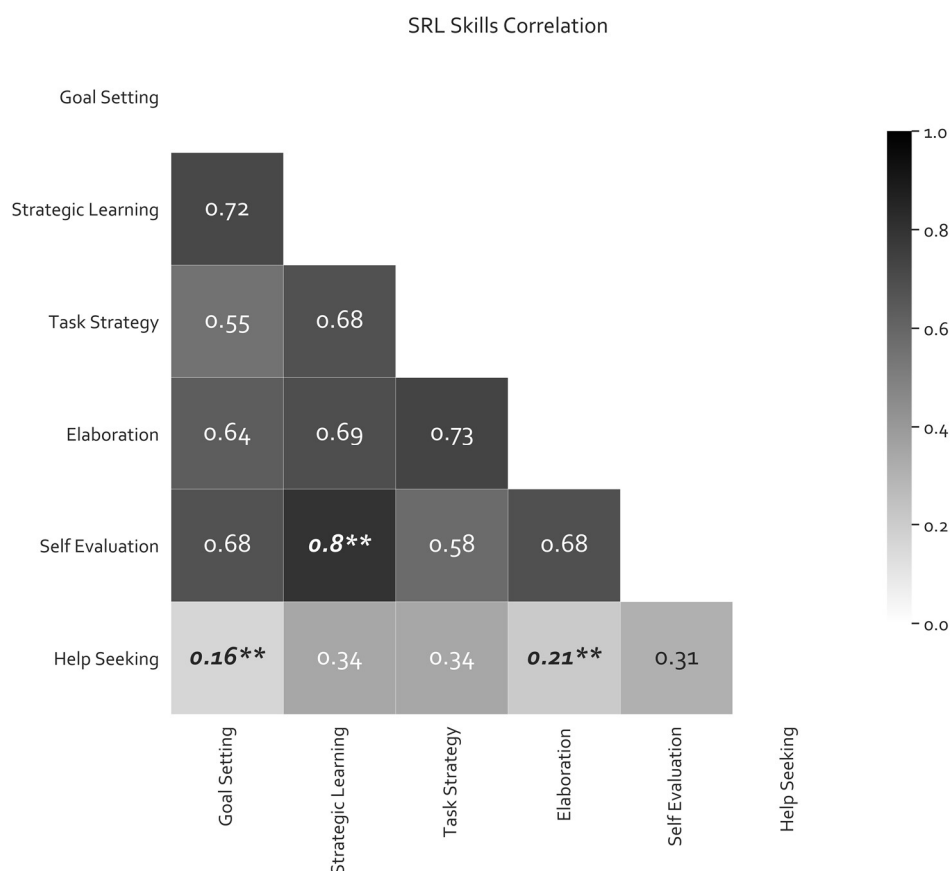


Figure 7.3.3: Pairwise correlation between SRL skills. (**) represents statistical significance. SRL skills are defined in detail in Chapter 3.3.1

From Figure 7.3.3, allowing healthcare providers using digital platforms like LIFE to frame and actively monitor their own learning objectives in order to gauge the required effort needed to realise them might possibly improve their SRL engagement. Findings from research question 3 ([Chapters 3.9](#), and [9](#)) generate understanding on how LICs healthcare providers perceive their SRL to be shaped by the nature of digital learning interventions like LIFE and their strategic learning needs.

7.3.4 Model diagnostics

The PFM models in Table 7.3.2 indicate that student factors account for 18.79% of the random variation in the PFM model A. This variation reduces to 12.19% when healthcare provider's SRL profile interaction with opportunities, time on task and spacing is adjusted for (Table 7.3.2, model B). How closely healthcare providers resemble each other regarding learning behaviours as indicated by PFM model reduces by 6.6% when their SRL is considered.

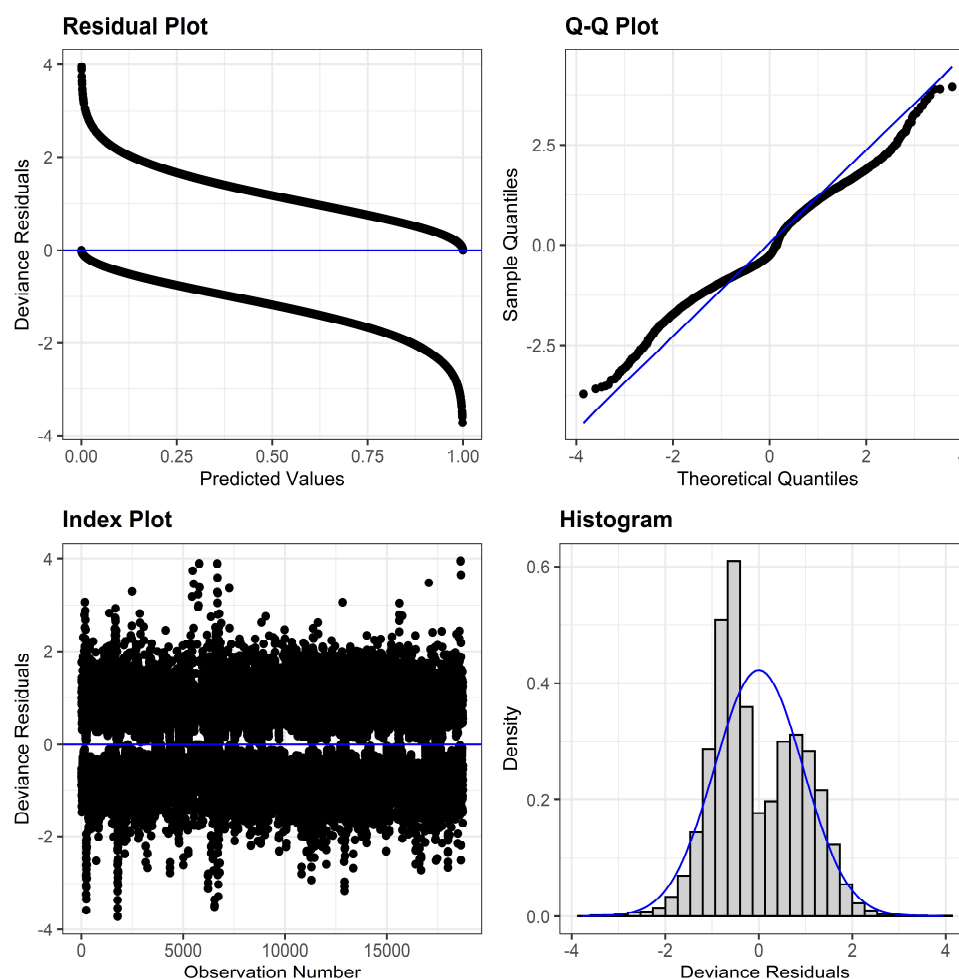


Figure 7.3.4: Model diagnostic plots for the PFM model

Further examination of the PFM model residuals was judged -with a high degree of certainty- to support the assumption of normally distributed dependent variable with a constant residual variance (Figure 7.3.4). Inclusion of these SRL interaction effects also reduced the deviance of the model A by around 50% as indicated by the decrease of Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Log Likelihood scores as illustrated by model B (Table 7.3.2), which indicates a better model fit. The proportion of variance in the PFM model that is explained by the fixed effects (i.e. marginal R^2) is roughly 53%. The fixed effects would be from all variables in Table 7.3.1 in addition to the learning tasks (i.e. KCs) difficulty ratings. At least 58% of the variance in the outcome (correct on first try in current learning session) is explained by both the fixed and the random effects. The student variable was implemented as having a random effect since student-step learning data is clustered around an individual learner.

7.4 Summary of findings

After adjusting for the moderating effect of the healthcare provider's self-regulated learning on both time on task, and previous learning opportunities, the odds of getting the next attempt correct on first try was significantly higher (OR: 4.13, 95% CI: 2.99–5.70, p-value:<0.001) when the learner received reflective feedback through an adaptive mechanism compared to when they received detailed feedback through a standard mechanism. When adjusting for the moderating effect of previous opportunities, the odds of getting the next learning attempt correct on first try after receiving detailed feedback was not significantly different from that of receiving no feedback or just minimal feedback through a standard mechanism. The feedback mechanism being referenced here has been explained in detail in Chapters 3.5 and 5.2.4-5.

The findings in this chapter suggest that the timing, specificity and frequency of reflective feedback, as determined by the adaptive mechanism, enhanced learning given the evidence that it moved the learner to a deeper understanding of the learning task (Greenwood, 1995; Lam et al., 2011; Rn, 2000). More feedback within LIFE did not always equate to better learning outcomes in future attempts, most

likely due to the limited capacity of cognition to make meaning during information overload (Brünken et al., 2010; Fournier, 1996; Lam et al., 2011; Rn, 2000).

From the analysis in this chapter, after accounting for the combination of multiple KCs of the LIFE learning content, and adopting an analysis strategy that minimises subject-treatment interaction bias and differential time-effect bias (Scott B Morris & DeShon, 2002), the Cohen d equivalent of the reflective feedback effect size is 0.782 (95% 0.604-0.960, p-value: <0.001) (Lenhard & Lenhard, 2016). As the resolution of the analysis unit gets more individualised, the effect of reflective feedback becomes more pronounced across the self-directed learning spacing options. This is illustrated in Table 7.4.1.

Table 7.4.1: Adaptive feedback effect on learning gain comparison				
Learning Spacing \ Unit of analysis	=< 1 hour	=< 1 day	=< 1 week	=< 1 month
Group level, Study arm (Chapter 5.2)	0.09			
Individual learner level (Chapter 5.3) [†]	0.644***	0.644***	0.364**	0.199***
Learning item level (Chapter 7.3) ^{††}	0.9155***	0.6474***	0.4872***	0.1294***
Note: *** : P value ≤ 0.001; ** : P value ≤ 0.01; Standardised Mean Difference (SMD) ≈ Common Language Effect Size. [†] SMD from standardised regression. ^{††} SMD = (log-odds * $\sqrt{3}$) / π				

8 LEARNING TO REPRESENT HEALTHCARE PROVIDERS KNOWLEDGE OF NEONATAL EMERGENCY CARE

8.1 Learning to represent healthcare providers knowledge of neonatal emergency care: Methodology

The aim of this chapter is to address research question four (i.e. *Given a healthcare provider's progress in the sequence of the emergency care tasks that they have prescribed, which knowledge tracing approach is better at predicting the outcome of the next caregiving task? How does this prediction vary by the healthcare provider's spacing of their own learning, experience and clinical role?*). Essentially, this chapter explores how well the common knowledge tracing approaches -which turn theoretical models to learning algorithms- embody the assumptions required for generalisation, together with comparison to novel approaches, as a preamble for providing more robust adaptive mechanisms for platforms such as LIFE.

8.1.1 Prior work on knowledge tracing

Given that LIFE is trying to capture aspects of a student's knowledge mastery over a sequence of conjunctive ordered tasks, the aim is to map and predict the learner input sequence $(x_1, x_2, \dots, x_T)$ to an output sequence $(y_1, y_2, \dots, y_T)$. That is, when a student attempts a learning task, a record is kept of whether the attempt was successful (y) linked to the learning task (x). Generally, the knowledge tracing task can be formalized as follows: given a learner's historical interactions $X = x_1, x_2, \dots, x_T$ up to time t on a Knowledge Component (KC), it predicts some aspects of their next interaction x_{t+1} (Corbett & Anderson, 1994).

Previously, Bayesian Knowledge Tracing (BKT) models have been used to model the knowledge states of KCs using Hidden Markov Models (HMM) (M. V. Yudelson, Koedinger, & Gordon, 2013). The hidden states in the HMMs represent the student's knowledge state which indicates whether they have mastered the KCs. However, some of the BKT modelling assumptions are considered impractical: BKT assumes that forgetting does not occur; the KCs are treated as being mutually independent; its typical implementation does not allow for learners to have different learning rates; it assumes that all students have the same probability of knowing a particular skill at their first opportunity (Pardos &

Heffernan, 2010); and it suffers from the problem of multiple global maxima when trying to estimate model parameters (Gong et al., 2011) and assumes that the latent states representing knowledge states are *discrete* and *independent* (Sapountzi A. et al., 2018).

To address some of the shortcomings of BKTs, Learning Factors Analysis (LFA) and their different modalities such as Performance Factors Models were proposed (PFMs) (Cen et al., 2006). They model student knowledge states using logistic regression models in order to deal with the multiple KCs issue while incorporating student ability into the model. They exploit the number of previous opportunities, -including the form of successes or failures- of a learner's attempt at a KC to predict whether the learner has acquired understanding about the KC. Although they can handle a learning task that is associated with multiple KCs, they are ill-equipped to deal with inherent dependency among KCs that are hierarchically related in the same learning content is varied (Pavlik Jr et al., 2009). Furthermore, PFMs tracking of prior successful or failed practices counts disregards the ordering of those practices (Gong et al., 2011) with AFMs assuming that all learners accumulate knowledge in the same manner while ignoring the outcome of their individual responses (Chi et al., 2011).

Recently, Recurrent Neural Network (RNN) models have been applied in an approach called Deep Knowledge Tracing (DKT) in efforts to address some of these shortcomings in knowledge tracing (Piech, Bassen, et al., 2015). These RNNs, have been found to robustly predict student performance on the next learning task, given prior performance (Khajah, Lindsey, & Mozer, 2016). I hypothesis that even in a “vanilla” form, RNNs are a suitable neural network architecture for the proposed analysis, given that they perform well on sequence modelling tasks in other domains (Khajah et al., 2016). To map the input $(x_1, x_2, \dots, x_T)$ to an output sequence $(y_1, y_2, \dots, y_T)$, the input vector undergoes a series of transformations via a hidden layer, which captures useful latent information in the form of a sequence of hidden states $(h_1, h_2, \dots, h_T)$. More concretely, at time-step t , the hidden state h_t is the encoding of the past information. Mathematically, this is represented as:

$$h_t = \tanh(W_{hx}x_t + W_{hh}h_{t-1} + b_h) \quad (8.1.1)$$

$$y_t = \sigma(W_{hy}h_t + b_y) \quad (8.1.2)$$

where both the hyperbolic tangent $\tanh(\dots)$ and the sigmoid function $\sigma(\dots)$ are applied in an element-wise manner. The model is parameterised by a weight matrix W and a bias vector b with appropriate dimensions. To control what information should be stored and used for predicting some aspects of the next interaction x_{t+1} , a popular variant of RNN known as Long Short-Term Memory (LSTM) RNN architecture is commonly adopted (Hochreiter & Schmidhuber, 1997). LSTM achieves this by incorporating three gates to mimic human memory: forget gate f_t , input gate i_t , and output gate o_t which control a memory state c_t . Mathematically, these are calculated based on the current input of x_t and the previous hidden state h_{t-1} as:

$$f_t = \tanh(W_f[x_t, h_{t-1}] + b_f) \quad (8.1.3)$$

$$i_t = \tanh(W_i[x_t, h_{t-1}] + b_i) \quad (8.1.4)$$

$$o_t = \tanh(W_o[x_t, h_{t-1}] + b_o) \quad (8.1.5)$$

Where [...] denotes concatenation. f_t decides what information to forget from the previous memory cell state c_{t-1} , while i_t decides what new information $\hat{c}_t$ is added to the recent cell state c_t . Therefore c_t depends on the previous cell state after forgetting with new information added from i_t . Eventually, the output gate o_t determines what information should be extracted from c_t to form the hidden state h_t . These can be expressed mathematically as follows:

$$\hat{c}_t = \tanh(W_c[x_t, h_{t-1}] + b_c) \quad (8.1.6)$$

$$c_t = f_t \otimes c_{t-1} + i_t \otimes \hat{c}_t \quad (8.1.7)$$

$$h_t = o_t \otimes \tanh(c_t) \quad (8.1.8)$$

Where $\otimes$ denotes elementwise multiplication. How LSTM utilises equations 8.1.1 – 8.1.9 is illustrated by Figure 8.1.1. This allows LSTMs to store information that occurred in the distant past, making it more robust in its capability to trace knowledge than a vanilla RNN. It also helps improve calculation

of gradients using the chain rule in RNN by overcoming the problem of vanishing gradients in backpropagation from using multiple RNN layers.

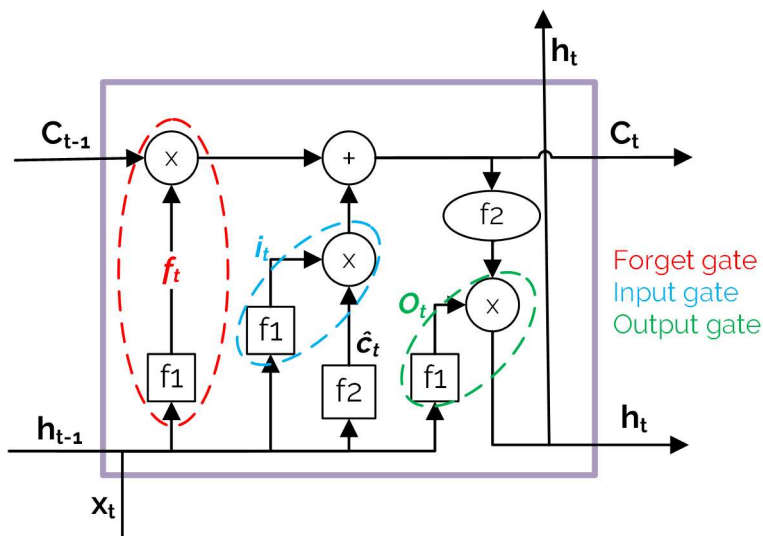


Figure 8.1.1: LSTM memory cell illustration

This is important because, it makes it possible to model “long-term” dependencies of the latent state being modelled i.e. LSTMS recall information for long periods of time as their default behaviour. The unfolded RNN explored for LIFE learning data to represent a high-level interpretation of DKT architecture is illustrated in Figure 8.1.2

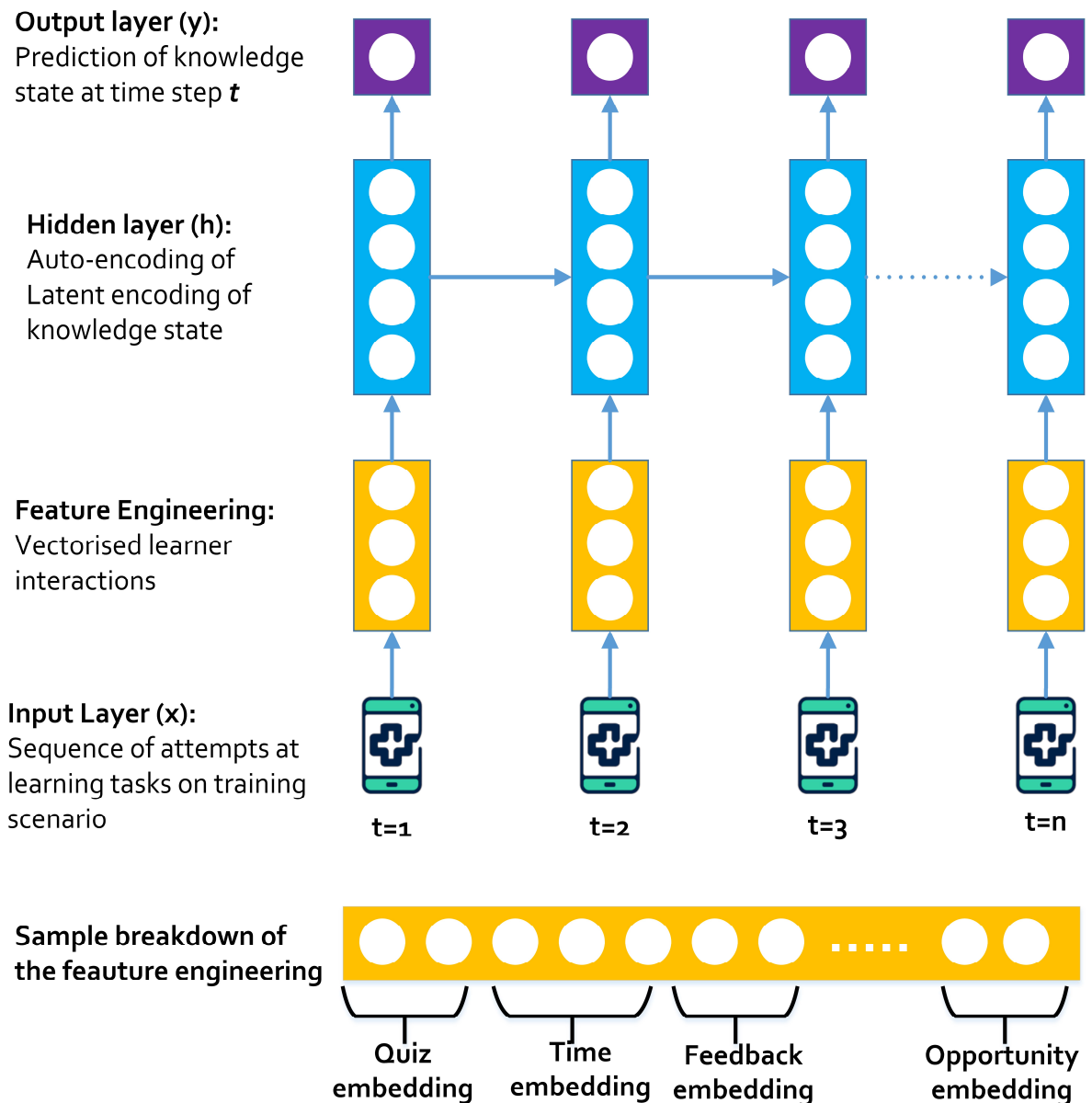


Figure 8.1.2: LIFE's proposed DKT model

For the DKT model proposed, autoencoding was implemented in the hidden layers (Figure 8.1.2). Autoencoder neural networks that use unsupervised learning technique for the task of **representation learning**. A DKT autoencoder model would seek to learn a compressed knowledge representation from learners' trajectories of learning task attempts. This is useful because when it creates a *bottleneck in the neural network (i.e. undercomplete)*, it forces the model to learn the most salient features of the original input features (student performance data). To avoid the model behaving like an identity function which duplicates the input features to the output features, regularisation is employed to force the DKT model to respond to unique statistical features of the input features.

Additionally, use of a loss function may be employed to encourage the model to have other properties besides the ability to replicate features i.e. insensitive to memorising input features but sensitive enough to build a reconstruction of data. Essentially, this usage of autoencoding achieves data denoising and dimensionality reduction. This is illustrated in Figure 8.1.3. If it was a linear network, (i.e. with no hidden layers, which have non-linear activation functions), this dimensionality reduction would be equivalent to Principal Components Analysis (PCA).

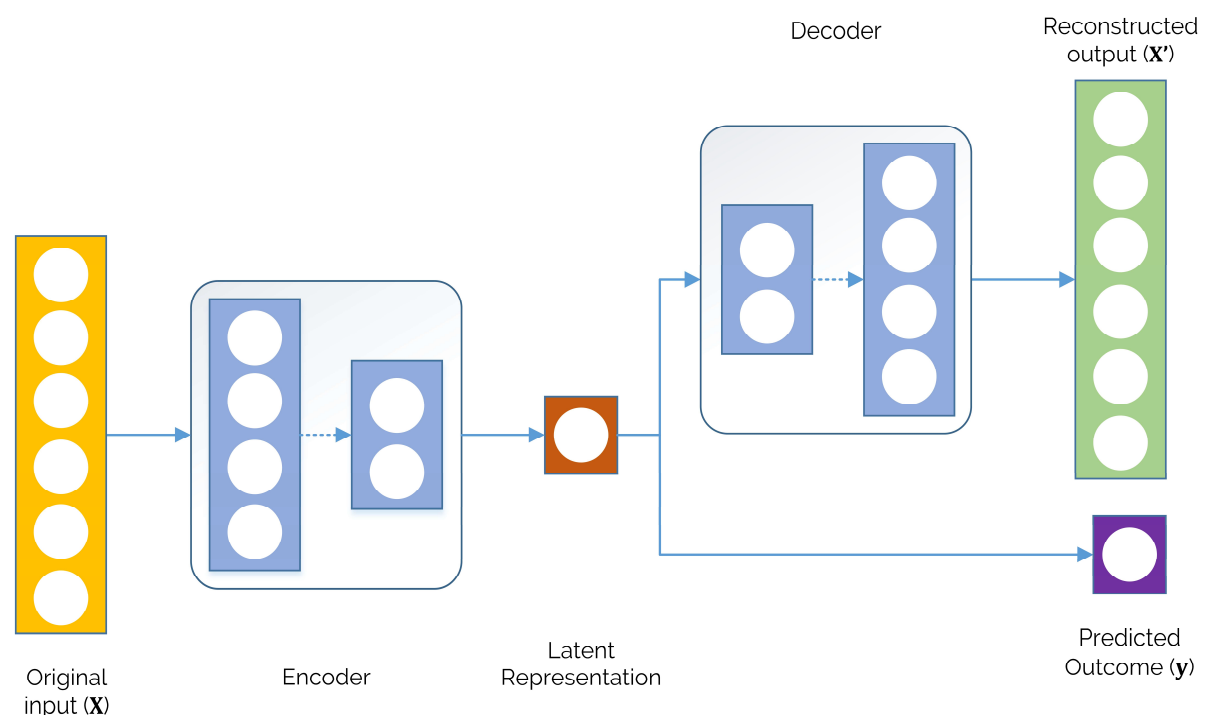


Figure 8.1.3: Architecture of the proposed Autoencoder

Using a DKT autoencoder is technically applying self-supervised model learning methods to student learning data to represent their knowledge. The best performing DKT model is posited to be a composite model that combines autoencoding of knowledge representation as continuous and high-dimensional together with prediction of students future performance (Sapountzi A. et al., 2018; Srivastava, Mansimov, & Salakhudinov, 2015) unlike a BKT model which discretises the latent knowledge state. Therefore, it does not assume that transition to the latent learning state is a discrete process like the BKT model.

LSTMs mimicking of memory is argued to better support knowledge tracing by accounting for learning ‘history’, based on the recency and outcome of completed learning tasks. This differs from BKT and LFA models’ “ballistic” approach where the use of the entire history of learner performance is enforced in the model fitting procedures (Deonovic et al., 2018).

There is scarce evidence of if and how the healthcare providers’ length of learning trajectories used in knowledge tracing affects the ability to predict some aspects of their next interaction for this conjunctive solution space of clinical training. Also, there is a scarcity of an empirical basis that explores prediction performance where explicitly addressing skill decay over time. The emphasis on skill decay over time rests in the risk implications for patient outcomes in this context. This study represents efforts to start addressing the aforementioned research gaps, in addition to evaluating whether this DKT’s design architecture, which is posited to better deal with the inherent dependency among KCs in a way that BKT and LFA models are unable to, outperforms the previously used knowledge tracing models.

The methodology in this section is aimed at addressing research question four and seeks to evaluate: (a) given a healthcare provider’s sequence thus far in the emergency care tasks they have prescribed, whether a DKT model can better predict the healthcare provider will successfully complete the next caregiving task compared to the BKT and PFMs, (b) how this prediction is affected by the healthcare providers’ spacing of their own learning, and (c) how this prediction varies by learners’ clinical role and experience. This research’s intended practical implication for the future is to advance evidence on how learning analytics can integrate a wholistic approach that maximises the impact of the formative learning evaluation on mobile platforms and increase both scale and access of adaptive learning interventions for clinical training across LICs contexts, especially those from under-represented contexts in learning analytics research e.g. Sub-Saharan Africa (Dawson et al., 2019).

8.1.2 Study Design, Setting And Participants

This study was an observational study of healthcare providers from both public and private hospitals in LICs, in clinical cadres such as nurses, clinical officers and medical doctors, with experience levels

varying from students to consultants. These are essentially the same participants as those that were recruited into the experimental evaluation of adaptive feedback (Chapter 5.2.3).

8.1.3 Study Variables

The learning scenario used in this study provides simulation training on the contextualised management of newborn resuscitation through a series of sixteen learning interactions that elicit responses from learners in the form of multiple-choice answers or performing interactive tasks. At the end of a successful completion of simulation tasks, the platform provides performance score feedback based on the learner's first attempt at each learning interaction. This is the same intervention used for the experimental evaluation of adaptive feedback (Chapter 5.2.4-5). For the purposes of the proposed analysis, the outcome of interest was specified as the next attempt at *any* learning task being correct on first try in current learning session, given the outcomes and features of the previous attempts *at the learning scenario* i.e. the assumption of inherent dependency among KCs is upheld but the prediction of student performance is not constrained to function within a single KC. The variables of interest were time spent on learning task, number of previous tries (i.e. opportunities) per learning task, and whether feedback had been provided for each unique try per learning task, and session-wise time spacing between learning sessions. Demographic data from study participants was optional and collected using an 'opt-in' mechanism in line with the approved ethics permissions. It included years of experience, clinical cadre (i.e. role) and the age-bracket of the study participants. Due to its limited availability (only 30% of participants had demographic data collected), to avoid biased modelling results, it would be used for describing model prediction patterns, but not included in the model fitting process.

8.1.4 Model Training

A student's trajectory consists of k learning task submissions, which represent how far they got in providing the emergency care needed (i.e. the number of attempts they made at learning tasks) within the LIFE game. Each task represents a question that they had to respond to. Additionally, time taken to complete the task, level of feedback provided on incorrect try, cumulative count of opportunities

at the specific learning task, time since last attempt, were captured with each submission. These data points were vectorised (vector representations of feature combinations from an individual learner's submissions) using within the LSTM model, similar to the one described in (Piech, Huang, et al., 2015) and illustrated in Figure 4.10.2. The input features were then temporalised i.e. if the model expected a sequence of length N , because each timestep t had its own input vector, and when fed to the model, it would be a three-dimensional matrix of size $(N_{samples}, N_{timestep}, N_{features})$. Predicted outcome (y') would be from the second timestep to the last i.e. a three-dimensional matrix of size $(N_{samples}, N_{timestep} - 1, 1)$.

To generate the predicted sequence of outcome for the next learning attempt, the temporalised input features are fed through the DKT neural network illustrated in Figure 4.10.5. The output y' from the prediction layer ("*pred_td*" layer in Figure 4.10.4) –has a sigmoid activation function and is an estimated probability distribution over two binary classes, indicating whether the healthcare provider would successfully solve the next task t given their past performance. Figure 8.1.4 is an elaboration of illustration provided by Figure 8.1.2 for the DKT model using four time-lagged student-steps to predict the outcome on the next learning task attempt.

Model: "Composite_AutoEncoder"

Layer (type)	Output Shape	Param #	Connected to
input_layer (InputLayer)	[(None, 4, 70)]	0	
in_layer_one (LSTM)	(None, 4, 32)	13184	input_layer[0][0]
encoder_layer (LSTM)	(None, 16)	3136	in_layer_one[0][0]
slice_repeat (RepeatVector)	(None, 4, 16)	0	encoder_layer[0][0]
out_layer_one (LSTM)	(None, 4, 32)	6272	slice_repeat[0][0]
strip_lagsteps (LSTM)	(None, 1)	72	slice_repeat[0][0]
reco_td (TimeDistributed)	(None, 4, 70)	2310	out_layer_one[0][0]
pred_td (Dense)	(None, 1)	2	strip_lagsteps[0][0]
Total params: 24,976			
Trainable params: 24,976			
Non-trainable params: 0			

Figure 8.1.4: Illustration of LSTM Autoencoder model structure with hidden layers structure

From the learning data, 30% was held-out as test dataset. Of the remaining 70%, in each of the 100 forward and backward passes over the training dataset (i.e. epochs), applying 10-fold cross-validation, the model would hold-out 10% of randomly selected samples from the training dataset as a validation dataset in each epoch. The use of multiple epochs aids in finding optimal model parameters which minimise the training and validation losses. I used Adam (Kingma & Ba, 2014) as the model's stochastic gradient optimizer together with a Rectified Linear Unit (ReLU) activation function (Nair & Hinton, 2010) for the deep layers of the model, with the Adam's learning rate set to 0.001.

To minimise overfitting and reduce generalisation error, I set the dropout rate of the hidden layer units during the transformation of both the hidden layer and the recurrent state to 20% in the deep layer (i.e. "in_layer_one" in Figure 4.10.4) and 50% in the prediction layer (i.e. "pred_lstm" in Figure 8.1.4) (Caruana, Lawrence, & Giles, 2001; Gal & Ghahramani, 2016). In this dropout regularisation mechanism, each hidden unit has a random probability of 20% (if in hidden layer) or 50% (if in the prediction layer) of being dropped, and thus the DKT neural network cannot rely on any single input node. Therefore, the neural network will be reluctant to give huge weights to certain input features,

because they might be dropped from the neural network. Instead the weights are spread across all features.

Furthermore, L2 kernel and recurrent regularization (i.e. ridge) was employed to smooth oscillations over training loss while also regularising autoencoding prediction, (Xu, Zhang, Wang, Chang, & Liang, 2010), with the a decay rate set to 0.001, thereby making the DKT model penalise large weights in order to minimize the model cost function. L2 regularisation was used instead of L1 (i.e. lasso) because unlike L1 regularisation, it provides a single and stable solution and its application in DKT was not to achieve feature selection. The detailed explanation of the L1 and L2 regularisation is provided elsewhere (Lo, 2013).

8.1.5 Model Evaluation

The Accuracy metric which calculated how often predictions matched labels in the validation dataset was employed as a model fitting optimising strategy. The models' performance was evaluated on the held-out test dataset (30% of the samples generated from sliding window approach which is explained in the next section), using a 10-fold cross validation strategy. More explanation of the cross-validation procedure is provided in Appendix G. Area Under a Curve (AUC)- also known as C-statistic - from Receiver Operating Characteristic curve (ROC) and Harmonic mean (i.e. the F1 score), Brier Score and accuracy were used as evaluation metrics. The Brier score provided a measure of the agreement between the observed binary outcome (correct on first try in current session) and the predicted probability of that outcome for the DKT model proposed and provided in comparison to the AFM and PFM model performance in Chapter 7.3. The Brier Score is a sum of both a calibration component and a discrimination component (Blattenberger & Lad, 1985; Brier, 1950), where lower scores indicate improved model accuracy. The Brier score is an improvement over the AUCROC, because it accounts for both the ordering and scale of predictions while the AUCROC, which evaluates how well a model discriminates, only accounts for the correct classification ordering (Assel, Sjoberg, & Vickers, 2017; N. R. Cook, 2008). Additionally, Brier score is a better metric when modelling an imbalanced outcome compared to AUCROC.

The modelling approach was designed to evaluate learner prediction performance based on temporal sequence of learners' input responses in LIFE. In efforts to train a model that uses the same dense function applied at every time (in LIFE's case, positional) step, a time-distributed layer was included, which allowed to make a prediction at every timestep t based on the hidden state at t . The performance would be evaluated using the already described metrics at every temporal slice provided in the data samples.

8.1.6 Statistical Methods, Missing Data, And Sensitivity Analyses

Data manipulation and statistical analyses were performed using *Python 3.7.3* and *Tensorflow 2.0.0* (Abadi et al., 2016; Van Rossum & Drake, 2011). A sliding window approach was used to determine the samples to use for modelling. In this approach, samples with varying time-lagged student-step sequence lengths, where the length in the lag varied from 3 to 10, were generated with each sample constrained to belong to a single learner. Given that the DKT model was specified to predict the outcome of the next attempt from current attempts learning features, predictions would be of length timestep-1 and begin from timestep two onwards. This is illustrated in Figure 8.1.5 where the index notation in the bottom right of the boxes illustrate sample counter and the number of circles illustrating single x input at timestep t . In Figure 8.1.5, the sliding window uses three student-steps, each column represents a student-step, and each row a concatenation of one-hot encoded input features for an individual student-step.

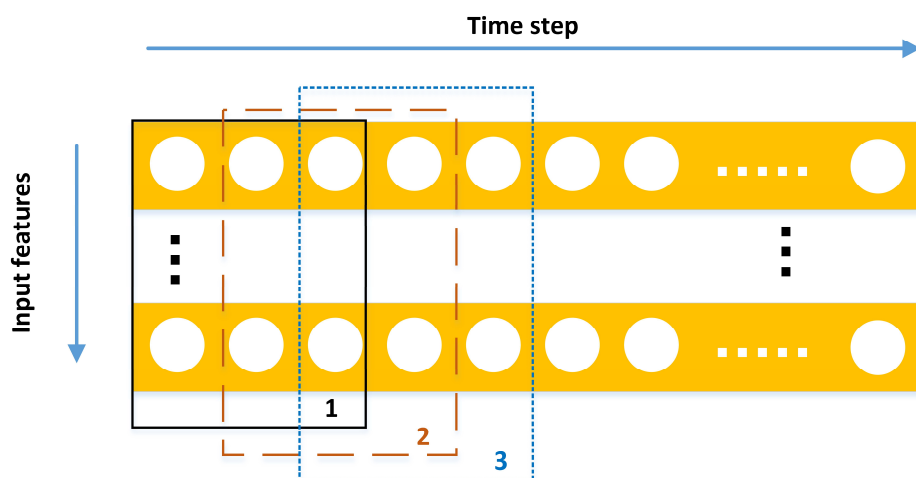


Figure 8.1.5: Sliding window sample generation based on varying length of timesteps for a single learner.

Varying the lengths of the timesteps in the samples was done to evaluate the sensitivity of the model performance in predicting y at timestep t based on how far back the history goes. Hence, the sensitivity of the predictions was evaluated against the length of the lagged student-steps input features used to predict the next step as illustrated in Figure 8.1.6.

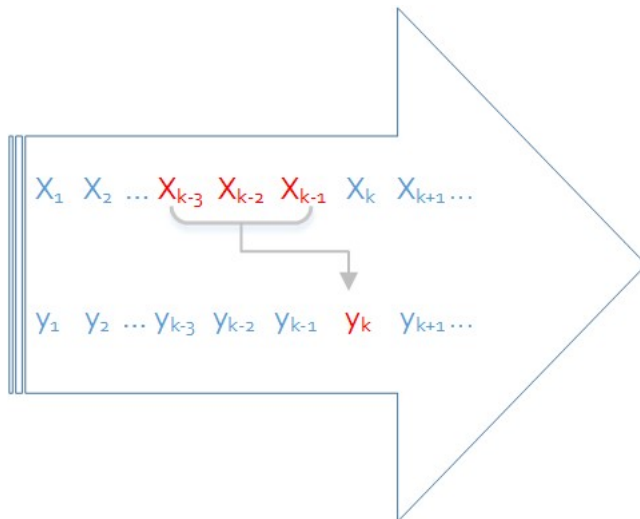


Figure 8.1.6: The DKT prediction setup for LIFE

As part of post-hoc analysis, a popular method for exploring high-dimensional data known as t-SNE was used to explore how well the latent encoding had been captured (Maaten & Hinton, 2008). The t-SNE parameters used were: learning rate=200, iterations = 5000, initialisation of embedding = PCA, and perplexity =10. The use of these metrics is explained in detail elsewhere (Maaten & Hinton, 2008; Wattenberg, Martin and Viégas, Fernanda and Johnson, & Ian, 2016). In the modelling of the samples generated from learning data using the sliding window approach, learners without enough data points were omitted from analysis as they would not have enough data for knowledge tracing. The ratio of omitted learners will be reported with findings and implications discussed. No evaluation was conducted of whether excluding observations with insufficient data would bias the results, and the limitations of this approach will be addressed in the discussions chapter. The DKT Autoencoder model performance was also compared to modelling approaches referenced in Chapters 7.1 – 7.3, with initial discussions provided of the implications for their future use.

8.1.7 Summary

The exploration and comparison of knowledge tracing models was continuous throughout the period of this thesis as learning data collection through LIFE was ongoing. This exploration was primarily done in effort to address shortcomings of earlier-used modelling approaches, but also to generate context relevant methodological contributions to the field of learning analytics in the area of knowledge tracing. It is important to note that as it pertains to knowledge tracing: (1) The latent construct being measured (i.e. knowledge) is not the product only of what LIFE was measuring, and (2) it is challenging to isolate how much of what you are looking at is constitutive of the latent construct in light of other factors that were not included. These challenges are addressed by findings reported in Chapter 9.

8.2 Results

8.2.1 Baseline characteristics of this study participants and variables

The data reported in this Chapter was observed between 1st February 2019 and 31st July 2019. This data is from the same healthcare providers as those reported in chapters 5.3.1. In brief, learning data from 572 healthcare providers from LICs who used LIFE game during this period is used. Table 5.3.1 in Chapters 5.3.1 provides a summary of variables of interest from all healthcare providers who used LIFE for learning, broken down into those who had a complete learning session and those who did not complete a learning session. In the DKT modelling using the samples generated from LIFE learning data using the sliding window approach, learners without enough data points were omitted from analysis. Table 8.2.1 gives a summary of the numbers omitted for each length of timestep used in DKT models.

Table 8.2.1: Learners excluded due to insufficient timestep data needed for the DKT model	
Length of sequence (timesteps) used	Learners excluded, N (%)
3	43 (7.52)
4	64 (11.19)
5	82 (14.34)
6	90 (15.73)
7	98 (17.13)
8	105 (18.36)
9	110 (19.23)
10	115 (20.1)

No evaluation was conducted of whether excluding observations with insufficient data would bias the results because there is no clear methodology to do so. Additionally, demographic data was not included in the model input features because the high level of missingness (i.e. only 30.59% of the participants provided this information) would affect model performance and bias the interpretation of findings. Instead, this information was used to explore patterns of learner performance prediction where it was available.

8.2.2 Summary of DKT model performance

Results from the running of the multiple LSTM-based DKT models demonstrated that using a longer time-lagged history of learner performance did not necessarily produce higher model accuracy, even though there was a consistent increase up to the fifth learning task attempt in the sequence⁴ (Figure 8.2.1). Overall, this modelling approach showed considerably better prediction performance compared to the alternatives (Michael Yudelson, 2019) when the length of sequence of predicted outcomes for the next learning task submissions was ≤ 5 , after which it started degrading.

⁴ The analysis code used for Chapter 8 can be found in the following GitHub Repository:
<https://github.com/AfrikanaCollective/Statistical-Programming-Projects/tree/master/Knowledge%20Tracing/DKT>

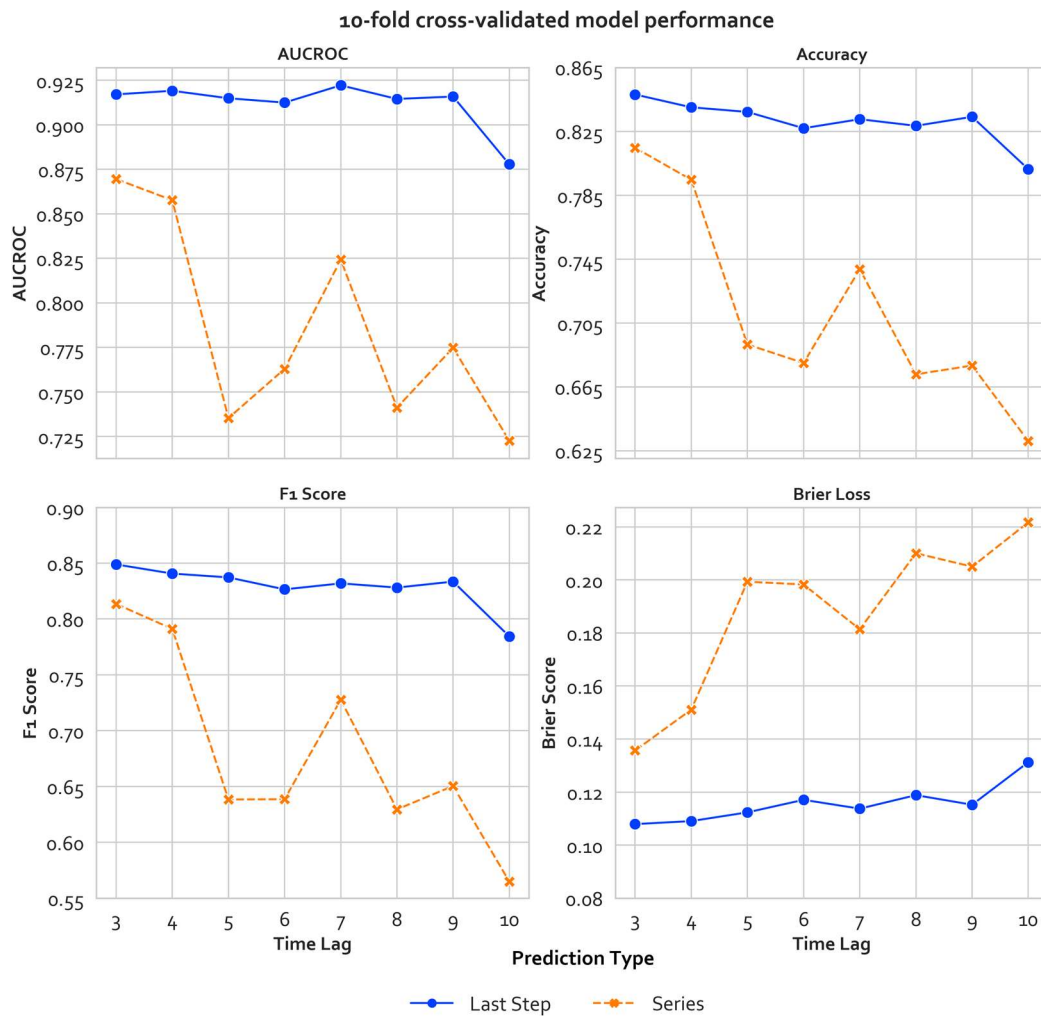


Figure 8.2.1: DKT model performance over length of sequence

Given that the binary outcome being modelled (i.e. correct on first try in current learning session) was imbalanced class distribution (7905, 42.18%), AUCROC curves were likely to give an overly optimistic indication of the DKT model performance. This is because ROC curves fundamental classifier performance does not change with the changes in class distribution. To overcome this challenge, Figure 8.2.1 is presented together with the Brier score and the F1 score (i.e. the harmonic mean accounting for both precision and recall of the model predictions), to help present an accurate prediction of future classification performance which instead evaluates the fraction of true positives among positive predictions.

This DKT approach accommodated for combining multiple ordered conjunctive learning task submissions which had unevenly spaced learning sessions. It also accommodated for combining learning task submissions within the input feature sequence which had drop-out and restarting.

Therefore, considering (a) the trade-off between the length of history of learner attempts the DKT model is ‘aware’ of, and (b) model performance, it would appear that for the LIFE content, with an Accuracy and AUC at least 82.5% and 90% respectively, using four lagged learning task submissions is robust and comparatively better in predicting the outcome of the next learner interaction compared to alternative learning models (Michael Yudelson, 2019). This predictive accuracy and knowledge representation are illustrated in Figure 8.2.2 below where how the distinctiveness of the latent factor captures the differences in, for example feedback provided in the next learning attempt step or outcome of that learning attempt, is evident.

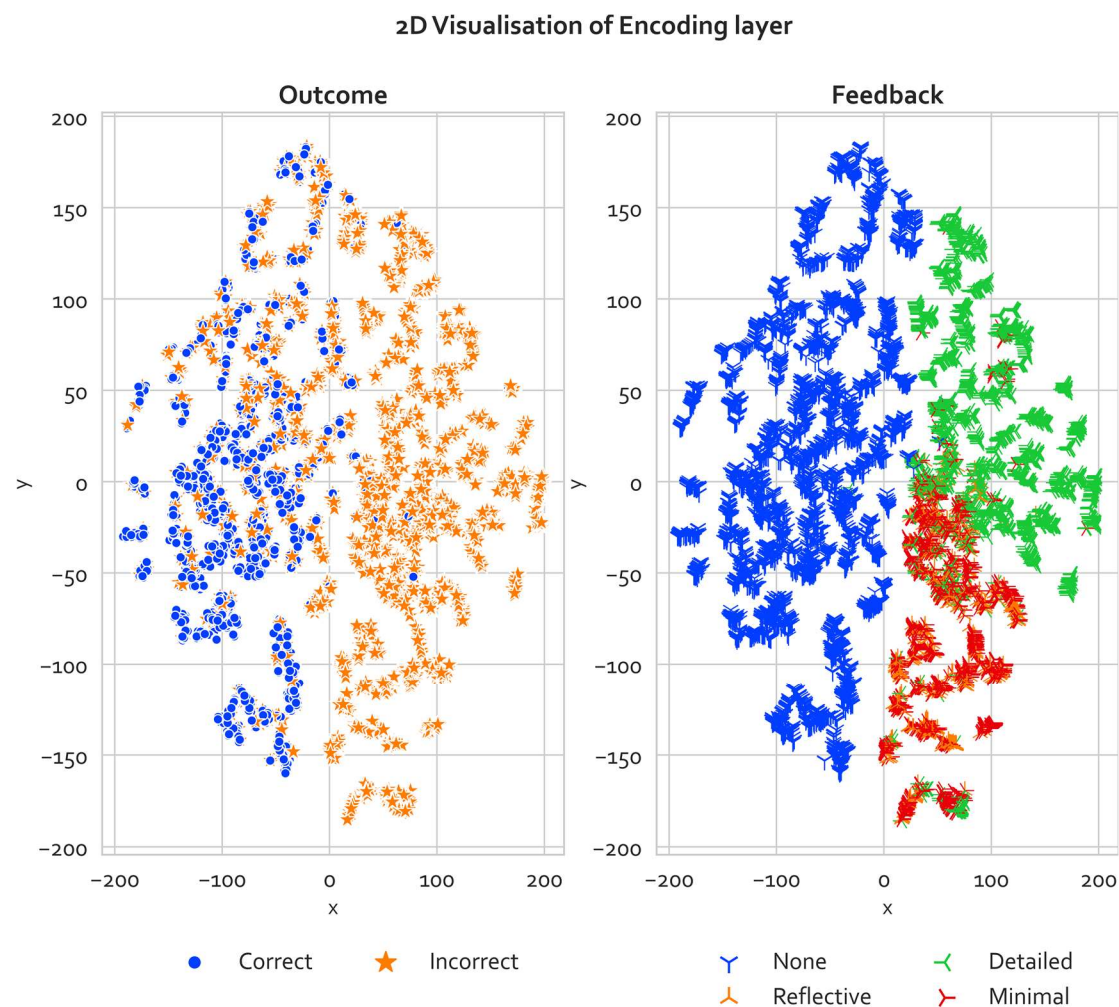


Figure 8.2.2: Visualisation of knowledge representation in the encoding layer of the DKT Auto-Encoder model

Within the Principal Components Analysis (PCA) representation of the encoding layer, in each learning attempt, Figure 8.2.2 illustrates how distinctively the DKT model differentiates the latent hidden nodes based on outcome for the *next* learning attempt, and the feedback to be provided.

Figure 8.2.2 is an example illumination on whether this DKT configuration can robustly represent, distinguish and latent representations from the ordered student-step sequences. It does not indicate what the patterns are: The researcher can subjectively infer it. While the DKT model's latent construct assumes "knowledge" to be what is being modelled, knowledge of neonatal resuscitation is not the product only of the measured input features (i.e. student step data). Therefore, it is challenging to isolate how much of what you are looking at is the encapsulation of "knowledge" versus other factors that were not included. Findings from Chapter 9 further elaborate on the conceptualisation of healthcare providers' perceptions of "knowledge" from smartphone-based clinical care training in LICs.

8.2.3 DKT model accuracy by learning spacing, healthcare provider features and overall performance

Consideration was given to how varied spacing of learning sessions would affect model performance given the current ongoing performance of the learner. From the DKT model using five timesteps (which was deemed to be the optimal length), there is consistent predictive performance of the DKT model across the different levels of the running performance of the healthcare provider (Figure 8.2.3). However, the model's prediction of next outcome at learning task for healthcare providers with a running performance score of between 50% to 80% has Brier scores that slightly higher than average, indicating poorer discrimination and calibration for these running learner scores. The concern here is that the less accurate DKT prediction performance for this learner performance range would bias any adaptive instructional intervention informed by the model for healthcare providers with scores that fall in that range.

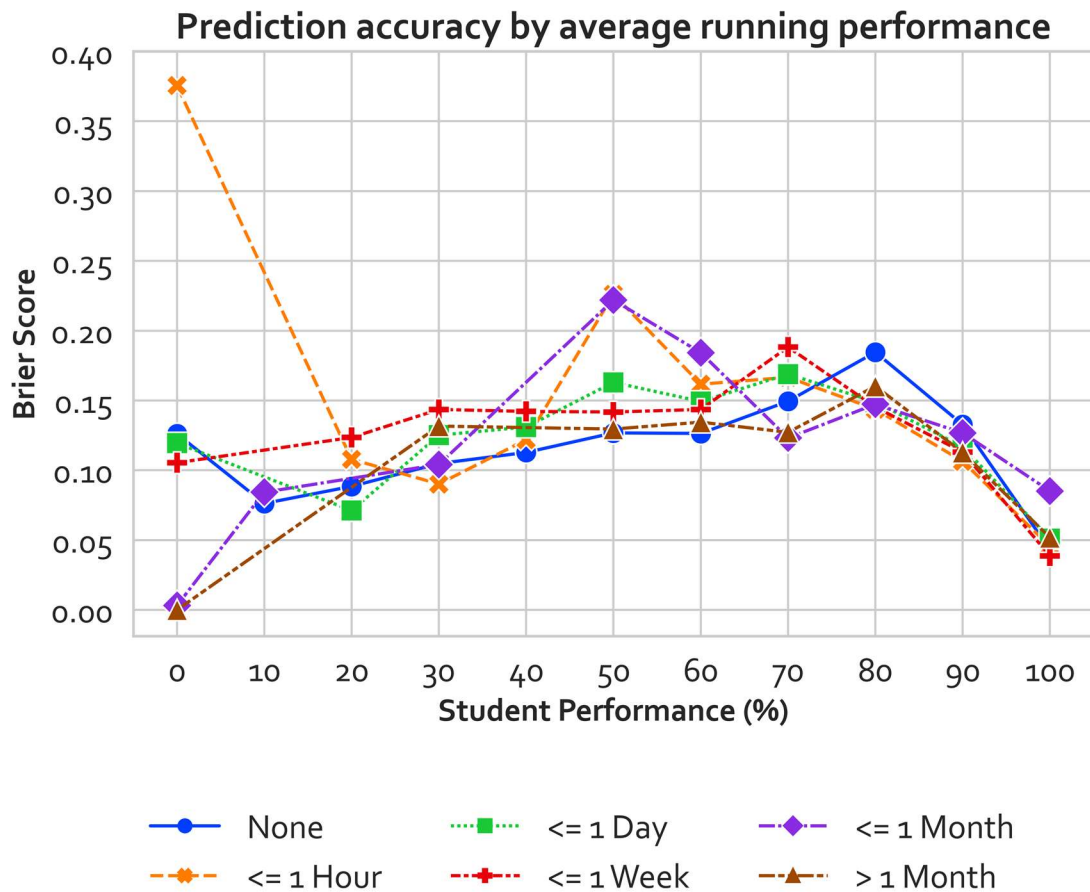


Figure 8.2.3: DKT model prediction accuracy across student performance and spacing

A plausible explanation for this poor DKT model performance within this score range is that the intersection of these categories has very few student-step data which reduces the level of accuracy in the inference from the DKT model (Appendix H, Table H.1). This in turn reduced the DKT model prediction accuracy when the learner's running average for these performance-defined sub-groups. However, the variance of the accuracy of learner performance prediction was neither affected by clinical cadre (i.e. role) nor the number of years of experience of the healthcare provider (Figure 8.2.4).

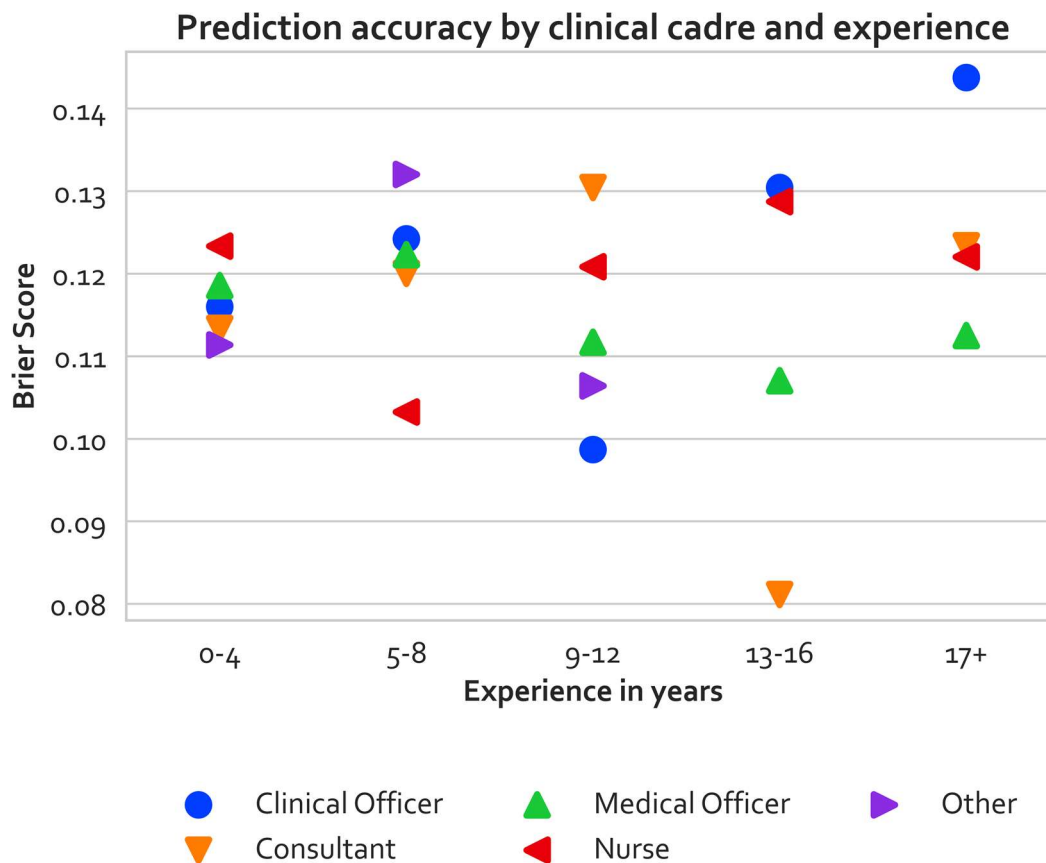


Figure 8.2.4: Prediction accuracy of performance by clinical role and years of experience

The only aberrant DKT performance was for Clinical Officers with more 16 years of experience, which is most likely due to lack of enough study participants in that group for the DKT model to learn from. Apart from that group, DKT model’s Brier score did not rise below 0.135, indicating that it is still highly accurate in predicting healthcare providers’ learning rates compared to the alternative approaches already highlighted in previous sections. This level of high accuracy in prediction of the outcome of the next learning task is still sustained when exposed to the SRL profiles of the healthcare providers from LICs except for those in the average SRL profile who are actively practicing and undergoing training (Figure 8.2.5).

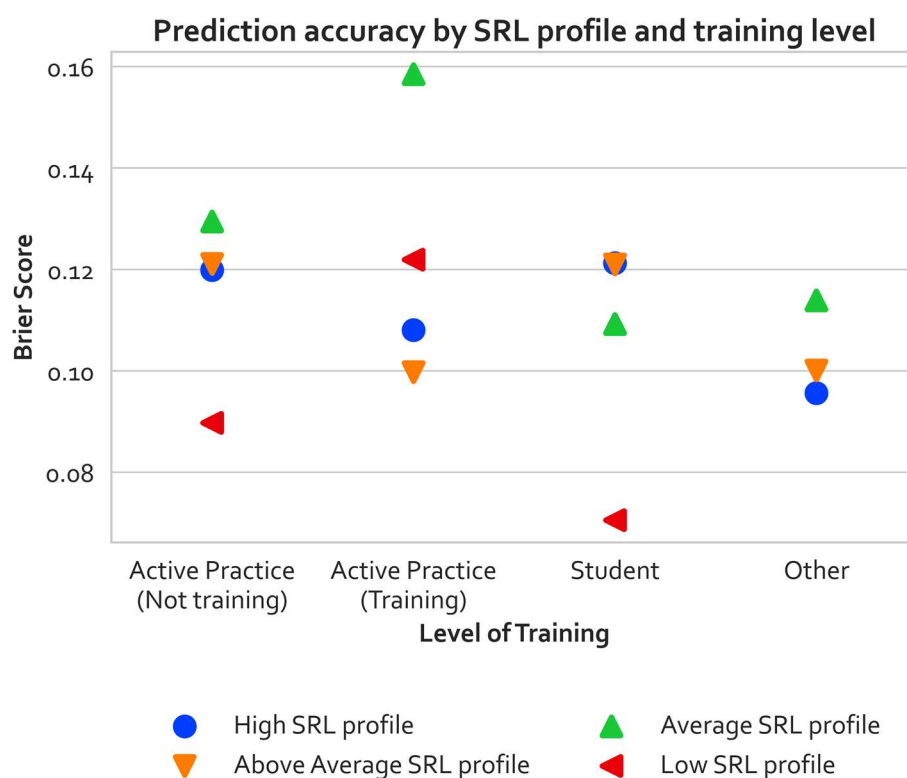


Figure 8.2.5: Prediction accuracy of the outcome of the next learning task based on SRL profile and level of training

8.2.4 Adaptive feedback's impact on knowledge mastery and wheel-spinning

Using the DKT model, mastery, calculated from successive attempts, was taken to be either where the predicted probability of mastery for all successive steps is ≥ 0.7 . This conceptualisation of mastery borrows from commonly used approaches for defining mastery (Beck & Gong, 2013) and conforms to the upper cut-off the ZPD used throughout this thesis (Chapter 5.2.5.4). The predicted probability is for the primary outcome (correct on first try in current learning session). Findings indicate that nurses and atypical healthcare providers (i.e. others) who received adaptive feedback had more streaks of knowledge mastery than those in the control arm. In all the other clinical carders, healthcare providers who received standardised feedback had more mastery streaks than those receiving adaptive feedback, with this difference being maximal in the specialist doctors (consultants). This is illustrated in Figure 8.2.6.

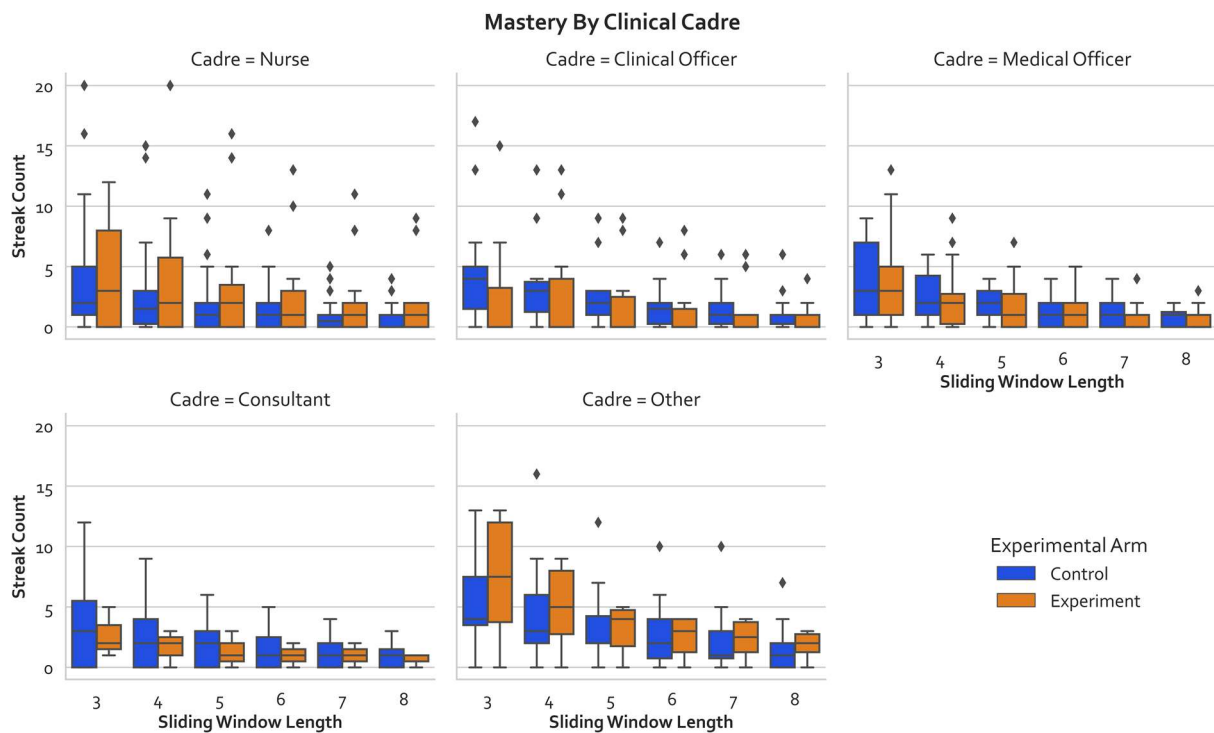


Figure 8.2.6: Knowledge mastery by healthcare provider role

The relationship of mastery to wheel spinning (i.e. defined as the learner spending considerable amount of time or effort struggling to learn a concept without gaining any mastery), was fascinating in that, healthcare providers most likely to have more counts of mastery streak had equally higher counts of wheel spinning (Figure 8.2.7).

Wheel spinning appeared to be highly correlated with higher counts of knowledge mastery streaks among the healthcare providers. Unexpectedly, for consultant physicians who had the least number of streaks, those who received adaptive feedback demonstrated almost no decay in their knowledge (Figure 8.2.8). This coincided with them also having an equally low amount of wheel spinning (Figure 8.2.7). Knowledge decay was conceptualised as the decline of the predicted probability of correct on first attempt on learning task given the time that had passed (Figure 8.2.8).

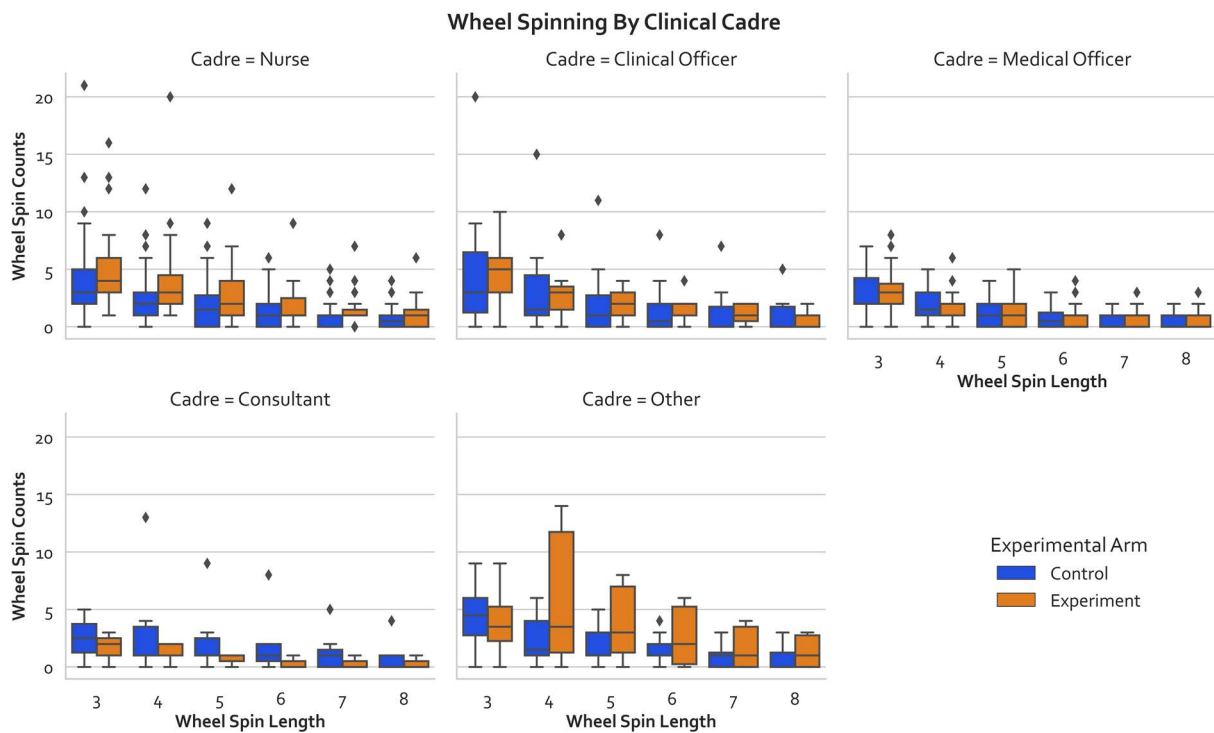


Figure 8.2.7: Wheel spinning by healthcare provider role

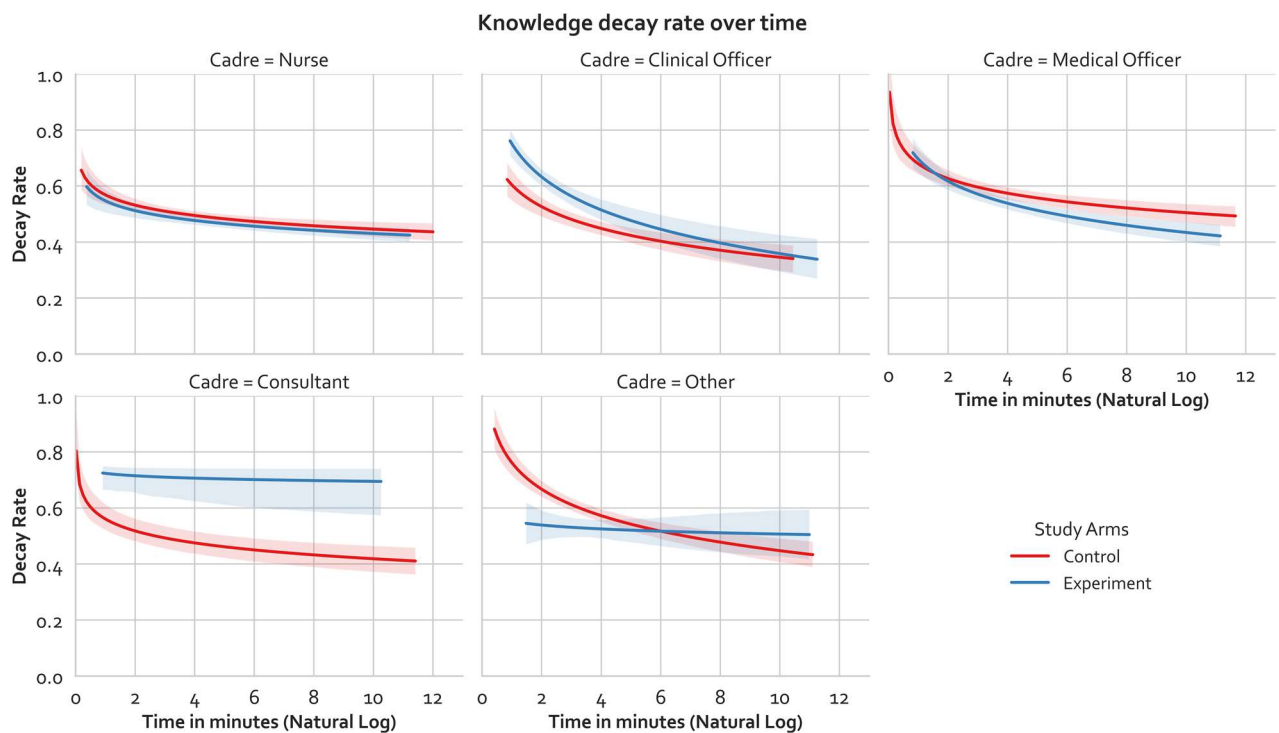


Figure 8.2.8: Knowledge decay rate by healthcare provider role

Further analysis was conducted to evaluate whether mastery streaks, wheel spinning, and dropouts covaried with the adaptive feedback provided and by clinical cadre. Applying Multiple Analysis of

Covariance i.e. MANCOVA, I explored whether the patterns demonstrated by Figures 8.2.6 – 8.2.8 were statistically significant. From fitting the MANCOVA model, the test-statistics of model assumptions (Shapiro-Wilks p-value < 0.001, Box's Test of Homogeneity of Covariance p-value < 0.001) were indicative that the experimental groups did not have equal variances, nor did they covary homogeneously. The experimental study arms, cadre and experience of the healthcare provider produced significant differences in the number of streaks, wheel spinning and dropouts. However, these findings, which are detailed in the Table H.2 and Table H.3 in Appendix H, should be interpreted cautiously due to the partial availability of demographic data and the violation of the MANCOVA model assumption. The MANCOVA model assumption violations might be indicative that the healthcare providers variances of the streaks, dropouts and wheel spinning for the adaptive feedback groups being compared are not equal. I.e. by participating in the study, the variances of the streaks, wheelspins and dropouts (which are artefacts of being in the study) were significantly modified to be unequal.

8.3 Comparison of knowledge tracing approaches so far

The challenges with the previous approaches to knowledge tracing have already been highlighted in Chapter 8.1.1: to recap, BKT modelling [impractically] assumes that forgetting does not occur, treats KCs as being mutually independent, assumes that all students have the same probability of knowing a particular skill at their first opportunity and suffers from the problem of multiple global maxima when trying to estimate model parameters (Gong et al., 2011; Pardos & Heffernan, 2010). On the other hand, although Additive and Performance Factors Models can handle a learning task that is associated with multiple KCs, they cannot deal with the inherent dependency among KCs (Pavlik Jr et al., 2009), their tracking of prior successful or failed practices counts disregards the ordering of those practices (Gong et al., 2011), or assume that all learners accumulate knowledge in the same manner while ignoring the outcome of their individual responses (Chi et al., 2011).

From fitting these models (including DKT) on LIFE student-step data, The DKT Auto-Encoder approach is comparatively a far better approach than the alternatives (Table 8.3.1) due to its better calibration in predicting the outcome of next learning task for healthcare providers from LICs using smartphones to learn.

Table 8.3.1: Model Metrics Comparing Different Knowledge Tracing Approaches					
Model	BKT	iBKT	AFM	PFM	DKT
Metric					
Accuracy	0.679	0.7642	0.689	0.767	0.837
AIC	22097.87	23366	9727.6	8235.2	N/A*
BIC	22411.41	41300	9994.7	8565.5	N/A*
AUCROC	<i>**Not provided</i>	<i>**Not provided</i>	0.748	0.85	0.915
Log likelihood	11008.94	9395.02	4825.8	4070.6	3353.32
Brier Score	<i>**Not provided</i>	<i>**Not provided</i>	0.3771	0.3099	0.1124
Note: <i>**The custom application for producing BKT/iBKT estimates does not provide these values.</i>					

The DKT model -treated as a mechanism for dynamic evaluation of how healthcare providers from LICs are learning through smartphone (e.g. formative assessment)- provides a more reliable avenue for instructional contingency (i.e. at what level will the next instructional intervention be), domain contingency (i.e. Given the content the learner selected, what support should be provided?) and temporal contingency (i.e. How long should the learner be left struggling?) in designing better learning experiences and interventions compared to alternatives. From Table 8.3.1, the extent to which the models accurately represent the learner’s knowledge influences the confidence with which targeted learning theories can be applied in instructional design to aid delivery of effective learning interventions.

A key concern in the models outlined in Table 8.3.1 is that with increasing model performance, there is also an increase in complexity in demonstrating how the model works, or, how to transfer the modelling approaches to other learning contexts. However, if the latent factor associated with “knowledge” is agreed to be complex and multi-dimensional (Cilliers, 2000), simpler knowledge

tracing models are more likely to harbor reductionism, which hinders the extent they can be useful for formative diagnosing of learning needs as illustrated by their poor calibration scores (Table 8.5.1). Another consideration is the implementation needs that make the choice of a knowledge tracing model to use more pragmatic; for this thesis, a huge consideration was the ability to function offline on low-end smartphone devices and integrated into other platforms i.e. Support for accessibility and scalability. At the time of running the experiment reported in Chapters 5, the standard BKT model lend itself well for this purpose. However, this consideration has been superseded by the recent release of machine learning platforms which allows for offline deployment of statistical models for real-time use on smartphone devices (Abadi et al., 2016; TensorFlow Team, 2019).

8.4 Summary of findings

Fundamentally, better content, presentation, and feedback structuring through use of adaptive learning pathways from a deep knowledge tracing approach is achievable with a high degree of accuracy for healthcare providers from LICs using smartphone to learn. This approach is robust to their years of experience, clinical roles, level of training, self-directed spacing of learning and student performance. The emphasis of these modelling approaches is not summative assessment, rather in understanding the how the learner is progressing, they create an avenue for intervening to ensure that the learner is consistently engaged in their learning, and increasingly being challenged. The DKT approach applied in this chapter only needs four sequential attempts on learning tasks to start making fairly accurate inferences about the knowledge of the healthcare provider.

However, caution must be given to the nature of data the models are trained on: class imbalances in the output feature being predicted, in combination with unrepresentative learner clusters (e.g. from LIFE's case, active healthcare workers who are in the average SRL profile and undergoing training, clinical officers with 17+ years of experience or, all healthcare providers with a running performance of between 50% and 30%) might negatively impact any gains to be made from adaptive learning design and implementation due to possible algorithmic bias. The idea of model cards for model reporting -

illustrating where the model is weakest in inference, and nature of data used to train the model- becomes a necessary feature to make sure those who implement such machine learning models understand where this bias is likely to be reproduced (Mitchell et al., 2019).

Additionally, it is quite clear that impractical assumptions made about how people learner are in turn associated with poorer model performance. As the statistical modelling approaches better reflect the learning behaviours of the intended learner, the better they perform in predicting learning outcomes.

9 EXPLORING HEALTHCARE PROVIDERS PERCEPTIONS OF GAMIFIED SELF-REGULATED LEARNING EXPERIENCES

9.1 Recap and Objectives

Although the instructional potential of game-based learning is recognised, there persists a lack of evidence to guide adaptive learning design, and support assessment (Baalsrud-Hauge et al., 2015). Little is known on what design and evaluation features of gamified digital learning might enhance SRL experience across different learning contexts (Sylvester Arnab et al., 2012; dit Dariel et al., 2013; Stuckless et al., 2014) more so for clinical training in LICs. Additionally, current frameworks that attempt to map pedagogical patterns to game-play features (e.g. badges, achievement and leader boards, difficulty level etc.) offer little evidence on how to evaluate adaptive game-play impact on the users' learning experience (Baalsrud-Hauge et al., 2015; Dawson et al., 2019).

The key challenge of how a gamified mobile training intervention might be better designed to support healthcare providers self-regulated learning of neonatal emergency-care management in LICs still remains, which is the purpose of this study. This is because research on gamified learning in contexts like those of SSA is in its early days and how to link this with SRL is even less well-developed. Additionally, while the effect of adaptive learning mechanism on learning outcomes has been evaluated (Chapter 5), and SRL behaviors of healthcare providers from LICs have been empirically determined (Chapter 6) and linked to learning gains (Chapter 7), how they perceive gamified learning on smartphone devices to influence their learning has not yet been elucidated upon in this thesis.

This chapter is aimed at addressing research question three i.e. its objectives are (1) *to explore the perceptions of healthcare providers on how to enhance the design and implementation of more engaging learning experiences on smartphone-based gamified emergency care training?* (2) *To explore how do healthcare providers from LICs perceive such contextualised learning experiences to influence their SRL efforts as they embark on bridging their knowledge gaps* (Mohan et al., 2016).

9.2 Methodology

9.2.1 Study design, setting, and participants and timelines

This qualitative study was conducted in three stages. The first stage used Nominal Group Technique (NGT) - a structured small-group consensus method (McMillan et al., 2016) which is illustrated further in Appendix I.1. I used NGT because it is well-suited for (1) building consensus among study participants for the purposes of idea-generation and (2) determining priorities of proposals for the design and implementation of more engaging learning experiences, (3) research questions geared towards informing learning experiences design (Baumann, 2018) and (4) aiding development of solutions that are testable using rapid-change cycles (e.g. regularly updated smartphone application content and user-interface design) (CHESS/NIATx, 2019). NGT was used to address the first research question whose aim was to build consensus with healthcare providers on how to enhance the design and implementation of more engaging gamified learning experiences.

The second stage was complimentary to the first and was carried out within a week of stage one and sought to evaluate whether the perceptions of healthcare providers were consistent across varied levels of seniority, training and experience. It consisted of brief semi-structured open-ended interviews and its purpose was to corroborate and validate the findings from stage one (Appendix I.2). Stage one and two of the qualitative study occurred in phase one of the thesis research (Chapter 4, Figure 4.1.1). The third stage consisted of semi-structured open-ended interviews (illustrated in Appendix I.3) was undertaken during phase two of this thesis research (Figure 4.1.1). Stage three's purpose was to evaluate how the contextualised smartphone learning experiences influenced healthcare providers' perceived SRL efforts. To overcome the challenge of having different types of participants in phase two from phase one, who might have very different learning needs, phase three participants was representative of participants in phase one and two. The research in this chapter was designed to produce insights that can be used to enhance SRL experiences design with deliberate repeated practice on smartphone-based gamified platforms; It was not designed to evaluate the

repeated use of the technology on learning outcomes, which we have reported in detail in chapters 5-7. A breakdown of the study participants in each stage is illustrated in Table 9.2.1.

Table 9.2.1: Summary of the Qualitative Research Stages			
Stage	One	Two	Three
Method	Nominal Group Technique	Semi-Structured Interviews	Semi-Structured Interviews
Time	Apr 2018	Apr 2018	1 st Feb 2019 – 31 st Jul 2019
Cadre	In-training paediatric nurses	Consultant and in-training paediatricians	Senior nurses, junior doctors, and in-training nurses
Location	Regional paediatric training centre (Nairobi)	Annual national paediatric conference (Mombasa)	Regional hospitals and medical training centres (Nairobi)
Level	Actively practicing	Actively practicing	Students, actively practicing
Recruitment	Medical training institution, local hospitals	Annual medical conferences	Professional paediatric social network managing the ETAT+ (Organization, 2005) course, peer referrals, regional clinical meetings, local hospitals
Date	April 2018	April 2018	February - July 2019

The selected participants varied by cadres and expertise to reflect different experiences at different levels of speciality training in neonatal emergency care (Table 9.2.1). In the first stage, participants were recruited from amongst practising nurses from different hospitals in the region undertaking advanced paediatric training at a private paediatric teaching hospital in Nairobi. In the second stage, all recruited participants were paediatricians drawn from different parts of the country attending the annual national Kenya Paediatric Association (KPA) conference. The third stage consisted of participants who were medical and nursing students and practising junior doctors and nurses. In this stage, these participants represented hospitals and medical training institutions in Nairobi County and its environs.

The participants in all the stages were selected using both convenience and purposive purposeful sampling techniques. They were identified to ensure a range of participants with varied training experience, covering both rural and urban areas, and a variety of the geographical location of clinical practice (i.e. rural and urban), hospital resource settings (i.e. private and public) and level of training. All study participants had completed at least one learning scenario in LIFE before being included into the qualitative study. Different sets of participants and methodological approaches was used to generate situated and contextually rich data about learning that would serve to corroborate, validate, contrast or expound on items generated from the NGT phase (stage one).

9.2.2 Data collection and analysis procedures

Stage one of the study – the NGT phase - took two hours. Using the neonatal resuscitation module within LIFE, the NGT's single purpose was to generate and prioritise ideas around how-to better design LIFE to support and motivate personalised learning. This would be derived from healthcare providers' insights about their learning experiences. The procedures adopted are outlined in detail in Appendix I.1. Data collection in the second phase was through semi-structured interviews, which were conducted with specialist paediatricians. These were analysed in juxtaposition to NGT analyses results, to highlight the varying views based on clinical cadre and experience. Each interview in stage two took between 20–25 minutes and was conducted in a designated space within the conference area. Data collection in the third stage was also through semi-structured interviews, each interview lasting 30-40 minutes. The NGT and interviews were all conducted in English.

I facilitated stage one NGT with the help of LF1⁵, and LF2⁶. Stages two and three interviews were conducted by me with LF2 aiding in participant recruitment. I digitally recorded all interviews, which were transcribed by a transcription service, but I checked all transcriptions for accuracy. During the

⁵ LIFE project team members

⁶ LIFE project team members

process of transcription, all participants were anonymised by assigning interviewee codes. The audio recordings and transcripts were stored in a password protected computer.

Because the individual in-depth interviews in stage three were being continuously interpreted together as the new interviews were occurring, questions that were not effective at eliciting perceptions and experiences related to clinical training were suppressed in favour of new ones. Thematic analysis (Braun & Clarke, 2006) informed by SRL theories (Winne H. P., 2011) and linked to the research aims was applied to the NGT and interview data after cross-checking and discussions with members of LIFE team.

SRL theory informed the analysis of the qualitative data but was not linked to the design of interview guides. SRL theory informed the analysis of how the immersive and interactive experiences that engaged the healthcare providers kept them motivated to continue learning. While I posit that LIFE provides SRL training opportunities, rather than explain the nature of SRL within the gamified platform, the analysis in this chapter focused on how healthcare providers perceptions of their learning experiences using LIFE aligned with SRL processes that have been explained in Chapter 3. This is because LIFE was designed to provide opportunities for autonomy and independence that can help set the stage for regulatory action and thinking. We did not explicitly ask about SRL experiences, so that the participants' level of knowledge of SRL concepts would be rendered inconsequential to the responses they provided.

9.2.3 Summary

Qualitatively evaluations of self-regulated learning experiences of LICs healthcare providers would help minimise bias in this thesis's discussion chapter from presuming meaning attached to the inferred benefit of learning approaches such as those used by LIFE. More importantly, it served to generate understanding of the ways learning analytics may be better leveraged in the field of clinical training on mobile devices for higher impact on patient outcomes in LICs (Dawson et al., 2019).

9.3 Results

9.3.1 Baseline characteristics of the qualitative study stages

For stage one, I targeted one cohort of nine study participants attending a continuous professional development nursing course. This had only female nurses. This gender disparity is not unusual as more than 75% of nurses in Kenya are female (Wakaba et al., 2014). Their median age and median experience in years was 34 and 6 respectively (Table 9.3.1).

In stage two, I interviewed five participants, all of whom were paediatricians or paediatricians-in-training, including three females and two males (see Table 9.3.1 for group characteristics). In stage three, I interviewed 19 participants, made up of junior doctors, in-charge paediatric nurses, and nursing students. The median age of participants in this stage was 25 years with three years of experience. Across all the stages of the qualitative study, participants were drawn from seventeen different health facilities' paediatric units and departments spanning private clinics to public referral hospitals, geographically representing different regions of Kenya and reflective of the various professional cadres.

Table 9.3.1: Summary characteristics of study participants					
Research Stage	Clinical Cadre	Institutions represented	N (% Female)	Age (years)	Experience* (years)
				Median (IQR)	Median (IQR)
One	Sub-speciality nurses in-training	8	9 (100%)	34.5 (30-38.75)	6 (3-8)
Two	paediatricians	4	5 (60%)	32 (31-33)	8 (7-8)
Three**	Senior nurses, junior doctors, and students	5	19 (52.6%)	25 (23.5-33.5)	3 (2-7.5)
Total		17	33 (66.7%)	31 (25-35.5)	5 (2-8)
Note: *Experience reflects cumulative years of front-line service, including internship. ** 4/19 of these participants were junior doctors.					

9.3.2 Results from Nominal Group Technique.

From the silent phase and item generation phase detailed of Nominal Group Technique (NGT) in Appendix I.1, the participants went through the listed items, discussing with and asking for clarification from each other, with the facilitator (TT) summarising the generated items as listed in Table 9.3.2. The participants tended to express their learning experiences using the app in a manner unlinked to

specific in-app feedback design instruments even though that was what I was asking them to do. An example of this is item C, G or I in Table 9.3.2, where stage one participants attached value to how LIFE influences their cognitive processes. Additionally, during NGT, I felt that the brevity in the descriptions of the items generated by the participants was purposely done to avoid in-depth interrogation by their peers.

Table 9.3.2: Summary of design components derived through consensus from item - generation phase of the Nominal Group Technique.	
Option	Description
A	Promote challenge using almost similar multiple-choice options
B	Accommodate differences in actual workplace practice in the scenarios
C	The timer adds pressure, anxiety, [<i>*viewed positively, to be retained as a feature</i>]
D	Give immediate feedback due to the "urgency" of the learning scenario [<i>*as opposed to providing delayed feedback at the end of the scenario</i>]
E	Improve the utility of feedback to help remember the correct answer [<i>*e.g. by adding more information links</i>]
F	Give immediate correct answer on first error screen but weight the final score when feedback message is provided to the wrong entry.
G	Sharpen <i>critical</i> thinking when playing the game [<i>*in reference to use of feedback messages in error dialogues, providing hints etc.</i>]
H	Add emojis to make enjoyable [<i>* e.g. When reflecting LIFE's evaluation of learner's progress</i>]
I	Challenging, if it's the learner's first time interacting with [<i>*ETAT+</i>] materials
J	Give a rationale for more <i>task</i> information. [<i>*Helps to build learner's appreciation of task importance</i>]
K	Add continuation to scenarios [<i>*quantity of training content</i>]
L	Cascade challenge and complexity level of the content
Note: <i>* Clarification added through discussions with study participants. **Two options (Give more demonstration of tasks; Reduce navigation difficulty;) were not discussed due to accidental omission by the session facilitator.</i>	

The identified thematic areas from the NGT discussions that generated Table 9.3.2 are expounded upon from the next section, and in juxtaposition to results from stage two and three analyses, for purposes of comparison and contrast. Participants in stage one ranked the items listed in Table 9.3.2 in order of decreasing priority, based on which items they were most inclined towards, impressed with, or would recommend being included (i.e. voting phase of NGT).

From the voting phase, each cell in Table 9.3.3 in rank columns represents the number of participants who voted for that item at that rank position. Given that there were 12 items to be voted upon, and participants were limited to only voting for their top 10 items, row-wise sums of the item votes could be less than or equal to the total number of participants. The weighted average rank, given as the sum of the ranks adjusted for the votes per rank, was calculated as illustrated in equation 9.3.1.

$$\text{Weighted Rank} = \frac{\sum(\text{row}(\text{cellwise votes} * \text{rank position}))}{\text{row}(\text{sum of votes})} \quad (9.3.1)$$

Table 9.3.3: Weighted rank of items generated from the voting phase of Nominal Group Technique												
Item	Item description*	Rank										Average Weighted Rank**
		1	2	3	4	5	6	7	8	9	10	
F	immediate corrective feedback on first incorrect try	1	1	1	2	1	0	1	0	0	0	3.71
J	Task rationale with more information	2	2	0	2	0	0	1	0	1	0	3.75
G	Sharpening critical thinking using feedback	2	1	0	2	1	0	1	1	0	0	4
I	Challenging Content	0	2	1	0	1	0	1	1	0	0	4.5
C	Task timing	1	1	2	0	0	1	1	2	0	0	4.75
D	Immediate feedback due to learning task urgency	2	1	0	1	1	0	2	1	1	0	4.89
E	Improved feedback utility	0	1	0	1	1	2	0	2	0	1	6.13
A	Promote challenge using almost similar multiple-choice options	1	0	0	0	1	1	1	0	2	0	6.17
K	Add continuation to scenarios	0	0	2	0	1	2	1	0	0	2	6.25
L	Cascade challenge and complexity level of the content	0	0	1	1	1	2	0	0	3	0	6.38
B	Scenarios accommodate differences in workplace practice	0	0	1	0	1	0	0	2	1	0	6.6
H	Add emojis to make enjoyable	0	0	1	0	0	1	0	0	1	5	8.5
Note: *The item description provided as shorthand of descriptions given in Table 9.2.1. **The average weighted rank is given as the sum of the ranks adjusted for the votes per rank, with the formula provided in equation 9.3.1												

The next sections of the results analyse responses from all the qualitative study stages through a thematic lens (Braun & Clarke, 2006; Winne H. P., 2011), reflecting how experiences within different cadres and levels of speciality training in neonatal emergency care might map onto personalised training using LIFE. The themes identified from semi-structured interviews with participants are discussed in the following sections and are linked to Winnie Hadwins' SRL model (Winne H. P., 2011).

9.3.3 Learning task definitions, elaboration and performance imperatives

Nursing and student participants interpretation of the aims of learning tasks within LIFE implied that from the game features provided within an assessed task, their attention was skewed towards performance (as indicated by their fixation with final score), rather than task elaboration. Specifically, time spent on a task appeared to be intrinsically associated with the urgency of the learning task, considered inseparable from performance goal. This correlates with the high rank item *F* (*Immediate corrective feedback*) in Table 9.3.2.

"...As you begin, it took a little bit of time because you had to read the explanations and the questions...So, at the end of it you find that your resuscitation took a bit of a long period. And even as you start the game, you know at the end you are being timed which again I think is a good thing because again with resuscitation, you are dealing with life and death and you have a very short period of time, so your interventions have to be really correct and they have to be timely..." LIFEEXP12

However, paediatrician participants in stage two tended towards downplaying time as a task strategy and instead, sought to use feedback to inspire goals linked to refining skill competencies. In stage two, timing was perceived to be relevant when conditioned on pre-existing learner knowledge or practice and is dealt with in a later section of this paper. This would be more reflective of the rank of item *I* in combination with item *C* (*Task timing*) in Table 9.3.2.

"...timing should not be [there], but when you have the pop-up message telling you 'this is a life-saving procedure, you need to move, you need to make a decision, you need to make a step' it's good enough..." KPA002

In contrast, perceptions of stage three participants combined the emphasis of elaboration of the learning task together with its time-linked consequences.

“...Ideally, as nurses, ...you need to know because if you mess, if you do not do the right things at the right time the first time, most of the times you will mess up with those children. So, it is good to understand, ‘why do I have to do this?’...” LIFEEXP06

Learning task elaboration (a “More information” button within the LIFE app) and as indicated by item *J* (*Task rationale*) in Table 9.3.2 was highly valued among the participants in stage one and three, hinged on the perception of LIFE not as a game first, but as a learning tool. This is illustrated in the quotes below.

“...They [feedback messages] are explaining why you should do that so I found it quite educative because there are times when you are not even sure, ‘why would I need to do this?’ but they are able to elaborate it further. If they tell you it is a neutral position, why should the head be in a neutral position? Then they explain to you why. ‘Why would I use this kind of mask rather than this other one?’, you know...” LIFEEXP06

However, stage two participants preferred a standardised learning process that was not seeking within-game tasks elaboration as the preferred emphasis of the game. Given that stage two participants were paediatricians, they would have been exposed to ETAT+ hence their agreement on this. They argued for elaboration of the learning goals to be articulated either before or after, but outside the LIFE game itself. This might be attributable to the differences in the level of training, prior experience and specialisation in the participants between the stages.

“...I think you can gauge how much people have internalised if you’ve given the talk and now use their scores and what mistakes they make to like correct things at the lab. So, I feel like it can still be used even without the ETAT setting⁷, if you were just given like a CME⁸ or something. Like do a talk and then tell them, ‘Okay fine I want you to play this game’, then give them their scores at the end, you go like, ‘I realised that this are some of the few things

⁷ ETAT Setting is a reference to ETAT+ trainings where practical assessment on a mannequin facilitated by qualified instructors is blended with class-based learning.

⁸ CME: Continuous Medical Education, A common routinely held hospital-based seminar series for healthcare providers. It is a brief session lasting around 20-60 minutes and usually delivered via power-point presentation.

that were hard to clear...’ I feel like it’s giving that talk before and giving them the goals of the game and what you expect them to achieve...” KPA003

Additionally, stage two participants appeared to be in agreement that an unintended consequence of tasks being elaborated in similar fashion to traditional educational tutorials diminishes the gaming experience, which I explain further in the *Games and engagement* results sub-section. Healthcare providers perceive limited internalisation of learning where there is no elaboration. The elaboration being referenced here not only emphasises the expected goal of the learning task being taught but also a nuanced breakdown of the consequences of [in]action or erroneous action.

9.3.4 Context of practice, cognitive adaptations and goal setting

Stage one and three participants’ perceptions on how LIFE might encourage personal evaluations were arguably a result of reflections on (1) context of resources and frontline work encountered so far, and (2) the cognitive effort put in weighing options available predicated on prior knowledge.

“...You know down there in the villages, things like a resuscitaire are not available and they actually have to do the resuscitation on a desk and not on a resuscitaire because the hospital doesn’t have a resuscitaire. Or even if they are available, they are not working. A normal resuscitaire should have a warmer at the bottom and the warmers are not even working in the hospital...” LIFEEXP03

“...Stimulate critical thinking...you think about resuscitation materials...I compare with what sometimes is practiced at place of work, like tickling the babies’ feet so I was like ‘I would like to tickle it also’...” NGTCONV

“...We have the scenario, like this is a new-born resuscitation, this is the flow chart, this triggers you, if at all you have done this before, it triggers you because you have the three Ss like shout for help, prepare your setting, wash your hands and all that then examine. So, when you are coming to the game itself like the scenario, when it comes to the choices you are like ‘this is off, this is off...’. But the way [LIFE] answers are, all the enticing ones are there: ‘this looks good, I will do this’. Then you are like ‘Oh! it is not correct’...” LIFEEXP11

Learning content relatability appears to influence self-evaluation of learning progression given a contextualised framing of learning objectives. Admittedly, it is worth noting that such necessary

equipment for providing quality healthcare might not be readily available in typical rural facilities, but there is a need to keep the app in line with what should be done so as to continually reinforce best practice (Shenouda, Davies, & Haq, 2018). Curiously, the typical learning task strategies especially utilised by participants who had limited prior knowledge of task was use of guesstimation (educated guess) approach as illustrated below.

“...Yeah maybe one or two I guessed, and got... I eliminated the wrong and I arrived on the right...” KPA002

But guesstimation, when linked to the need for elaboration earlier addressed can be viewed as learning by understanding what not to do. Healthcare providers evaluated this learning strategy to be successful if it produced perfect conformity to the clinical care guidelines as evidenced in the quotes below.

“... [The score] represents the kind of management I am doing. Because if you do not score well that means the care you are giving is not right... let's say you have scored 20%, what is going to be the outcome of this baby? You are going to have a baby with severe cerebral palsy. You are managing in terms of, you know... to come out with somebody who is alive and not having any major complications...” LIFEEXP01

“...I think scoring 70% means you never understood the full training. It doesn't mean you can't provide all the care, because even scoring that 70% is still actually part of the training. But a person who scores 100%, there is a tendency for this person to be better trained as compared to the person who is scoring 70. But it doesn't mean that the person who is scoring 70 cannot take part in resuscitation...” LIFEEXP03

Even in the gamified smartphone-based training linked to clinical scenarios, there is limited perception of internalisation of learning where performance is not 'automatic' (i.e. characterised by very quick response to the learning task) and have a perfect score. Performance, while being regarded as not being exclusively constitutive of knowledge mastery, cannot be dissociated from the perception of knowledge mastery. Healthcare providers concern for patient outcomes was heightened by how they

perform, but they do not consider this to be representative of their knowledge mastery, which they implicitly referred to as not being directly observable i.e. latent.

Stage two participants made a distinction of whether personal learning goals needed to be set based on (1) previous exposure to ETAT+, and (2) the number of previous plays on LIFE. These they argued, would mitigate any negative effects stemming from learner's first experience with ETAT+ content, and once acclimatised to LIFE and ETAT+, focus on implications of erroneous knowledge on clinical outcomes. Where there was no previous knowledge of the content, the emphasis was on an exploratory learning approach, with goal-setting exercises building on previous rounds of play.

"...For beginners, maybe no, but [maybe] if someone has either played it before or has done it before. I feel like because you, you know the sequence, you'd be like, 'Okay fine the next step is I want to do it in less time' ... for the purpose of this learning, it, it wouldn't be that useful...doing the game the second time, I think that will improve on your time and like improve on your skills. Yeah maybe it will be useful for those people, for first-timers it is discouraging..."
KPA 003

"...Well for me, I think getting a perfect score, translated to how well I understood the scenarios... like I was reminding myself what I had forgotten about ETAT+. So for me the motivation was other than the high score, whenever I scored higher than I did initially, I was like 'Wow, I think I am getting better', you know? it is like I am beginning to showcase whatever I have learnt in that particular scenario and putting it into practice now, like '...how well had I gotten the concept and how well I'm replicating it in the scenario now...' LIFEEXP10

At a global level, LIFE's goal-oriented imperative of saving lives at birth was well understood and internalised by the participants in all the qualitative study stages in general, with the clear intended consequence being building confidence to provide care.

"...If I had a life, in front of me that means I would have gotten it wrong when I'm doing an emergency procedure that I need to know, so that was scary... I know it's more of goal-oriented, to have more confidence, you need to pass, you need to get better..." KPA002

These strategic learning adaptations informed by personalised learning goals which constantly evolve based on perceptions of knowledge mastery have implications on the type of instructional support expected and to what end. I expound on this in the next section.

9.3.5 Feedback and motivation, and transfer to practice

Feedback is the key mechanism for maintaining high SRL behaviour by allowing learners to monitor their engagement and attainment of their learning goals (Butler & Winne, 1995). In LIFE's case, it served to highlight the extrinsic motivation for learning, which healthcare providers understood to be the adequate management of the neonatal emergency cases. It also unearthed self-efficacy concerns from healthcare providers' reflections:

".... I am a healthcare worker, how do I miss that, how do I get that wrong in resuscitating a new-born? But it's good it's there because it helped me know, 'you are wrong on this one'. But on the other hand, ... I think I'm very bad if I can get a wrong on this..." KPA002

They perceived their self-efficacy concerns to be inseparable from their extrinsic motivation linked to a higher game score. Feedback needed to build up motivation for personal goal achievement at the individual learning task level in addition to encouraging the global goal of perfect automatic performance. This was reflected in stage one by items **D** (*Immediate feedback due to the learning scenario "urgency"*) and **E** (*Improved feedback utility*) in Table 9.3.2.

"Q: Having played it, in terms of the type of feedback you got from LIFE, what would you like to see?

A: Maybe for them [the feedback messages] to say what I could do better, Okay, the overall point of the game was that we did it and the baby cried, I don't know if it's possible for them to, to say maybe he should have bagged faster, is there an option for that?..." KPA001

"...maybe it's to say that 'you've gotten at least X right ...It's essential that you get this one also right...' A way to imply that 'Yes, you're wrong, still, but try... we need you to improve'. What is the motivation after I get the wrong things? That's what I'm thinking, what is the motivation? ..." KPA002

"...you can just give feedback at the end of it and tell me 'at this stage went wrong and this was supposed to be the answer' and then after revising something and then you go and start

doing it, after getting where you went wrong, I will try and remember. Instead if you are done with this step and it tells you immediately... you feel stupid. Personally, I felt so stupid and I was like 'you don't know this thing?'..." LIFEEXP02

This illustrates how the corrective feedback messages (a) motivate individual learners to *keep reflecting in and on their action* and (b) might encourage healthcare providers to apply what they learn through LIFE in routine hospital settings as illustrated in the comments below.

"...[LIFE] is better than nothing because when we see that and when you call your consultant over the phone because you can be doing a resuscitation and you are like 'this child is actually not responding to the bagging that I am actually doing', once they talk to you over the phone and tell you that you need to actually give thirty breaths per minutes, if you have actually gone through the [LIFE] app ...like three times probably you would have seen the messages somewhere, and it would be better than that someone who has never heard about thirty breaths per minutes..." LIFEEXP03

"...But you know, at first before I played the game, I could have killed so many babies, but right now actually I am so confident, so I can handle it..." LIFEEXP02

Within gamified learning, targeted approaches leveraging on heightened learner emotions are needed to enhance learner engagement (Özhan & Kocadere, 2020). Through such targeting, gamification makes it easy to repeat cycles of game play, evaluation and reflection on learning performance, and adapting the learning strategy for next iteration, free from associated real-life risks associated with clinical learning content.

"...Because I understand that with such games, some of them actually they are more intense in the fact that they tell you that the baby has died... I've seen it before in other games and I would think when on emotional level it's somehow useful, you would want to save the life again. You know, through applications and games, it's the only time you can revive, resurrect someone to try and save them again..." KPA004

"...You know when you play it [LIFE], it gets so interesting and then you are just like 'why did I get 75 and I want to get to 100?'...That was my goal! And then the more you play it and the more you want to hate it, but the more you are learning from it. And the more it gets easier..." LIFEEXP02

The reflection and adaptation within the repeated learning cycles are exemplified in the previously mentioned 'guesstimation' strategy where healthcare providers with partial understanding of care giving procedures try to deduce the gap in their knowledge by a partial elimination mechanism when responding to learning tasks, or reflect on their experience in routine clinical practice. Healthcare providers perceive their smartphone based SRL to be limited when their individual motivation for reflective learning is not targeted by the platform nor does learning address confidence and courage to act in the routine clinical setting. This perception of limited learning reflects on their views of whether they are acquiring knowledge mastery (despite good performance) through SRL using gamified platforms.

It is peculiar that this subsection was barely mentioned in stage one, except as primarily intended to produce better scores instead of better capacity to respond to personal growth areas in delivering emergency care. These differences might be partially attributable to differing level of experience mediated by specialisation training, or due to the methodological differences between stage one NGT and the complimentary stages two and three.

9.3.6 Game experience and engagement

As a gamified smartphone-application, healthcare providers provided their views on the trade-off between content, design, and provision of supplementary learning resources in enhancing the game experience. In this regard, while stage one participants would want more task elaboration, stage two participants reflected on its perceived negative impact on the gamified experience.

"...You don't want to go so far from being in the game and actually is a bit of a game and a massive scientific quiz and research and information overload. Because people can get information from books and things...keep the game as a 'this is how we resuscitate babies, lets practice doing it, let's get into a routine, this is the protocol to follow', not digging so much in to the science behind it...." KPA005

"...I feel like if you added anything it would be too long, and someone would get bored in-between, but it would be a tough balance to add something and still keep people interested..." KPA003

However, the effort healthcare providers applied in learning tended to be low due to perceiving the platform as not requiring utmost serious-mindedness while using it. This could be attributed to a limited experience of immersion while using LIFE (Özhan & Kocadere, 2020). Such a learning posture towards gamified learning might hamper its acceptability as a novel modality of teaching if does not target to deeply immerse the learners (Özhan & Kocadere, 2020).

"...You know like the way I was playing it, at first, I was playing it while watching movies but now like when someone tells you play this game there is a qualification you will do it with a lot of seriousness in it... I would play it with a lot of care and a lot of concentration..." LIFEEXP02

"...Probably I would have been keener because I told you that in one of the scenarios, class was going on and I was just here [playing it]. So, I wasn't concentrating, so probably I should have been keen and taken it as more of life and death situation. As opposed to treating it as just a game. So, I think if I would have been keener... I would have been more meticulous..." LIFEEXP10

"...I think there will be some initial resistance from people trained on mannequins and real babies...people saying, 'is this [form of] paediatrics training better?' ..." KPA005

Effort regulation will produce higher goal attainment where gamified platforms such as LIFE ensure heightened keenness and sombreness of healthcare providers learning through it. Additional incentives to foster such postures towards SRL behaviours need to be further explored.

From the key findings, internalisation of the learning task would be maximised in the presence of feedback that elaborated on the consequences of [in]action or erroneous action. Such internalisation would be in response to learning content that was relatable due to its reflection of the clinical context and would help foster an engaging learning journey. Healthcare providers perceived themselves to have not internalised the learning content where their performance was not perfect nor were their responses to the learning tasks automatic. They also perceived achieving a perfect score as not being fully constitutive of knowledge mastery. Heightening the keenness and emotions of healthcare providers while learning was perceived to lead to enhanced learning and optimised effort regulation

(Özhan & Kocadere, 2020). From the study participants, the validation that they had met their learning objectives was when they perceived that their courage to act in a routine clinical setting had been enhanced from the confidence in the knowledge gained produced in a healthcare provider. These key findings, which represent how the interviewed healthcare providers perceive SRL to be linked to their personal learning objectives, are illustrated by Figure 9.3.1.

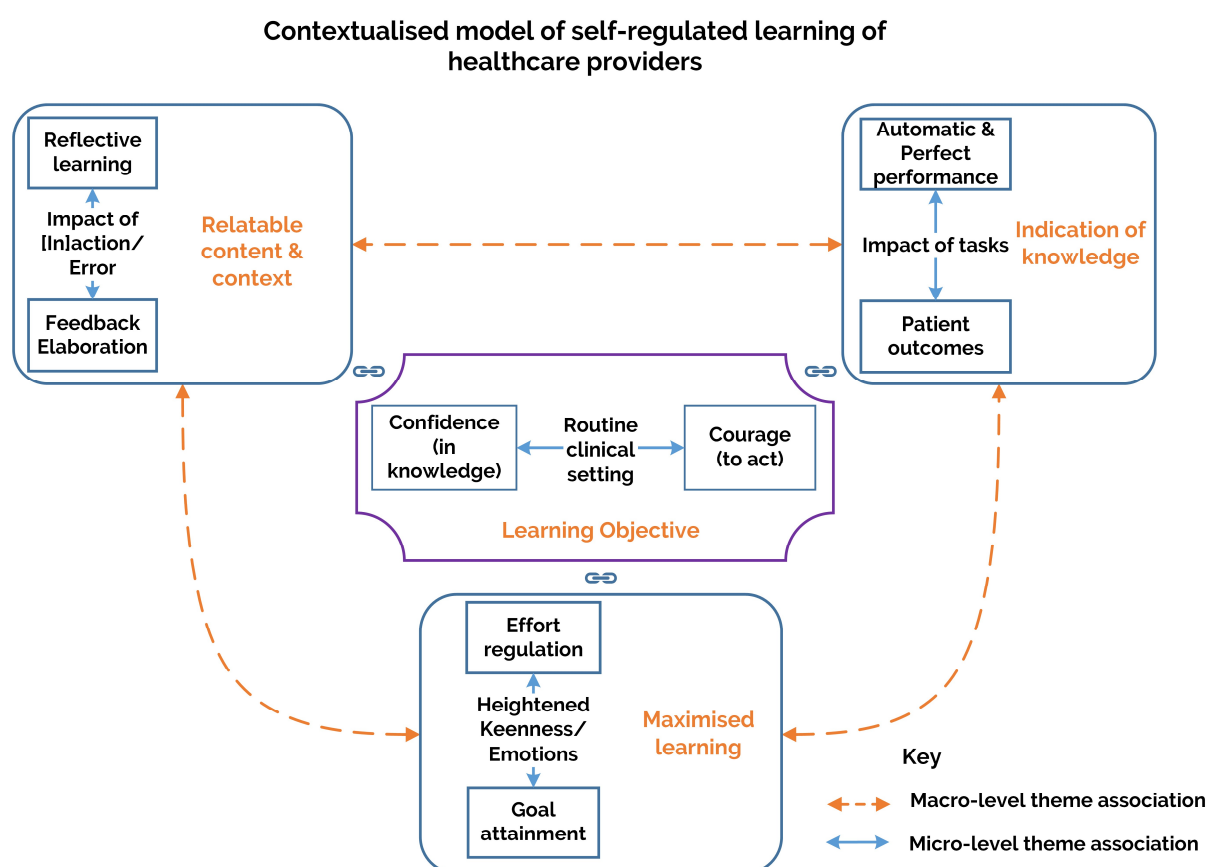


Figure 9.3.1: Illustration of the perceptions of self-regulated learning on gamified scenario-based training on digital platforms of the study participants

9.4 Summary of findings

The qualitative research sought to explore the perceptions of Kenyan healthcare providers on the use of a gamified smartphone-based application platform (LIFE) for emergency care training in their context of clinical practice. Specifically, the interest was in how they perceived such platforms might be better implemented to enhance more engaging and fruitful self-regulatory learning experiences. The study participants were geographically from diverse regions of Kenya and included different clinical cadres involved in paediatric care at different levels of practice, experience and seniority.

The qualitative research highlighted differences in focus between participants in the different phases of the studies; for example, healthcare providers with less clinical domain specialisation concentrating on the time-mediated performance and learning task-elaboration while those with more domain-specialisation being more concerned with minimising risk on clinical outcomes associated with erroneous knowledge.

10 DISCUSSION

10.1 Implications of findings

The finding that at the group level, the effect of adaptive feedback on healthcare provider learning gains was not significant, but the opposite is true when evaluated at an individual level might imply the intervention effect is strongly mediated by other factors (some of which might be external to the smartphone application). Chief among them being the healthcare providers' self-directed spacing of learning repetition. Together, with the inclusion of demographic characteristics, this increased the explainable variance of adaptive feedback on individual learning gains by 18.31%. Linking the level of adaptive feedback provided to healthcare providers' individualised learning repetitions conditioned on their level of experience and clinical role might explain the stronger intervention effect at the individual learner level.

This difference between the group and individual learning gain might also imply that the decision of when to collect post-intervention or repeated measures is a significant factor in determining what effective learning intervention is. Considering spaced learning, intuitively, the intervention effect becomes more reflective of a mechanism that works on optimising the recall rate as opposed to the level of internalising/understanding content. Despite efforts to encourage the latter mechanism by using reflective cues, from findings of the qualitative evaluation for this thesis, an alternative model using elaborative cues on why the choices of the healthcare providers were wrong would arguably enhance internalisation of the learning content.

The high drop-off rate with each subsequent round of play despite a lack of maximising on the learning gains was disconcerting and limits assessment of learners' skill mastery over time. The high drop-off rate might imply that either (1) effort regulation and motivation for self-directed learning using LIFE is very low, or (2) healthcare providers self-assessed learning needs lead them to believe that they have attained content mastery and thus are no longer in need of this training, or (3), other factors that are external to this experiment affect the use of the LIFE app (e.g. uninstalling the LIFE application to make

more space for other apps on learner's smartphone). The qualitative interviews will be used to explore whether any of these explanations are plausible and how they can be leveraged to encourage low drop-off rates.

The findings for LPA analysis in Chapter 9 should be interpreted cautiously for a number of reasons including the known disparities between the learning platforms, the pragmatic constraints on recruitment methods, and the lack of available information to better differentiate between the respondents' context of learning. Nevertheless, the findings raise important issues for the role of technology in healthcare training in low-income countries.

Learners in the low SRL profile are more likely to be of higher clinical cadre (Chapter 6, Table 6.3.5). They demonstrate the earliest drop-out during learning just after they achieve maximum learning gain (Figure 6.3.4). Apart from the possible explanation that the content might not be as challenging for learners in this profile, another possible explanation for the low SRL behaviour of these learners might be that their learning *context -if conceptualised as a set of situational stimuli-* needs more restructuring to improve both their learning experience and regulation (de la Fuente-Arias, 2017). This might be because the clinical practice context which informs their learning is likely *a-regulatory*, i.e. it does not promote self-regulation in learning but instead needs more external regulation (de la Fuente-Arias, 2017). To what extent this is true in LIC context remains unclear and is being investigated qualitatively. For regulated learning to be more effective for the low SRL learners, it might need to be restructured in a way that caters for the role of the interaction between an *individual* learner and *contextual* characteristics. In exploring the plausibility of these a-regulatory learning behaviours, future research needs to investigate what healthcare providers' expectations of success in SRL are and how they value SRL in view of their contextual contingencies. This will generate better understanding of their learning experiences that can inform the implementation of learning platforms that better engage and challenge them as they continue making gains in their learning.

Additionally SRL measurement tools might be incongruent to a healthcare provider's learning efficiency (Rovers, Clarebout, Savelberg, de Bruin, & van Merriënboer, 2019), but is necessary to address in future developments in education (de Bruin & van Merriënboer, 2017). Efficiency in learning, which might be another reason of the differences observed in learning gains between the SRL groups (Figure 3), might be better addressed using other learning theories e.g. the Cognitive Load Theory (Phan, Ngu, & Yeung, 2017). Future research can seek to evaluate how to combine the alternative learning theories linking efficiency in learning and focus on learner's attempts to deal with complex information efficiently with SRL theory, which focuses on learner's effort to monitor and control their learning (de Bruin & van Merriënboer, 2017).

From findings of RQ2(*"What influence these SRL strategies may have on healthcare providers' learning gains"*) in Chapter 9, the implication here is that scaffolding of content and instructional support might possibly play a significant role in ensuring the continuous professional development of healthcare providers in low-income contexts, by ensuring that their motivation to engage in constant learning is targeted by their SRL profile characteristics (Duffy & Azevedo, 2015). Alternatively, based on knowledge gaps that learners are keen to address and their SRL profile learners, learners could be given access to a wider range of learning scenarios and be allowed to also structure and customise the learning content they receive to reflect clinical scenarios they struggle with. This can be explicitly targeted during the digital scaling up of the Low-Dose-High-Frequency (LDHF) learning approach commonly used (and considered effective) in face-to-face clinical training in LICs (Atukunda & Conecker, 2017) using digital learning platforms. Platforms similar to LIFE that use LDHF training approach would help maximise learning gains if combined with cascaded content challenge and instruction support levels to encourage a more engaging learning experience. Additionally, it would serve as a means for facilitating the refreshing and updating of healthcare providers' knowledge in keeping up with the ever-evolving clinical practice guidelines.

Given that task strategy and help-seeking had low item means across the profiles, this might imply that healthcare providers generally have low effort-regulation when it comes to learning on digital

devices in this context. Additionally, help-seeking -while being lowly regarded in clinical context due to being misconstrued as evidence of incompetence- might not be an easily improvable SRL strategy (Dueñas et al., 2018; Maibach, Schieber, & Carroll, 1996).

LDHF scenario-based clinical training conditioned on skill difficulty and learner proficiency, produce higher cumulative learning gains where feedback provided is conditioned on previous practice at the learning task and the inferred level of the clinician. Additionally, such clinical education interventions might need to accommodate learners whose training needs require more complex scenarios, in addition to reflective feedback being provided. While this is not yet true for most (if not all) LDHF clinical training in LICs, using knowledge-tracing models might help differentiate learning needs with 84% accuracy, and guide better instructional support design. For interventions such as LIFE, where the content (neonatal resuscitation) is implicitly time-sensitive, time spent on learning task might not necessarily reflect learners' adoption of that assumption. Rather, it might be more indicative some of the learners individualised strategic learning objectives which are not linked to getting high performance in the shortest time possible but rather, taking time to reflect on the learning concept(s). This, however, was not equally true for all healthcare providers. From this thesis findings, a cognitive model for clinical training in low-resource settings using smartphone devices might be better served if it encourages repeated practice on a weekly basis, while providing learners with adaptive reflective feedback and content, conditioned on the inferred level of knowledge from previous practice opportunities. In a context where the healthcare providers tend to be significantly overworked, underpaid, and under-resourced, and therefore overstretched, interventions such as LIFE ought to optimise when and the way they encourage them to carve out time to refresh their clinical training knowledge. From this thesis findings, this would allow for learners who are activated to continue to learn in the presence of both past successes and complex scenarios, while offering more support to those in most need in the form of adaptive feedback.

Additional findings from the evaluation of knowledge tracing approaches, leveraging information about time spent on an individual learning task, the level of feedback provided in each attempt, and spacing between learning sessions, an Auto-Encoding DKT approach only needs four ‘time-lagged’ sequential attempts on learning tasks to start making comparatively more accurate inferences about the knowledge of the healthcare provider compared to common knowledge tracing alternatives. This allows DKT approach to dispense away with ballistic approaches of BKT and LFA models, where model-fitting has to take into account huge amounts of longitudinal learner data to generate insight (Deonovic et al., 2018). The DKT Auto-Encoding approach overcomes three fundamental challenges in providing adaptive learning experiences:

1. Fluctuating knowledge states of the healthcare providers are assessed by a modelling approach that treats learning as a continuous latent process (unlike BKT), and overcome the inherent cross-sectional nature of LFA models (Deonovic et al., 2018; Sapountzi A. et al., 2018).
2. Up-to-date, accurate predictions, inferences and sensible recommendations can be made available after four learning attempts into the first learning session are made by a new learner, i.e. where scant learning data from the healthcare provider is available (Ekanadham & Karklin, 2017).
3. Inference engines and frameworks are progressively implementing functionality that can work offline i.e. without internet, which greatly increases the capacity to increase their use for learning contexts in LICs. The DKT Auto-Encoding presented in this thesis is implementable in an offline form (Abadi et al., 2016; TensorFlow Team, 2019).

As a knowledge tracing intervention that is linked to skill performance mastery, while the evaluation of knowledge tracing approaches on LIFE platform provides a necessary starting point for understanding knowledge gaps in healthcare providers, adding a layer of multi-modal learning using *in situ* high fidelity simulation training provided by platforms such as Virtual Reality (VR), or Mixed Reality (MR) might arguably enhance hands-on experiential learning: Such platforms can transform

the learning experience into an even more meaningful one aimed at building confidence and self-efficacy of the healthcare provider in performing necessary life-saving care (Sheik-Ali, Edgcombe, & Paton, 2019; Winters et al., 2019). The knowledge tracing approach reported in this thesis sets the stage for progression to a multimodal evaluation of healthcare training experiences research in LICs, being cognisant that (a) learning processes are not independent, neither is a healthcare provider's learning constrained to only one mode at any given time and (b) dependencies of different pedagogical approaches are still remain relatively unexplored using the same knowledge tracing approach (Deonovic et al., 2018). Future work into knowledge tracing for multimodal learning in clinical training needs to evaluate how situated learner behaviours in LICs context are best targeted and catered for.

Learning also takes place in social practice, and therefore tutors and colleagues are vital for any healthcare provider's self-regulated learning in their context of practice. However, in this thesis, the focus was on how individual healthcare providers perceived themselves to have overt control over their self-regulated learning on digital platforms as opposed to how others in the clinical context of practice influence them (Berkhout, Helmich, Teunissen, van der Vleuten, & Jaarsma, 2017). Future research may address if and how the nature of learning in social and collaborative contexts might influence SRL behaviours (e.g. Help-seeking) of healthcare providers' in LICs. This may explain, for example how external learning regulation from using clinical peer influence for learning support (Saque-Rockoff, Ciardiello, & Schubert, 2019; van Schaik, Plant, Diane, Tsang, & O'Sullivan, 2011) after catering for individual learning motivation (Duffy & Azevedo, 2015) on their preferred digital learning platform. It might also help address the a-regulatory learning behaviours of healthcare providers that I have previously alluded to.

This thesis has highlighted qualitative differences in the learning perceptions and SRL experiences of healthcare providers in its different phases and stages. Healthcare providers with relatively less clinical domain specialisation concentrated on the time-mediated performance and learning task-

elaboration, while those with more domain-specialised were concerned with minimising the risk associated with erroneous knowledge (and clinical outcomes) and standardising knowledge. As already indicated, self-regulated learning in gamified digital platforms within this context ought to cater to differentiated goal-achievement orientation by using novel and elaborative feedback mechanisms. However, these novel feedback designs while helping to support learners to diagnose, recognise, recover from errors, and ensure adherence to emergency-care standards (Magoulas, Chen, & Papanikolaou, 2003), should not minimise the 'game' experience lest they render the platform to be similar to traditional learning tools which minimise the posited benefits of gamified learning on learner engagement (Kiili et al., 2012).

Such novel elaborative feedback designs need to be underpinned by a clear theory of change of how at the learning-task level, motivation to continue learning is enhanced (Kiili et al., 2012), and at the transfer-to-practice level, how such motivation might produce the intended action. These theories of change also need to (1) Highlight how they intend to create and maintain heightened keenness in healthcare providers SRL experience while being cognisant of the learning content and context, and (2), How they aim to enhance repeated learning so that as a skill particular to an individual increases, the game might adapt learning content and instructional support to their change in starting status. The former will aid in minimising flippancy in self-directed use of gamified digital platforms, while the latter, coupled with the knowledge tracing findings of this thesis, is possible with a relatively high degree of accuracy.

Another key implication of measurement in learning, even on gamified digital platforms, is being able to distinguish knowledge from performance, and in efforts to satisfy the healthcare providers interest in both, give an indication of both. The learners (who are healthcare providers in this case) make this distinction since they do not consider perfect performance to be wholly constitutive that they have grasped the concepts well. Informing them how they perform might not suffice if they cannot reconcile how they perceive their knowledge. The emphasis on the adoption of knowledge tracing

approaches for clinical training addresses this conundrum by being able to provide inference of the healthcare providers knowledge apart from indicating their performance. This is important because change in healthcare provider's practice behaviour might not be observed until they are confident to act based on their perceived knowledge. Different approaches to building this confidence -whether through feedback messages, prompts for practical action linked to performance, provision of a fully immersive VR experience which can move them closer to real-life experience, or knowledge mastery approximation- ought to be considered and this thesis lays the groundwork for doing so in LICs.

10.2 Summary of the key findings

SUMMARY OF KEY FINDINGS

- ❖ The effect of adaptive feedback on healthcare providers' learning gain was found to be substantively and statistically insignificant when analysed at group level after two learning sessions. However, when individualised subject-treatment interaction and differential time effect was adjusted for, this effect increased significantly to 0.644 (95% CI: 0.35 - 0.94, p-value <.001) with immediate repetition -which is a moderate learning effect- but reduced significantly by 0.28 within a week.
- ❖ The overall learning gain from LIFE use in both experiment arms was relatively large that it might have obscured a direct effect of the adaptive feedback at the group level. However, the learning effect was not obscured at the individual or item level.
- ❖ Latent Profile Analysis (LPA) findings demonstrated that there are four SRL profiles in this population of learners: high, above-average with low help-seeking, average, and low SRL profiles. Consultants had three times the odds of being in the low SRL group, and online computer-based learners had 13 times the odds of being the average SRL profile. The above-average with low help-seeking SRL learners were 89% less likely to have used online computer-based learning and at least 65% less likely to be clinical officers or consultants.
- ❖ A unit increase in the years of experience in providing healthcare significantly increased the odds of healthcare providers being in the high SRL profile by 1.5 times.
- ❖ Compared to healthcare providers in the low SRL profile (who were mostly consultants), the other three SRL profiles had lower normalised gains which were never maximised even with more reinforcement learning cycles

SUMMARY OF KEY FINDINGS (CONTINUED)

- ❖ After controlling for the moderating effect of the healthcare provider's self-regulated learning on both time on task, and previous practice opportunities, the odds of getting the next attempt correct on first try was significantly higher (OR: 5.58, 95% CI: 4.29–7.26, p-value:<0.001) when the learner received reflective feedback through an adaptive mechanism compared to when they received detailed feedback through a standard mechanism.
- ❖ Wheel spinning (i.e. defined as the learner spending considerable amount of time or effort struggling to learn a concept without gaining any mastery), appeared to be highly and positively correlated with higher counts of knowledge mastery streaks among the healthcare providers. Wheel spinning is not indicative of a lack of knowledge mastery.
- ❖ As the analysis unit gets more individualised and granular, the magnitude of the effect of reflective feedback (and the adaptive mechanism by extension) becomes more pronounced across the varied self-directed learning spacing options.
- ❖ For LIFE learning data from healthcare providers from LICs, the extent to which the knowledge tracing model models make increasingly reasonable assumptions about contextual learning, the more accurately they represent and predict learners' knowledge and performance.
- ❖ Leveraging information about time on learning task, feedback provided, and spacing between learning sessions, an Auto-Encoding DKT approach only needs four 'time-lagged' sequential attempts on learning tasks to start making more accurate inferences about the knowledge of the healthcare providers compared to alternative common knowledge tracing approaches.

10.3 Relation to published literature and studies

10.3.1 Experimental findings

The experiment in this thesis differs from (and addresses gaps in) previous similar studies in digital education for clinical practice in that it: (1) used mobile devices for delivery of digital education interventions, (2) evaluated novel educational modalities enabling simulated learning such as adaptive feedback, (3) provided essential methodological information necessary for comparability purposes, (4) reported relative learning gains which is uncommon (digital education studies in health tend to report differences of post-intervention scores in the study arms), and (5) was done in low-income settings such as Sub-Saharan Africa or South-East Asia (Car et al., 2019). In the three studies that dealt with resuscitation identified by the systematic review by Car *et al.* (Car et al., 2019), only one study reported relative mean change but found no significant effect of different formulations of online content on learning gain (Leszczyński et al., 2018) which is similar to this experiment's findings, even though the experiment's primary outcome is calculated differently from theirs. In that study, a variant of individual students' normalised gain was used, like the secondary outcome I used in this thesis's experiment where I found the effect on individual student learning gains was 0.644 (95% CI: 0.347 to 0.941, p-value <.001) when considering individual learner characteristics.

This study has addressed some of the recommendations from a recent evidence review into gamified education in health which proposed use of rigorous experimental design to evaluate learning interventions, and involving participants from low- and middle-income countries (S. V. Gentry et al., 2019) which are currently arguably non-existent. Even when considering interventions looking into adaptive feedback in digital education (Bimba, Idris, Al-Hunaiyyan, Mahmud, & Shuib, 2017), none that I know of, are in the health domain making this thesis's experiment unique. Given that the implementation of the learning intervention was available to all clinical cadres involved with bedside care provision, who had varying levels of experience (from students to consultants), from multiple low-income countries, representing a varied mix of geographical and resource settings, the diverse population of the clinical taskforce in this experiment ensures that those who would most benefit

from using the presented learning tool for training on emergency care delivery are well-represented. Furthermore, as the overall learning gain from LIFE use in both study arms was large, given that the learning platform minimised elements that would not be typically available in routine application settings (Peters et al., 2013), this experiment's findings are more generalisable to emergency care training of the healthcare provider population from low-income countries using smartphone-based platforms.

10.3.2 Self-regulated learning of LICs healthcare providers on digital platforms

Unlike previous studies that have examined LPA for adult learners on digital devices, I found that within healthcare providers in LICs, four substantive latent SRL profiles may exist as opposed to three (Abar & Loken, 2010; Vanslambrouck et al., 2019). I also found that a majority of healthcare providers from low-income contexts fall under above-average SRL profile membership with low-help seeking while both high and average SRL profiles having almost equal membership unlike other studies which would indicate that a majority of the learners may tend to fall within the average SRL profile (Abar & Loken, 2010; Vanslambrouck et al., 2019). The LPA analysis also sought to ensure that a diverse population of healthcare providers from LICs not limited by clinical speciality, cadre, or digital learning platform were included in the analysis. The included healthcare provider population is substantively more diverse compared to other empirical studies looking at SRL in clinical training (Cho, Marjadi, Langendyk, & Hu, 2017; Keçeci, 2017). In line with the most recent review of SRL in clinical settings, evidence suggests that further research is needed to unravel and reconceptualise the sub-processes of SRL that are relevant to the clinical context to contribute to more elaborate context-specific SRL frameworks (van Houten-Schat et al., 2018). The analysis reported in Chapters 6 and 9 serve as formative response to that call, is especially relevant to LICs, and makes a good starting point for future implementation of context-relevant SRL-informed digital learning interventions for healthcare providers in this region.

10.3.3 Effect of adaptive feedback on practice opportunities

In appreciating that reinforcement of learning in clinical domain is useful in enhancing clinical outcomes, interventions in emergency care training in low resource settings tend towards face-to-face group training and, more recently, have used low-dose high-frequency (LDHF) in-person training in group settings (but did not utilise technology) (Chaudhury et al., 2016; Willcox et al., 2017). The emphasis on using LDHF model in for emergency care training is not new but has been recently introduced in SSA (Atukunda & Conecker, 2017). LDHF training is commonly done to mitigate against forgetfulness of knowledge and skills within health workers in LICs (Willcox et al., 2017). How different clinical training interventions use cognitive scaffolds to improve knowledge gains is not new (Feyzi-Behnagh et al., 2014; Veredas, Ruiz-Bandera, Villa-Estrada, Rufino-González, & Morente, 2014; V. Wong et al., 2015), but hardly present in research from low resource settings (Opiyo & English, 2015). This thesis goes further by demonstrating that for LICs, learning gains can be and are significantly higher where the practice opportunities in LDHF scenario-based training is matched with reflective feedback predicated on the inferred learner's knowledge.

Additionally, I found that low self-regulating healthcare providers -who have higher learning gains with fewer learning opportunities- would arguably benefit more from challenging scenarios in addition to reflective feedback (Figure 6.3.4, Table 7.3.2). Moreover, while using adaptive learning demonstrates significantly better knowledge gains than alternatives (Feyzi-Behnagh et al., 2014; Veredas et al., 2014; V. Wong et al., 2015), the current LDHF training models –which are not adaptive to individual learner needs nor do they utilise smartphone platforms- are still the most commonly implemented models of learning, usually face-to-face, and at a very high cost (Atukunda & Conecker, 2017; Chaudhury et al., 2016; Willcox et al., 2017). This thesis findings generates evidence into how the use of smartphone devices to deliver clinical training using LDHF simulated scenario-based learning activities, can utilise self-directed spacing to facilitate self-evaluation self-regulated learning, which tends to be ignored in SRL studies in clinical settings (van Houten-Schat et al., 2018).

10.3.4 Knowledge tracing for clinical learning interventions

There are very limited cases of use of DKTs together with forgetting curves (i.e. learning decay) in healthcare provider training. An example of learning decay analysis in a clinical care setting has been in radiography skills where it was mitigated against by exposure to repeated testing every two months (Boutis et al., 2019). However, typical learning intervention studies in clinical training using digital platforms rarely employ the use of knowledge tracing approaches, and the paucity of knowledge tracing evidence in clinical education grows bigger when LICs contents are considered (Fontaine et al., 2019). From applying knowledge tracing approaches on LIFE data, this thesis findings indicate that a Deep Knowledge Tracing (DKT) approach is comparatively better than the alternative commonly used knowledge tracing approaches (Sapountzi A. et al., 2018) when applied to smartphone-based LDHF training model commonly used in LICs for emergency care training (Atukunda & Conecker, 2017; Chaudhury et al., 2016; Willcox et al., 2017).

In the health domain, LSTMs neural networks have been used in clinical settings to predict the occurrence of clinical events and clinical diagnoses (Kaji et al., 2019; Lipton, Kale, Elkan, & Wetzel, 2015). But in this thesis's use-case, LSTMs application has been extended to tracing healthcare providers knowledge in providing emergency care using simulated clinical scenarios. This thesis's emphasis on neonatal emergency care in LICs is due to both the need for scalable access to training for emergency care training in LICs, and the subsequent impact it promises on better neonatal patient outcomes by reducing avoidable deaths (Sandra Barteit et al., 2019; Couper et al., 2018). How different clinical training interventions use cognitive scaffolds to improve healthcare providers' knowledge gain is not new (Feyzi-Behnagh et al., 2014; Veredas et al., 2014; V. Wong et al., 2015), but this evidence has been considerably opaque as to how knowledge tracing was achieved, and hardly presents research evidence from low resource settings (Opiyo & English, 2015) making its impact limited to the Global North. Such evidence does not improve understanding of situated learning from diverse contexts such as LICs (Dawson et al., 2019).

Moreover, while adaptive learning produces significantly higher knowledge gains than alternatives (Feyzi-Behnagh et al., 2014; Veredas et al., 2014; V. Wong et al., 2015), the current training models used in LICs -which are not adaptive to individual learner needs- are typically face-to-face, and offered at a very high cost (Atukunda & Conecker, 2017; Chaudhury et al., 2016; Willcox et al., 2017). The findings from the knowledge tracing component of this thesis, begin to address how the use of smartphone-based clinical training for short simulation-based learning activities can begin to accommodate self-regulated personalised learning over time, which arguably optimises learning outcomes in the health domain (van Houten-Schat et al., 2018).

10.3.5 Qualitative exploration healthcare providers perceptions on gamified self-regulated learning on smartphones

I have initiated the implementation of the recommendations from previous studies to “...find a point where the intrinsic and extrinsic motivation of the players can converge...” (Drummond et al., 2017) by starting to generate insights into self-regulated learning experiences necessary for improving gamified learning in clinical training. Such insights address a gap in research evidence from studies that involve healthcare providers from low-income countries (Car et al., 2019; S. V. Gentry et al., 2019). With scarcity of comparable evidence, I provide theory-informed considerations for enhancing learning experiences design and implementation of digital clinical education, specifically in the Global South (Wang et al., 2016). Typical studies looking into the effectiveness of gamified digital clinical training do not highlight how learning theories inform their implementation or provide a framework for interpreting findings (S. V. Gentry et al., 2019). I provide theoretical underpinning for understanding how healthcare providers from SSA perceive (1) Their learning; (2) Interaction with the learning platforms that enhances their SRL, and (3) The theory of change for how their SRL-linked learning gains might be positioned for transfer to routine clinical practice. I offer a more nuanced explanation of the healthcare professionals’ learning behaviours linked to related evaluation studies reported in Chapter 5-8, unlike previous studies in this area (Car et al., 2019; S. V. Gentry et al., 2019).

I also have sought to thread the needle in this topic by providing a baseline conceptual framework that can inform the design and evaluation of similar interventions in contexts like Kenya.

10.4 Thesis strengths and limitations

In the randomised experiment, I used the standard Bayesian Knowledge Tracing modelling approach - which is explained in detail in the Chapter 3.4.1, Appendix C.2 and elsewhere (M. V. Yudelson et al., 2013) - for determining the adaptive feedback threshold cut-offs. While LFA models such as AFMs and PFM are posited to be highly predictive, they are argued to be difficult to optimise *subsets* of instructional content presented to learners according to their historic performance (Sapountzi A. et al., 2018) and or implement online let alone offline (Pelánek, 2016). Additionally, evidence on knowledge tracing approaches suggested that there was no clear winner between BKTs, AFMs and PFM for student modelling at the beginning of this thesis, and there performance posited to be context-specific (Chi et al., 2011). With these considerations in mind, I opted to use BKT as the mechanism for determining at what point to offer adaptive support- given its added ability to overcome the challenge of offline use of the model for prediction of performance using individual learner's data. This was achievable because a pre-trained standard BKT model can be implemented using a decision-tree structure that adjust model parameters for each individual learner while completely offline (Figure 3.4.1).

However, BKT has a set of problematic assumptions: it assumes that forgetting does not occur; the knowledge components (i.e. the learning tasks in LIFE's case) are treated as being mutually independent; its typical implementation does not allow for learners to have different learning rates; it assumes that all students have the same probability of knowing a particular skill at their first opportunity; and it suffers from the problem of multiple global maxima when trying to estimate model parameters (Gong et al., 2011; M. V. Yudelson et al., 2013). In addition, I used learning data from 187 healthcare providers from the beta phase to train the model, which is a relatively small number. Together with the moderate predictive accuracy of the healthcare provider's knowledge (Chapter

5.2.5), it is highly probable that the prescription of the level of feedback might have been biased, thereby misconstruing the effect of the intervention.

From interviews of the physically accessible study participants, those who were from lower clinical cadres found the one-off phone data cost necessary for downloading the smartphone application as being too high. To mitigate against burdening participants with this extra personal cost due to participating in the study, they could request for reimbursement within a few weeks of using the smartphone application. This skewed the study participants numbers towards a specific country where the reimbursements were made available (Table E.1 in Appendix E). Within these resource contexts, such costs play a role in generating study participants who might not be representative of those with the intention to participate rather representing those with the “economic” ability to participate, which is hard to mitigate against in a multi-country study. It is challenging to determine if the reimbursement favoured any study arm because of the blinded participant allocation to the study arms coupled with a lack of linking participants (whose data was collected as anonymised at source) to reimbursement.

The risk of performance bias across arms was moderate given that allowing for the “real-world” use of the smartphone application in a self-directed manner would also make it difficult to comprehensively control for exposure to factors outside the intervention of interest. This is especially true where some of the participants were in the same peer-groups, hospitals, or social circles and might have collaboratively used the smartphone application. Even though randomisation allocation did not result in equal number of participants in the study arms, from the post-hoc analysis, I do not believe that this biased this experiment’s findings.

While LPA model had reasonably good performance on the survey data unlike the planned confirmatory factor analysis, the relatively low explained variance in the models is disconcerting (Table 6.3.5). This is most likely because of the number of healthcare providers analysed, making it hard to give more precise estimations. Nevertheless, the study was looking at a largely unexplored topic and

population (i.e. latent quantitative analysis of SRL in low-income contexts such as a Sub-Saharan Africa). The findings have helped provide nuanced understanding of SRL especially for digital learning interventions used in clinical training where there is promise of much positive gains in patient outcomes in this region. I argue that these findings, internalised as a case-study underpinned by Self-Regulated Learning theory, follow the analytical and transferability model of generalisability (to context) as opposed to statistical generalisability (Polit & Beck, 2010). This type of generalisability can –and has been- enhanced by use of an in-depth qualitative study, which is reported in Chapter 9 in efforts to aid better understanding and explanation of the SRL findings.

And while the sample of healthcare providers used in this study is not statistically generalisable to the population of all healthcare providers in low income contexts, its make-up (from those in training to specialists, at all healthcare provision categories) makes it a highly informative and situated source of understanding self-regulated learning of healthcare providers in resource-constrained settings for clinical training on digital platforms, a gap recognised by most recent systematic reviews in these areas (Car et al., 2019; S. V. Gentry et al., 2019). I am yet to find any comparable studies on this subject in this context for this domain.

A key challenge this thesis aimed to address was optimising the gamified self-regulated learning on mobile devices. However, it still does not address the issue on whether such an approach can be introduced into health systems. This is because the study's interest was on how individual healthcare providers experienced and perceived themselves to have overt control over their self-regulated learning on digital platforms as opposed to how health system's organisation influences their learning.

A key strength in this research is the use of methodological triangulation to validate and strengthen findings. This was further increased by using varied participants with differing experience, specialisation levels and expansive geographical representation. However, the lack of gender diversity is a limitation of this study. For qualitative work in stage two, while every effort was made to include more participants, limited conference time made it difficult to recruit. This was further exacerbated

by the conference environment having moderate to extreme noise levels depending on the time of day making it difficult to conduct longer interviews. Given that findings across the qualitative study stages cohere, any bias due to interview length, and number of recruited participants is likely to be minimal and I concluded that saturation was achieved as evidenced by convergence of themes across stages (Saunders et al., 2018). It is also difficult to track impact on healthcare providers' practice behaviour due to lack of availability of high-quality clinical care process data from routine clinical setting linked to the providers.

11 CONCLUSIONS

This study set out to evaluate the effect of adaptive feedback within a smartphone-based serious game on the learning gains of healthcare providers from low-income countries. From 247 healthcare providers, the effect on learning gain was found to be $g=0.09$ (95% CI: -0.31 to 0.46, p -value = .474). When subject-treatment interaction and differential time effect was controlled for, the effect of the adaptive feedback on learning gains increased significantly to 0.644 (95% CI: 0.347 to 0.941, p -value: $<.001$). From this thesis findings, a key suggestion is that linking the level of adaptive feedback provided to healthcare providers to how they space their learning and their clinical level might yield a larger intervention effect both at group and individual learner level. For the feedback content itself, as an alternative to using reflective hints on what the right answers might be, elaborating why the healthcare providers' responses were wrong might enhance understanding of the learning content. As the level of analysis shifted from group to individual to task, the effect of the adaptive feedback mechanism grew in magnitude and statistical significance.

Healthcare providers' perceptions on the use of gamified learning on digital platforms for clinical training in SSA contexts like Kenya was that it ought to provide feedback elaborating the consequences of [in]action or erroneous action, which is hinged on relatable contextualised learning content. Participants felt this would help improve their learning experiences, including heightened engagement and effort regulation. Use of theory of change that validates the self-directed learning of healthcare providers by building confidence in the knowledge gained and produces the courage to act in a routine clinical setting is necessary if self-regulated learning is to be effective in SSA contexts. I posit that the context-reflective conceptual model emerging from this study will encourage further development of smartphone-based self-regulated learning platforms in efforts to make gains in better patient outcomes in SSA.

Self-regulated learning is critical for encouraging voluntary upskilling of healthcare providers in LICs, especially when using digital learning interventions. Although insight into healthcare providers' SRL in

digital environments is vital for enabling personalised instructional support necessary for better patient outcomes, research from low-income contexts is lacking. This thesis contributes to reducing this evidence gap by adding to empirical evidence on the SRL of clinicians in these contexts by generating insights into how these healthcare providers self-direct their learning on digital platforms. In addressing the second research question, the results of this thesis offer evidence of the differences in SRL between healthcare providers from these contexts who use digital learning environments, as they indicate four SRL profiles: *Low*, *Average*, *Above Average with Low Help Seeking* and *High* SRL profile. From these profiles, healthcare providers in the high SRL profile were likely to be junior doctors and nurses, had relatively high endorsement of Goal Setting, Strategic Learning, Task Strategy, Elaboration, Self-Evaluation, and Help-Seeking SRL skills with the converse being true for those in Low SRL profile who were more likely to be consultants. Healthcare providers in average SRL profile had average and relatively monotonic response to all these skills while those in the above-average SRL profile demonstrated low affinity for help-seeking behaviours while having above-average to high response on all other SRL skills. Longer experience in providing healthcare significantly increased the odds of healthcare providers being in the high SRL profile while healthcare providers in the average SRL profiles were more likely to be users of computer-based digital learning compared to smartphone-based learning. Additionally, from exploring the SRL profiles' association to healthcare providers' learning gains, the two average and the high SRL profiles had lower normalised learning gains but more repeated learning cycles compared to healthcare providers in the low SRL profile, even though it is only learners in the low SRL profile that were able to attain maximum learning gains. The implications of my findings are that healthcare providers from these low resource contexts who are using digital learning environments are typically, at least moderately self-regulating their learning, but that scaffolding instructional design might play a bigger role in enhancing task strategy and persistence in the SRL of healthcare providers from LICs when their motivation to engage in continuous but challenging learning, is targeted.

Future research into SRL on digital platforms for clinical training utilising scenario-based modular content (e.g. for low-dose-high-frequency model that is commonly found in LICs resource settings) might explore how to better enhance content complexity apart from instructional support scaffolding. This would be important in enhancing maximised learning over spaced repetition of learning when informed by -for example- healthcare providers' SRL task and help-seeking strategies in effort regulation.

The Deep Knowledge Tracing model -treated as a mechanism for dynamic evaluation of how healthcare providers from LICs are learning through smartphone (e.g. formative assessment)- provides a more reliable (and accurate) avenue for instructional contingency (i.e. at what level will the next instructional intervention be), domain contingency (i.e. Given the content the learner selected, what support should be provided?) and temporal contingency (i.e. How long should the learner be left struggling?) in designing better learning experiences and interventions compared to common alternatives. The extent to which the knowledge tracing models accurately represent the learner's knowledge influences the confidence with which targeted learning theories can be applied in instructional design to aid delivery of effective learning interventions.

A key concern in the knowledge tracing models is that with increasing model performance, there is also an increase in complexity in demonstrating how the model works, or, how to transfer the modelling approaches to other learning contexts. However, if the latent factor associated with "knowledge" is agreed to be complex and multi-dimensional, simpler knowledge tracing models are more likely to harbor reductionism, which hinders the extent they can be useful for formative diagnosing of the self-regulated learning needs of healthcare providers in LICs.

REFERENCES

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., . . . Isard, M. (2016). *Tensorflow: A system for large-scale machine learning*. Paper presented at the 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16).
- Abar, B., & Loken, E. (2010). Self-regulated learning and self-directed study in a pre-college sample. *Learning and individual differences, 20*(1), 25-29.
- Abdul, J., Azita, I., & Felicia, P. (2015). Gameplay engagement and learning in game-based learning: A systematic review. *Review of Educational Research, 85*(4), 740-779.
- Abraham, I., Cherubini, A., Cozzolino, F., De Florio, R., Luchetta, M. L., Rimland, J. M., . . . Orso, M. (2015). Deviation from intention to treat analysis in randomised trials and treatment effect estimates: meta-epidemiological study. *Bmj, 350*, h2445.
- Al, S. (2012). Need a Game Idea? A List of Game Mechanics and a Random Mechanic Mixer. *The Invent with Python Blog*. Retrieved from <https://inventwithpython.com/blog/2012/07/30/need-a-game-idea-a-list-of-game-mechanics-and-a-random-mechanic-mixer/>
- Aleven, V., McLaren, B., Roll, I., & Koedinger, K. (2006). Toward meta-cognitive tutoring: A model of help seeking with a Cognitive Tutor. *International Journal of Artificial Intelligence in Education, 16*(2), 101-128.
- Aleven, V., Roll, I., McLaren, B. M., & Koedinger, K. R. (2016). Help helps, but only so much: Research on help seeking with intelligent tutoring systems. *International Journal of Artificial Intelligence in Education, 26*(1), 205-223.
- All, A., Castellar, E., & Van Looy, J. (2014). Measuring Effectiveness in Digital Game-Based Learning: A Methodological Review. *International Journal of Serious Games, 1*(2).
- Arnab, S., de Freitas, S., Bellotti, F., Lim, T., Louchart, S., Suttie, N., . . . De Gloria, A. (2012). Pedagogy-driven design of Serious Games: An overall view on learning and game mechanics mapping, and cognition-based models. *Serious Games Institute*.
- Arnab, S., Lim, T., Carvalho, M., Bellotti, F., Freitas, S., Louchart, S., . . . De Gloria, A. (2015). Mapping learning and game mechanics for serious games analysis. *British Journal of Educational Technology, 46*(2), 391-411.
- Assel, M., Sjöberg, D. D., & Vickers, A. J. (2017). The Brier score does not evaluate the clinical utility of diagnostic tests or prediction models. *Diagnostic and prognostic research, 1*(1), 19.
- Atukunda, I. T., & Conecker, G. A. (2017). Effect of a low-dose, high-frequency training approach on stillbirths and early neonatal deaths: a before-and-after study in 12 districts of Uganda. *The Lancet Global Health, 5*, S12.
- Ayieko, P., Ntoburi, S., Wagai, J., Opondo, C., Opiyo, N., Migiro, S., . . . English, M. (2011). A Multifaceted Intervention to Implement Guidelines and Improve Admission Paediatric Care in Kenyan District Hospitals: A Cluster Randomised Trial. *PLOS Medicine, 8*(4), e1001018. doi:10.1371/journal.pmed.1001018
- Baalsrud-Hauge, J. M., Stanescu, I. A., Arnab, S., Ger, P. M., Lim, T., Serrano-Laguna, A., . . . Ninaus, M. (2015). Learning through analytics architecture to scaffold learning experience through technology-based methods. *International Journal of Serious Games, 2*(1), 29-44.
- Barnard-Brak, L., Paton, V. O., & Lan, W. Y. (2010). Profiles in self-regulated learning in the online learning environment. *The International Review of Research in Open and Distributed Learning, 11*(1), 61-80.
- Barnard, L., Lan, W. Y., To, Y. M., Paton, V. O., & Lai, S.-L. (2009). Measuring self-regulation in online and blended learning environments. *The internet and higher education, 12*(1), 1-6.
- Barnard, L., Paton, V., & Lan, W. (2008). Online self-regulatory learning behaviors as a mediator in the relationship between online course perceptions with achievement. *The International Review of Research in Open and Distributed Learning, 9*(2).

- Barrow, D., Mitrovic, A., Ohlsson, S., & Grimley, M. (2008). *Assessing the impact of positive feedback in constraint-based tutors*. Paper presented at the International Conference on Intelligent Tutoring Systems.
- Barteit, S., Hoepffner, P., Huwendiek, S., Karamagi, A., Munthali, C., Theurer, A., & Neuhaus, F. (2015). Self-directed e-learning at a tertiary hospital in Malawi--a qualitative evaluation and lessons learnt. *GMS Z Med Ausbild*, 32(1), Doc7. doi:10.3205/zma000949
- Barteit, S., Jahn, A., Banda, S. S., Bärnighausen, T., Bowa, A., Chileshe, G., . . . Malunga, G. (2019). E-Learning for Medical Education in Sub-Saharan Africa and Low-Resource Settings. *Journal of medical Internet research*, 21(1).
- Barteit, S., Jahn, A., Bowa, A., Luders, S., Malunga, G., Marimo, C., . . . Neuhaus, F. (2018). How Self-Directed e-Learning Contributes to Training for Medical Licentiate Practitioners in Zambia: Evaluation of the Pilot Phase of a Mixed-Methods Study. *JMIR Med Educ*, 4(2), e10222. doi:10.2196/10222
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.
- Baumann, B. (2018). The Modern Learner and Learning Experience Design. *Content Development*. Retrieved from <https://trainingindustry.com/articles/content-development/the-modern-learner-and-learning-experience-design/>
- Beck, J. E., & Gong, Y. (2013). *Wheel-spinning: Students who fail to master a skill*. Paper presented at the International conference on artificial intelligence in education.
- Bediang, G., Bagayoko, C. O., Raetz, M. A., & Geissbuhler, A. (2011). Relevance and usability of a computerized patient simulator for continuous medical education of isolated care professionals in sub-saharan Africa. *Stud Health Technol Inform*, 169, 666-670.
- Bediang, G., Stoll, B., Geissbuhler, A., Kohn, A. M., Stuckelberger, A., Nko'o, S., & Chastonay, P. (2013). Computer literacy and E-learning perception in Cameroon: the case of Yaounde Faculty of Medicine and Biomedical Sciences. *BMC Med Educ*, 13, 57. doi:10.1186/1472-6920-13-57
- Ben Bolker. (2019). Test fitted model for (near) singularity. Retrieved from <https://www.rdocumentation.org/packages/lme4/versions/1.1-21/topics/isSingular>
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological bulletin*, 107(2), 238.
- Bergeron, B. (2006). *Developing Serious Games*. Game Development Series. Charles River Media, Inc., Massachusetts.
- Bergin, A., Ford, E., & Hess, D. (1993). Patterns of motivation and social behavior associated with microcomputer use of young children. *Journal of Educational Psychology*, 85(3), 437.
- Berkhout, J. J., Helmich, E., Teunissen, P. W., van der Vleuten, C. P., & Jaarsma, A. D. C. (2017). How clinical medical students perceive others to influence their self-regulated learning. *Medical education*, 51(3), 269-279.
- Bimba, A. T., Idris, N., Al-Hunaiyyan, A., Mahmud, R. B., & Shuib, N. L. B. M. (2017). Adaptive feedback in computer-based learning environments: a review. *Adaptive Behavior*, 25(5), 217-234.
- Blanchard, E. G. (2015). Socio-cultural imbalances in AIED research: Investigations, implications and opportunities. *International Journal of Artificial Intelligence in Education*, 25(2), 204-228.
- Blattenberger, G., & Lad, F. (1985). Separating the Brier score into calibration and refinement components: A graphical exposition. *The American Statistician*, 39(1), 26-32.
- Bleakley, A. (2006). Broadening conceptions of learning in medical education: the message from teamworking. *Medical education*, 40(2), 150-157.
- Bollinger, R., Chang, L., Jafari, R., O'Callaghan, T., Ngatia, P., Settle, D., . . . Al Shorbaji, N. (2013). Leveraging information technology to bridge the health workforce gap. *Bulletin of the World Health Organization*, 91, 890-892.
- Boud, D. (2013). *Enhancing learning through self-assessment*: Routledge.
- Boutis, K., Pecaric, M., Carriere, B., Stimec, J., Willan, A., Chan, J., & Pusic, M. (2019). The effect of testing and feedback on the forgetting curves for radiograph interpretation skills. *Med Teach*, 41(7), 756-764. doi:10.1080/0142159x.2019.1570098

- Boyle, A., Connolly, T., Hailey, T., & Boyle, J. (2012). Engagement in digital entertainment games: A systematic review. *Computers in Human Behavior*, 28(3), 771-780.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2), 77-101.
- Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly weather review*, 78(1), 1-3.
- Brünken, R., Seufert, T., & Paas, F. (2010). Measuring cognitive load.
- Brusilovsky, P. (2007). Adaptive navigation support. *The adaptive web*, 263-290.
- Brusilovsky, P., & Peylo, C. (2003). Adaptive and intelligent web-based educational systems. *International Journal of Artificial Intelligence in Education (IJAIED)*, 13, 159-172.
- Buliva, N. (2019). *Reaching Sub-Saharan Africa Learners Innovatively: The Role of Mobile Technologies. A Narrative Review of Literature*. Paper presented at the Society for Information Technology & Teacher Education International Conference.
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, 65(3), 245-281.
- Car, L. T., Soong, A., Kyaw, B. M., Chua, K. L., Low-Beer, N., & Majeed, A. (2019). Health professions digital education on clinical practice guidelines: a systematic review by Digital Health Education collaboration. *BMC medicine*, 17(1), 139.
- Carenini, G., Conati, C., Hoque, E., Steichen, B., Toker, D., & Enns, J. (2014). *Highlighting interventions and user differences: informing adaptive information visualization support*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.
- Caruana, R., Lawrence, S., & Giles, C. L. (2001). *Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping*. Paper presented at the Advances in neural information processing systems.
- Cen, H., Koedinger, K., & Junker, B. (2006). *Learning factors analysis—a general method for cognitive model evaluation and improvement*. Paper presented at the International Conference on Intelligent Tutoring Systems.
- Cen, H., Koedinger, K., & Junker, B. (2008). *Comparing two IRT models for conjunctive skills*. Paper presented at the International Conference on Intelligent Tutoring Systems.
- Chaiklin, S. (2003). The zone of proximal development in Vygotsky's analysis of learning and instruction. *Vygotsky's educational theory in cultural context*, 1, 39-64.
- Chaudhury, S., Arlington, L., Brenan, S., Kairuki, A. K., Meda, A. R., Isangula, K. G., . . . Msemo, G. (2016). Cost analysis of large-scale implementation of the 'Helping Babies Breathe' newborn resuscitation-training program in Tanzania. *BMC health services research*, 16(1), 681.
- Cheng, A., Lang, T., Starr, S., Pusic, M., & Cook, D. A. (2014). Technology-enhanced simulation and pediatric education: a meta-analysis. *Pediatrics*, 133(5), e1313-e1323.
- CHESS/NIATx. (2019). The Nominal Group Technique (NGT) is designed to promote group participation in the decision-making process. *Nominal Group Technique*. Retrieved from <https://niatx.net/Content/ContentPage.aspx?NID=147>
- Chi, M., Koedinger, K. R., Gordon, G. J., Jordon, P., & VanLahn, K. (2011). Instructional factors analysis: A cognitive model for multiple instructional interventions.
- Cho, K. K., Marjadi, B., Langendyk, V., & Hu, W. (2017). The self-regulated learning of medical students in the clinical environment – a scoping review. *BMC Med Educ*, 17(1), 112. doi:10.1186/s12909-017-0956-6
- Chounta, I.-A., Albacete, P., Jordan, P., Katz, S., & McLaren, B. M. (2017). *The "Grey Area": a computational approach to model the zone of proximal development*. Paper presented at the European Conference on Technology Enhanced Learning.
- Cilliers, P. (2000). Knowledge, complexity, and understanding. *Emergence, A Journal of Complexity Issues in Organizations and Management*, 2(4), 7-13.
- Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital games, design, and learning: A systematic review and meta-analysis. *Review of Educational Research*, 86(1), 79-122.

- Conn, K. M. (2017). Identifying effective education interventions in sub-Saharan Africa: A meta-analysis of impact evaluations. *Review of Educational Research*, 87(5), 863-898.
- Cook, D. A., Hamstra, S. J., Brydges, R., Zendejas, B., Szostek, J. H., Wang, A. T., . . . Hatala, R. (2013). Comparative effectiveness of instructional design features in simulation-based education: systematic review and meta-analysis. *Medical teacher*, 35(1), e867-e898.
- Cook, D. A., Hatala, R., Brydges, R., Zendejas, B., Szostek, J. H., Wang, A. T., . . . Hamstra, S. J. (2011). Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. *Jama*, 306(9), 978-988.
- Cook, D. A., & Reed, D. A. (2015). Appraising the Quality of Medical Education Research Methods: The Medical Education Research Study Quality Instrument and the Newcastle–Ottawa Scale-Education. *Academic Medicine*, 90(8), 1067-1076.
- Cook, N. R. (2008). Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve. *Clinical chemistry*, 54(1), 17-23.
- Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction*, 4(4), 253-278.
- Couper, I., Ray, S., Blaauw, D., Ng'wena, G., Muchiri, L., Oyungu, E., . . . Fonn, S. (2018). Curriculum and training needs of mid-level health workers in Africa: a situational review from Kenya, Nigeria, South Africa and Uganda. *BMC health services research*, 18(1), 553. doi:10.1186/s12913-018-3362-9
- Crisp, N., Gawanas, B., & Sharp, I. (2008). Training the health workforce: scaling up, saving lives. *The Lancet*, 371(9613), 689-691. doi:10.1016/S0140-6736(08)60309-8
- Daktari Media Africa LLC. (2019). Daktari Online. Retrieved from <https://www.daktarionline.co.ke/index.php>
- Dale, H. S. (1991). *Learning theories: an educational perspective*: Macmillan.
- Dawson, S., Joksimovic, S., Poquet, O., & Siemens, G. (2019). *Increasing the Impact of Learning Analytics*. Paper presented at the Proceedings of the 9th International Conference on Learning Analytics & Knowledge.
- de Bruin, A. B., & van Merriënboer, J. J. (2017). Bridging cognitive load and self-regulated learning research: A complementary approach to contemporary issues in educational research. *Learning and Instruction*, 51, 1-9.
- de la Fuente-Arias, J. (2017). Theory of Self- vs. Externally-Regulated Learning(TM): Fundamentals, Evidence, and Applicability. *Frontiers in psychology*, 8, 1675-1675. doi:10.3389/fpsyg.2017.01675
- Dearing, J. W. (2009). Applying diffusion of innovation theory to intervention development. *Research on social work practice*.
- Deci, E. L., & Ryan, R. M. (2011). Self-determination theory. *Handbook of theories of social psychology*, 1(2011), 416-433.
- Delaval, M., Michinov, N., Le Bohec, O., & Le Hénaff, B. (2017). How can students' academic performance in statistics be improved? Testing the influence of social and temporal-self comparison feedback in a web-based training environment. *Interactive Learning Environments*, 25(1), 35-47.
- Deonovic, B., Yudelson, M., Bolsinova, M., Attali, M., & Maris, G. (2018). Learning meets assessment. *Behaviormetrika*, 45(2), 457-474.
- Deutskens, E., De Ruyter, K., Wetzels, M., & Oosterveld, P. (2004). Response rate and response quality of internet-based surveys: An experimental study. *Marketing letters*, 15(1), 21-36.
- Dewey, J. (1998). *Experience and education*: Kappa Delta Pi.
- dit Dariel, O. J. P., Raby, T., Ravaut, F., & Rothan-Tondeur, M. (2013). Developing the Serious Games potential in nursing education. *Nurse education today*, 33(12), 1569-1575.
- Dror, I. E. (2008). Technology enhanced learning: The good, the bad, and the ugly. *Pragmatics & Cognition*, 16(2), 215-223.

- Drummond, D., Hadchouel, A., & Tesnière, A. (2017). Serious games for health: three steps forwards. *Advances in Simulation*, 2(1), 3. doi:10.1186/s41077-017-0036-3
- Dueñas, M., Salazar, A., Sánchez, M., De Sola, H., Ojeda, B., & Failde, I. (2018). Relationship between using clinical practice guidelines for pain treatment and physicians' training and attitudes toward patients and the effects on patient care. *Pain Practice*, 18(1), 38-47.
- Duffy, M. C., & Azevedo, R. (2015). Motivation matters: Interactions between achievement goals and agent scaffolding for self-regulated learning within an intelligent tutoring system. *Computers in Human Behavior*, 52, 338-348.
- Dumville, J. C., Torgerson, D. J., & Hewitt, C. E. (2006). Reporting attrition in randomised controlled trials. *BMJ (Clinical research ed.)*, 332(7547), 969-971. doi:10.1136/bmj.332.7547.969
- Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British journal of psychology*, 105(3), 399-412.
- Edgcombe, H., Paton, C., & English, M. (2016). Enhancing emergency care in low-income countries using mobile technology-based training tools. *Arch Dis Child*. doi:10.1136/archdischild-2016-310875
- Effeney, G., Carroll, A., & Bahr, N. (2013). Self-Regulated Learning: Key strategies and their sources in a sample of adolescent males. *Australian Journal of Educational & Developmental Psychology*, 13.
- Ekanadham, C., & Karklin, Y. (2017). T-skirt: Online estimation of student proficiency in an adaptive learning system. *arXiv preprint arXiv:1702.04282*.
- Ellis P.D. (2009). Thresholds for interpreting effect sizes. Retrieved from http://www.polyu.edu.hk/mm/effectsizefaqs/thresholds_for_interpreting_effect_sizes2.html
- Ertmer, P. A., & Newby, T. J. (2013). Behaviorism, cognitivism, constructivism: Comparing critical features from an instructional design perspective. *Performance Improvement Quarterly*, 26(2), 43-71.
- Etikan, I., Alkassim, R., & Abubakar, S. (2016). Comparison of snowball sampling and sequential sampling technique. *Biometrics and Biostatistics International Journal*, 3(1), 55.
- Fabijan, A., Dmitriev, P., McFarland, C., Vermeer, L., Holmström Olsson, H., & Bosch, J. (2018). Experimentation growth: Evolving trustworthy A/B testing capabilities in online software companies. *Journal of Software: Evolution and Process*, 30(12), e2113.
- Fay, M. P., & Proschan, M. A. (2010). Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules. *Statistics surveys*, 4, 1.
- Felder, R. M., & Silverman, L. K. (1988). Learning and teaching styles in engineering education. *Engineering education*, 78(7), 674-681.
- Fernández-Toro, M., & Furnborough, C. (2014). Feedback on feedback: Eliciting learners' responses to written feedback through student-generated screencasts. *Educational Media International*, 51(1), 35-48.
- Feyzi-Behnagh, R., Azevedo, R., Legowski, E., Reitmeyer, K., Tseytlin, E., & Crowley, R. S. (2014). Metacognitive scaffolds improve self-judgments of accuracy in a medical intelligent tutoring system. *Instructional science*, 42(2), 159-181.
- Fitzgerald, K. (1997). Instructional methods: Selection, use, and evaluation. *Nurse as educator: Principles of teaching and learning*. Sudbury, MA: Jones and Bartlett, 26186.
- Fitzmaurice, G. M., & Ravichandran, C. (2008). A primer in longitudinal data analysis. *Circulation*, 118(19), 2005-2010.
- Fontaine, G., Cossette, S., Maheu-Cadotte, M.-A., Mailhot, T., Deschênes, M.-F., Mathieu-Dupuis, G., . . . Dubé, V. (2019). Efficacy of adaptive e-learning for health professionals and students: a systematic review and meta-analysis. *BMJ open*, 9(8), e025252.
- Fournier, J. F. (1996). *Information Overload and Technology Education*. Paper presented at the Society for Information Technology & Teacher Education International Conference.

- Frehywot, S., Vovides, Y., Talib, Z., Mikhail, N., Ross, H., Wohltjen, H., . . . Scott, J. (2013). E-learning in medical education in resource constrained low-and middle-income countries. *Human resources for health*, 11(1), 4.
- Frenk, J., Chen, L., Bhutta, Z. A., Cohen, J., Crisp, N., Evans, T., . . . Zurayk, H. (2010). Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. *The Lancet*, 376(9756), 1923-1958. doi:10.1016/S0140-6736(10)61854-5
- Gal, Y., & Ghahramani, Z. (2016). *A theoretically grounded application of dropout in recurrent neural networks*. Paper presented at the Advances in neural information processing systems.
- Galyardt, A., & Goldin, I. (2014). *Recent-performance factors analysis*. Paper presented at the Educational Data Mining 2014.
- Gatignon, H. (2010). Confirmatory factor analysis *Statistical analysis of management data* (pp. 59-122): Springer.
- Gee, J. P. (2003). What video games have to teach us about learning and literacy. *Computers in Entertainment (CIE)*, 1(1), 20-20.
- Gentry, S., L'Estrade-Ehrstrom, B., Gauthier, A., Alvarez, J., Wortley, D., van Rijswijk, J., . . . Nikolaou, C. K. (2016). Serious Gaming and Gamification interventions for health professional education. *The Cochrane Library*.
- Gentry, S. V., Gauthier, A., L'Estrade Ehrstrom, B., Wortley, D., Lilienthal, A., Tudor Car, L., . . . Car, J. (2019). Serious Gaming and Gamification Education in Health Professions: Systematic Review. *Journal of medical Internet research*, 21(3), e12994-e12994. doi:10.2196/12994
- George E.H. (1991). Constructivist Learning Theory. *The Museum and the Needs of People*. Retrieved from <https://www.exploratorium.edu/education/ifi/constructivist-learning>
- George, P. P., Zhabenko, O., Kyaw, B. M., Antoniou, P., Posadzki, P., Saxena, N., . . . Lockwood, C. (2019). Online digital education for postregistration training of medical doctors: systematic review by the Digital Health Education Collaboration. *Journal of medical Internet research*, 21(2), e13269.
- Gibbons, J. D., & Chakraborti, S. (1991). Comparisons of the Mann-Whitney, Student's t, and alternate t tests for means of normal distributions. *The Journal of Experimental Education*, 59(3), 258-267.
- Glass, G. V., Smith, M. L., & McGaw, B. (1981). *Meta-analysis in social research*: Sage Publications, Incorporated.
- Gong, Y., Beck, J. E., & Heffernan, N. T. (2011). How to construct more accurate student models: Comparing and optimizing knowledge tracing and performance factor analysis. *International Journal of Artificial Intelligence in Education*, 21(1-2), 27-46.
- Graafland, M., Dankbaar, M., Mert, A., Lagro, J., De Wit-Zuurendonk, L., Schuit, S., . . . Schijven, M. (2014). How to systematically assess serious games applied to health care. *JMIR Serious Games*, 2(2).
- Grady, C., Cummings, S. R., Rowbotham, M. C., McConnell, M. V., Ashley, E. A., & Kang, G. (2017). Informed Consent. *New England Journal of Medicine*, 376(9), 856-867. doi:10.1056/NEJMr1603773
- Graf, S., Viola, S. R., Leo, T., & Kinshuk. (2007). In-depth analysis of the Felder-Silverman learning style dimensions. *Journal of Research on Technology in Education*, 40(1), 79-93.
- Greenhalgh, T. (2001). Computer assisted learning in undergraduate medical education. *Bmj*, 322(7277), 40-44.
- Greenwood, J. (1995). Two strategies for promoting clinical competence in pre-registration nursing students. *Nurse education today*, 15(3), 184-189.
- Gross, S., Mokbel, B., Hammer, B., & Pinkwart, N. (2015). Learning feedback in intelligent tutoring systems. *KI-Künstliche Intelligenz*, 29(4), 413-418.
- Gutierrez, F., & Atkinson, J. (2011). Adaptive feedback selection for intelligent tutoring systems. *Expert Systems with Applications*, 38(5), 6146-6152.

- Hategekimana, C., Shoveller, J., Tuyisenge, L., Kenyon, C., Cechetto, D. F., & Lynd, L. D. (2016). Correlates of performance of healthcare workers in emergency, triage, assessment and treatment plus admission care (ETAT+) course in Rwanda: context matters. *PloS one*, 11(3), e0152882.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.
- Heather, C. (2009). Zone of proximal development. *Learn NC*. retrieved February, 2, 2012.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780.
- Holmes, W., Anastopoulou, S., Schaumburg, H., & Mavrikis, M. (2018). Technology-enhanced Personalised Learning: Untangling the Evidence.
- Holmes, W., Mavrikis, M., Hansen, A., & Grawemeyer, B. (2015). *Purpose and Level of Feedback in an Exploratory Learning Environment for Fractions*, Cham.
- Hu, L. t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55. doi:10.1080/10705519909540118
- Ian Walker. (2008). Null hypothesis testing and effect sizes. *Statistics for Psychology*. Retrieved from <http://staff.bath.ac.uk/pssiw/stats2/page2/page14/page14.html>
- Imlig-Iten, N., & Petko, D. (2018). Comparing Serious Games and Educational Simulations: Effects on Enjoyment, Deep Thinking, Interest and Cognitive Learning Gains. *Simulation & Gaming*, 1046878118779088.
- IntraHealth (PRIME II Project) , & JHPIEGO. (2002). Transfer of Learning: A Guide for Strengthening the Performance of Health Care Workers. Retrieved from <https://www.who.int/workforcealliance/knowledge/toolkit/52/en/>
- Irimu, G., Wamae, A., Wasunna, A., Were, F., Ntoburi, S., Opiyo, N., . . . English, M. (2008). Developing and Introducing Evidence Based Clinical Practice Guidelines for Serious Illness in Kenya. *Arch Dis Child*, 93(9), 799-804. doi:10.1136/adc.2007.126508
- Jansen, R. S., Van Leeuwen, A., Janssen, J., Kester, L., & Kalz, M. (2017). Validation of the self-regulated online learning questionnaire. *Journal of computing in higher education*, 29(1), 6-27.
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational researcher*, 33(7), 14-26.
- Kaji, D. A., Zech, J. R., Kim, J. S., Cho, S. K., Dangayach, N. S., Costa, A. B., & Oermann, E. K. (2019). An attention based deep learning model of clinical events in the intensive care unit. *PloS one*, 14(2), e0211057.
- Kaufmann, E., Cappé, O., & Garivier, A. (2014). *On the complexity of A/B testing*. Paper presented at the Conference on Learning Theory.
- Keçeci, A. (2017). Self-regulated learning in nursing: A study from a health education course. *Journal of Human Sciences*, 14(4), 3830-3842.
- Khajah, M., Lindsey, R. V., & Mozer, M. C. (2016). How deep is knowledge tracing? *arXiv preprint arXiv:1604.02416*.
- Kiili, K., De Freitas, S., Arnab, S., & Lainema, T. (2012). The design principles for flow experience in educational games. *Procedia Computer Science*, 15, 78-91.
- Kinfu, Y., Dal Poz, M. R., Mercer, H., & Evans, D. B. (2009). The health worker shortage in Africa: are enough physicians and nurses being trained? : SciELO Public Health.
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Kirkpatrick, D. (1996). Revisiting Kirkpatrick's four-level model. *Training and Development*, 50(1), 54-59.
- Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2017). Self-regulated learning strategies predict learner behavior and goal attainment in Massive Open Online Courses. *Computers & Education*, 104, 18-33.

- Kohavi, R. (1995). *A study of cross-validation and bootstrap for accuracy estimation and model selection*. Paper presented at the Ijcai.
- Kohavi, R., & Longbotham, R. (2017). Online controlled experiments and a/b testing. *Encyclopedia of machine learning and data mining*, 922-929.
- Kolb, A. Y., & Kolb, D. A. (2005). Learning styles and learning spaces: Enhancing experiential learning in higher education. *Academy of management learning & education*, 4(2), 193-212.
- Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: a meta-analytic review. *Review of Educational Research*, 86(1), 42-78.
- Kuljis, J., & Liu, F. (2005). A Comparison of Learning Style Theories on the Suitability for elearning. *Web Technologies, Applications, and Services*, 2005, 191-197.
- Kumar, P., Paton, C., & Kirigia, D. (2016). I've got 99 problems but a phone ain't one: Electronic and mobile health in low and middle income countries. *Arch Dis Child*, 101(10), 974-979.
- Lajoie, S., Poitras, E., Lajoie, S., Wiese, E. S., Long, Y., Aleven, V., . . . Goldberg, B. (2014). Enhancing Self-Regulated Learning through Metacognitively-Aware Intelligent Tutoring Systems.
- Lajoie, S. P., & Azevedo, R. (2006). Teaching and learning in technology-rich environments.
- Lam, C. F., DeRue, D. S., Karam, E. P., & Hollenbeck, J. R. (2011). The impact of feedback frequency on learning and task performance: Challenging the "more is better" assumption. *Organizational Behavior and Human Decision Processes*, 116(2), 217-228.
- Landers, R. N. (2014). Developing a theory of gamified learning: Linking serious games and gamification of learning. *Simulation & Gaming*, 45(6), 752-768.
- Lavavej, S. T. (2013). rand() Considered Harmful. Retrieved from <https://channel9.msdn.com/Events/GoingNative/2013/rand-Considered-Harmful>
- Lawn, J. E., Blencowe, H., Oza, S., You, D., Lee, A. C., Waiswa, P., . . . Christian, P. (2014). Every Newborn: progress, priorities, and potential beyond survival. *The Lancet*, 384(9938), 189-205.
- Lee, J., & Park, O. (2008). Adaptive instructional systems. In J. M. Spector, M. D. Merrill, J. J. G. van Merriënboer, & M. Driscoll (Eds.), *Handbook of research on educational communications and technology* (3rd ed.). New York, NY: Taylor & Francis.
- Lee, J. I., & Brunskill, E. (2012). *The Impact on Individualizing Student Models on Necessary Practice Opportunities*. Paper presented at the 5th International Conference on Educational Data Mining (EDM 2012), Chania, Greece.
- Lee, M. K. (2015). Effects of mobile phone-based app learning compared to computer-based web learning on nursing students: pilot randomized controlled trial. *Healthcare informatics research*, 21(2), 125-133.
- Lenhard, W., & Lenhard, A. (2016). Transformation of the effect sizes d, r, f, Odds Ratio, and Common Language Effect Size (CLES). *Computation of Effect Sizes*. Retrieved from https://www.psychometrica.de/effect_size.html
- Leszczyński, P., Charuta, A., Łaziuk, B., Gałązkowski, R., Wejnarski, A., Roszak, M., & Kołodziejczak, B. (2018). Multimedia and interactivity in distance learning of resuscitation guidelines: a randomised controlled trial. *Interactive Learning Environments*, 26(2), 151-162.
- Levinson, S. E., Rabiner, L. R., & Sondhi, M. M. (1983). An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition. *Bell System Technical Journal*, 62(4), 1035-1074.
- Lipton, Z. C., Kale, D. C., Elkan, C., & Wetzal, R. (2015). Learning to diagnose with LSTM recurrent neural networks. *arXiv preprint arXiv:1511.03677*.
- Littlejohn, A., & Milligan, C. (2015). Designing MOOCs for professional learners: Tools and patterns to encourage self-regulated learning. *eLearning Papers*, 42.
- Lo. (2013). Differences between L1 and L2 as Loss Function and Regularization. Retrieved from <http://www.chioka.in/differences-between-l1-and-l2-as-loss-function-and-regularization/>
- Long, L. A., Pariyo, G., & Kallander, K. (2018). Digital Technologies for Health Workforce Development in Low- and Middle-Income Countries: A Scoping Review. *Glob Health Sci Pract*, 6(Suppl 1), S41-s48. doi:10.9745/ghsp-d-18-00167

- López, X., Valenzuela, J., Nussbaum, M., & Tsai, C.-C. (2015). Some recommendations for the reporting of quantitative studies. *Computers & Education*, 91, 106-110. doi:<https://doi.org/10.1016/j.compedu.2015.09.010>
- Luckin, R. (2009). *Learning, context and the role of technology*: Institute of Education, University of London.
- Luckin, R. (2010). *Re-designing learning contexts: Technology-rich, learner-centred ecologies*: Routledge.
- Luckin, R., & Du Boulay, B. (1999). Ecolab: The development and evaluation of a Vygotskian design framework. *International Journal of Artificial Intelligence in Education*, 10(2), 198-220.
- Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). Intelligence unleashed: An argument for AI in education.
- Lyytinen, K., & Damsgaard, J. (2001). *What's wrong with the diffusion of innovation theory?* Paper presented at the Working Conference on Diffusing Software Product and Process Innovations.
- Ma, W. (2017). *Intelligent Tutoring Systems and Learning Outcomes: Two Systematic Reviews*. Education: Faculty of Education.
- Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology*, 106(4), 901.
- Maaten, L. v. d., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of machine learning research*, 9(Nov), 2579-2605.
- MacKinnon, D. P. (2011). Integrating mediators and moderators in research design. *Research on social work practice*, 21(6), 675-681.
- Magoulas, G. D., Chen, S. Y., & Papanikolaou, K. A. (2003). *Integrating Layered and Heuristic Evaluation for Adaptive Learning Environments*. Paper presented at the 9th International Conference on User Modeling UM2003, Pittsburgh.
- Maheu-Cadotte, M.-A., Cossette, S., Dubé, V., Fontaine, G., Mailhot, T., Lavoie, P., . . . Mathieu-Dupuis, G. (2018). Effectiveness of serious games and impact of design elements on engagement and educational outcomes in healthcare professionals and students: a systematic review and meta-analysis protocol. *BMJ open*, 8(3), e019871.
- Maibach, E. W., Schieber, R. A., & Carroll, M. F. (1996). Self-efficacy in pediatric resuscitation: implications for education and performance. *Pediatrics*, 97(1), 94-99.
- Marx, J. D., & Cummings, K. (2007). Normalized change. *American Journal of Physics*, 75(1), 87-91.
- Maybin, J., Mercer, N., & Stierer, B. (1992). Scaffolding learning in the classroom. *Thinking voices: The work of the national oracy project*, 186-195.
- McMillan, S. S., King, M., & Tully, M. P. (2016). How to use the nominal group and Delphi techniques. *International Journal of Clinical Pharmacy*, 38(3), 655-662. doi:10.1007/s11096-016-0257-x
- Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., . . . Gebru, T. (2019). *Model cards for model reporting*. Paper presented at the Proceedings of the Conference on Fairness, Accountability, and Transparency.
- Mitrovic, A., Mayo, M., Suraweera, P., & Martin, B. (2001). Constraint-based tutors: a success story. *Engineering of intelligent systems*, 931-940.
- Mödrischer, F., Garcia-Barrios, V. M., & Gütl, C. (2004). The Past, the Present and the Future of adaptive E-Learning. *Proceedings of ICL 2004*.
- Mohan, D., Schell, J., & Angus, D. (2016). Not Thinking Clearly? Play a Game, Seriously! *Jama*, 316(18), 1867-1868.
- Moher, D., Schulz, K. F., Altman, D., & Group, C. (2001). The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *Jama*, 285(15), 1987-1991.
- Molloy, E., Boud, D., & Henderson, M. (2020). Developing a learning-centred framework for feedback literacy. *Assessment & Evaluation in Higher Education*, 45(4), 527-540.

- Moore, S., Tassé, A.-M., Thorogood, A., Winship, I., Zawati, M. n., & Doerr, M. (2017). Consent Processes for Mobile App Mediated Research: Systematic Review. *JMIR mHealth and uHealth*, 5(8), e126. doi:10.2196/mhealth.7014
- Morris, S. B. (2008). Estimating Effect Sizes From Pretest-Posttest-Control Group Designs. *Organizational Research Methods*, 11(2), 364-386. doi:10.1177/1094428106291059
- Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological methods*, 7(1), 105.
- Mousavinasab, E., Zarifsanaiey, N., R. Niakan Kalhori, S., Rakhshan, M., Keikha, L., & Ghazi Saeedi, M. (2018). Intelligent tutoring systems: a systematic review of characteristics, applications, and evaluation methods. *Interactive Learning Environments*, 1-22. doi:10.1080/10494820.2018.1558257
- Mundell, W. C., Kennedy, C. C., Szostek, J. H., & Cook, D. A. (2013). Simulation technology for resuscitation training: a systematic review and meta-analysis. *Resuscitation*, 84(9), 1174-1183.
- Muthén, B. (2004). Latent variable analysis. *The Sage handbook of quantitative methodology for the social sciences*, 345(368), 106-109.
- Muthén, B. O. (2001). Latent variable mixture modeling *New developments and techniques in structural equation modeling* (pp. 21-54): Psychology Press.
- Nair, V., & Hinton, G. E. (2010). *Rectified linear units improve restricted boltzmann machines*. Paper presented at the Proceedings of the 27th international conference on machine learning (ICML-10).
- Nakamura, J., & Csikszentmihalyi, M. (2014). The concept of flow *Flow and the foundations of positive psychology* (pp. 239-263): Springer.
- Narkhede, S. (2018). Understanding AUC-ROC Curve. *Towards Data Science*, 26.
- Nicoll, P., MacRury, S., van Woerden, H. C., & Smyth, K. (2018). Evaluation of Technology-Enhanced Learning Programs for Health Care Professionals: Systematic Review. *Journal of medical Internet research*, 20(4), e131. doi:10.2196/jmir.9085
- Niemi, H., Nevgi, A., & Virtanen, P. i. (2003). Towards self-regulation in web-based learning. *Journal of Educational Media*, 28(1), 49-71.
- Nolan, T., Angos, P., Cunha, A. J., Muhe, L., Qazi, S., Simoes, E. A., . . . Pierce, N. F. (2001). Quality of hospital care for seriously ill children in less-developed countries. *Lancet*, 357(9250), 106-110. doi:10.1016/s0140-6736(00)03542-x
- North, N., Shung-King, M., & Coetzee, M. (2019). The children's nursing workforce in Kenya, Malawi, Uganda, South Africa and Zambia: generating an initial indication of the extent of the workforce and training activity. *Human resources for health*, 17(1), 30.
- Nye, B. D. (2015). Intelligent tutoring systems by and for the developing world: A review of trends and approaches for educational technology in a global context. *International Journal of Artificial Intelligence in Education*, 25(2), 177-203.
- Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(4), 535-569. doi:10.1080/10705510701575396
- O'Brien, H. L., & Toms, E. G. (2008). What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the American society for Information Science and Technology*, 59(6), 938-955.
- Okuda, Y., Bryson, E. O., DeMaria, S., Jr., Jacobson, L., Quinones, J., Shen, B., & Levine, A. I. (2009). The utility of simulation in medical education: what is the evidence? *Mt Sinai J Med*, 76(4), 330-343. doi:10.1002/msj.20127
- Olenka, B. (2009, January 2011). Scaffolding. Retrieved from <http://www.educ.ualberta.ca/staff/olenka.bilash/best%20of%20bilash/scaffolding.html>
- Olivera-Aguilar, M., & Rikoon, S. H. (2018). Assessing measurement invariance in multiple-group latent profile analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(3), 439-452.

- Opiyo, N., & English, M. (2015). In-service training for health professionals to improve care of seriously ill newborns and children in low-income countries. *The Cochrane database of systematic reviews*(5), 1.
- Organization, W. H. (2005). Emergency triage assessment and treatment (ETAT) course. *World Health Organization: Geneva, Switzerland*.
- Oxford University. (2015). Life-saving Instruction for Emergencies (LIFE). Retrieved from <https://oxlifeproject.org/>
- Özhan, Ş. Ç., & Kocadere, S. A. (2020). The effects of flow, emotional engagement, and motivation on success in a gamified online learning environment. *Journal of Educational Computing Research*, 57(8), 2006-2031.
- Panadero, E. (2017). A review of self-regulated learning: six models and four directions for research. *Frontiers in psychology*, 8, 422.
- Pardos, Z. A., & Heffernan, N. T. (2010). *Modeling individualization in a bayesian networks implementation of knowledge tracing*. Paper presented at the International Conference on User Modeling, Adaptation, and Personalization.
- Pastor, D. A., Barron, K. E., Miller, B., & Davis, S. L. (2007). A latent profile analysis of college students' achievement goal orientation. *Contemporary Educational Psychology*, 32(1), 8-47.
- Pavlik Jr, P. I., Cen, H., & Koedinger, K. R. (2009). Performance Factors Analysis--A New Alternative to Knowledge Tracing. *Online Submission*.
- Pawson, R., & Tilley, N. (1997). Realist evaluation. *Changes*.
- Peddycord-Liu, Z., Harred, R., Karamarkovich, S., Barnes, T., Lynch, C., & Rutherford, T. (2018). *Learning Curve Analysis in a Large-Scale, Drill-and-Practice Serious Math Game: Where Is Learning Support Needed?* Paper presented at the International Conference on Artificial Intelligence in Education.
- Pekrun, R., Frenzel, A. C., Goetz, T., & Perry, R. P. (2007). The control-value theory of achievement emotions: An integrative approach to emotions in education *Emotion in education* (pp. 13-36): Elsevier.
- Pelánek, R. (2016). Applications of the Elo rating system in adaptive educational systems. *Computers & Education*, 98, 169-179.
- Peters, D. H., Adam, T., Alonge, O., Agyepong, I. A., & Tran, N. (2013). Implementation research: what it is and how to do it. *Bmj*, 347, f6753.
- Phan, H. P., Ngu, B. H., & Yeung, A. S. (2017). Achieving optimal best: Instructional efficiency and the use of cognitive load theory in mathematical problem solving. *Educational Psychology Review*, 29(4), 667-692.
- Pickering, J. D. (2017). Developing the evidence-base to support the integration of technology-enhanced learning in healthcare education. *Medical Science Educator*, 27(4), 903-905.
- Pickering, J. D., & Joynes, V. C. (2016). A holistic model for evaluating the impact of individual technology-enhanced learning resources. *Med Teach*, 38(12), 1242-1247. doi:10.1080/0142159x.2016.1210112
- Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., & Sohl-Dickstein, J. (2015). *Deep knowledge tracing*. Paper presented at the Advances in neural information processing systems.
- Piech, C., Huang, J., Nguyen, A., Phulsuksombati, M., Sahami, M., & Guibas, L. (2015). Learning program embeddings to propagate feedback on student code. *arXiv preprint arXiv:1505.05969*.
- Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning *Handbook of self-regulation* (pp. 451-502): Elsevier.
- Polit, D. F., & Beck, C. T. (2010). Generalization in quantitative and qualitative research: Myths and strategies. *International journal of nursing studies*, 47(11), 1451-1458.
- Polly, D., Allman, B., Casto, A., & Norwood, J. (2017). Sociocultural Perspectives of Learning. *Foundations of Learning and Instructional Design Technology*.

- Popescu, E. (2010). Adaptation provisioning with respect to learning styles in a Web-based educational system: an experimental study. *Journal of Computer Assisted Learning*, 26(4), 243-257.
- Porayska-Pomsta, K. (2016). AI as a methodology for supporting educational praxis and teacher metacognition. *International Journal of Artificial Intelligence in Education*, 26(2), 679-700.
- Porter, S. R., Whitcomb, M. E., & Weitzer, W. H. (2004). Multiple surveys of students and survey fatigue. *New Directions for Institutional Research*, 2004(121), 63-73.
- Poushter, J. (2016). Smartphone ownership and internet usage continues to climb in emerging economies. *Pew Research Center*, 22, 1-44.
- Prensky, M. (2001). The Games Generation: How Learners have changed *Digital game-based learning* (Vol. 1, pp. 21-21): McGraw-Hill.
- Pusic, M. V., Boutis, K., Hatala, R., & Cook, D. A. (2015). Learning curves in health professions education. *Academic Medicine*, 90(8), 1034-1042.
- QuantStart. (2019). Random Number Generation via Linear Congruential Generators in C++. Retrieved from <https://www.quantstart.com/articles/Random-Number-Generation-via-Linear-Congruential-Generators-in-C>
- R Core Team. (2013). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>
- Rawlings, A. M., Tapola, A., & Niemivirta, M. (2017). Predictive effects of temperament on motivation. *International Journal of Educational Psychology: IJEP*, 6(2), 148-182.
- Reeves, S., Fletcher, S., McLoughlin, C., Yim, A., & Patel, K. D. (2017). Interprofessional online learning for primary healthcare: findings from a scoping review. *BMJ open*, 7(8), e016872.
- Reid, A.-M., Brown, J. M., Smith, J. M., Cope, A. C., & Jamieson, S. (2018). Ethical dilemmas and reflexivity in qualitative research. *Perspectives on medical education*, 7(2), 69-75.
- Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological bulletin*, 138(2), 353.
- Rn, P. A. G. (2000). 'Feedback. I listened, reflected and utilized': Third year nursing students' perceptions and use of feedback in the clinical setting. *International Journal of Nursing Practice*, 6(5), 247-252.
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC bioinformatics*, 12(1), 77.
- Rovers, S. F., Clarebout, G., Savelberg, H. H., de Bruin, A. B., & van Merriënboer, J. J. (2019). Granularity matters: comparing different ways of measuring self-regulated learning. *Metacognition and Learning*, 14(1), 1-19.
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist*, 55(1), 68.
- Saad, A., & Zaghloul, A. R. M. (2002). *A Constructivist Knowledge Base Architecture for Adaptive Intelligent Tutoring Systems*. Paper presented at the International Conference on Intelligent Tutoring Systems.
- Sailer, M., & Homner, L. (2019). *The gamification of learning: A meta-analysis*: Springer.
- Säljö, R. (2010). Digital tools and challenges to institutional traditions of learning: technologies, social memory and the performative nature of learning. *Journal of Computer Assisted Learning*, 26(1), 53-64.
- Sangrà, A., Vlachopoulos, D., & Cabrera, N. (2012). Building an inclusive definition of e-learning: An approach to the conceptual framework. *The International Review of Research in Open and Distributed Learning*, 13(2), 145-159.
- Sapountzi A., Bhulai, S., Cornelisz, I., & van Klaveren, C. (2018). *Dynamic Models for Knowledge Tracing & Prediction of Future Performance*. Paper presented at the Data Analytics 2018: The Seventh International Conference on Data Analytics, Athens, Greece.

- Sage-Rockoff, A., Ciardiello, A. V., & Schubert, F. D. (2019). Low-Fidelity, In-Situ Pediatric Resuscitation Simulation Improves RN Competence and Self-Efficacy. *Journal of Emergency Nursing*.
- Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., . . . Jinks, C. (2018). Saturation in qualitative research: exploring its conceptualization and operationalization. *Quality & quantity*, 52(4), 1893-1907.
- Schneider D.K. (2012, 9th January 2018). Mixed methods. Retrieved from http://edutechwiki.unige.ch/en/Mixed_methods
- Schultz, B. B. (1985). Levene's test for relative variation. *Systematic Zoology*, 34(4), 449-456.
- Schunk, D. H. (2005). Self-regulated learning: The educational legacy of Paul R. Pintrich. *Educational Psychologist*, 40(2), 85-94.
- Seabold, S., & Perktold, J. (2010). *Statsmodels: Econometric and statistical modeling with python*. Paper presented at the Proceedings of the 9th Python in Science Conference.
- Shah, T. (2017). About Train, Validation and Test Sets in Machine Learning. *Medium*. Retrieved from <https://towardsdatascience.com/train-validation-and-test-sets-72cb40cba9e7>
- Sheik-Ali, S., Edgcombe, H., & Paton, C. (2019). Next-generation Virtual and Augmented Reality in Surgical Education: A Narrative Review. *Surgical technology international*, 35.
- Shenouda, J. E., Davies, B. S., & Haq, I. (2018). The role of the smartphone in the transition from medical student to foundation trainee: a qualitative interview and focus group study. *BMC Med Educ*, 18(1), 175.
- Sherry, J. L. (2004). Flow and media enjoyment. *Communication theory*, 14(4), 328-347.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153-189.
- Silver, L., & Johnson, C. (2018). *Internet Connectivity Seen as Having Positive Impact on Life in Sub-Saharan Africa: But Digital Divides Persist*: Pew Research Center.
- Smith, P. L., & Ragan, T. J. (1999). *Instructional design*: Wiley New York.
- Soloman, B. A., & Felder, R. M. (2005). Index of learning styles questionnaire. *NC State University*. Available online at: <http://www.engr.ncsu.edu/learningstyles/ilsweb.html> (last visited on 14.05.2010), 70.
- Sottolare, R. A., Graesser, A. C., Hu, X., Olney, A., Nye, B., & Sinatra, A. M. (2016). *Design Recommendations for Intelligent Tutoring Systems: Volume 4-Domain Modeling* (Vol. 4): US Army Research Laboratory.
- Spurrier, J. D. (2003). On the null distribution of the Kruskal–Wallis statistic. *Nonparametric Statistics*, 15(6), 685-691.
- Srivastava, N., Mansimov, E., & Salakhudinov, R. (2015). *Unsupervised learning of video representations using lstms*. Paper presented at the International conference on machine learning.
- Stanley, L., Kellermanns, F. W., & Zellweger, T. M. (2017). Latent profile analysis: Understanding family firm profiles. *Family Business Review*, 30(1), 84-102.
- Steenbergen-Hu, S., & Cooper, H. (2013). A meta-analysis of the effectiveness of intelligent tutoring systems on K–12 students' mathematical learning. *Journal of Educational Psychology*, 105(4), 970.
- Steenbergen-Hu, S., & Cooper, H. (2014). A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning. *Journal of Educational Psychology*, 106(2), 331.
- Stuckless, P., Hogan, M., & Kapralos, B. (2014). Virtual simulations and serious games in community health nursing education: a review of the literature *Virtual, Augmented Reality and Serious Games for Healthcare 1* (pp. 145-158): Springer.
- Sutton, R. M., Niles, D., Meaney, P. A., Aplenc, R., French, B., Abella, B. S., . . . Nadkarni, V. (2011). Low-dose, high-frequency CPR training improves skill retention of in-hospital pediatric providers. *Pediatrics*, 128(1), e145-e151.
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction*, 4(4), 295-312.

- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International journal of medical education*, 2, 53.
- Taylor, D. C. M., & Hamdy, H. (2013). Adult learning theories: Implications for learning and teaching in medical education: AMEE Guide No. 83. *Medical teacher*, 35(11), e1561-e1572.
- TensorFlow Team. (2019). TensorFlow 2.0 is now available! Retrieved from <https://blog.tensorflow.org/2019/09/tensorflow-20-is-now-available.html>
- Triantafillou, E., Pomportsis, A., & Demetriadis, S. (2003). The design and the formative evaluation of an adaptive educational system based on cognitive styles. *Computers and Education*, 41(1), 87-104.
- Tuti, T., Winters, N., Muinga, N., Wanyama, C., English, M., & Paton, C. (2019). Evaluation of Adaptive Feedback in a Smartphone-Based Serious Game on Health Care Providers' Knowledge Gain in Neonatal Emergency Care: Protocol for a Randomized Controlled Trial. *JMIR research protocols*, 8(7), e13034.
- Tuyisenge, L., Kyamanywa, P., Van Steirteghem, S., Becker, M., English, M., & Lissauer, T. (2014). Knowledge and skills retention following Emergency Triage, Assessment and Treatment plus Admission course for final year medical students in Rwanda: a longitudinal cohort study. *Arch Dis Child*. doi:10.1136/archdischild-2014-306078
- UNICEF. (2018). *Levels and trends in child mortality: report 2018*. Retrieved from New York:
- Unreal Engine. (2018). Shipping With Android. *Android Game Development*. Retrieved from <https://docs.unrealengine.com/en-us/Platforms/Mobile/Android/ShippingWithAndroid#androidsdkversion>
- Valle, A., Cabanach, R. G., Núñez, J. C., González-Pienda, J., Rodríguez, S., & Piñeiro, I. (2003). Cognitive, motivational, and volitional dimensions of learning: An empirical test of a hypothetical model. *Research in higher education*, 44(5), 557-580.
- van Houten-Schat, M. A., Berkhout, J. J., van Dijk, N., Endedijk, M. D., Jaarsma, A. D. C., & Diemers, A. D. (2018). Self-regulated learning in the clinical context: a systematic review. *Medical education*, 52(10), 1008-1015.
- Van Rossum, G., & Drake, F. L. (2011). *The python language reference manual*: Network Theory Ltd.
- van Schaik, S. M., Plant, J., Diane, S., Tsang, L., & O'Sullivan, P. (2011). Interprofessional Team Training in Pediatric Resuscitation: A Low-Cost, In Situ Simulation Program That Enhances Self-Efficacy Among Participants. *Clinical Pediatrics*, 50(9), 807-815. doi:10.1177/0009922811405518
- Vandewaetere, M., Cornillie, F., Clarebout, G., & Desmet, P. (2013). Adaptivity in Educational Games: Including Player and Gameplay Characteristics. *International Journal of Higher Education*, 2(2), 106-114.
- Vandewaetere, M., Desmet, P., & Clarebout, G. (2011). The contribution of learner characteristics in the development of computer-based adaptive learning environments. *Computers in Human Behavior*, 27(1), 118-130.
- VanLehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education*, 16(3), 227-265.
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197-221.
- VanLehn, K., Jordan, P., & Litman, D. (2007). *Developing pedagogically effective tutorial dialogue tactics: Experiments and a testbed*. Paper presented at the Workshop on Speech and Language Technology in Education.
- Vanslambrouck, S., Zhu, C., Pynoo, B., Lombaerts, K., Tondeur, J., & Scherer, R. (2019). A latent profile analysis of adult students' online self-regulation in blended learning environments. *Computers in Human Behavior*, 99, 126-136.
- Veredas, F. J., Ruiz-Bandera, E., Villa-Estrada, F., Rufino-González, J. F., & Morente, L. (2014). A web-based e-learning application for wound diagnosis and treatment. *Computer methods and programs in biomedicine*, 116(3), 236-248.

- Voerman, L., Meijer, P. C., Korthagen, F. A., & Simons, R. J. (2012). Types and frequencies of feedback interventions in classroom interaction in secondary education. *Teaching and Teacher Education*, 28(8), 1107-1115.
- Wakaba, M., Mbindyo, P., Ochieng, J., Kiriinya, R., Todd, J., Waudu, A., . . . English, M. (2014). The public sector nursing workforce in Kenya: a county-level analysis. *Human resources for health*, 12(1), 6.
- Wall, S. N., Lee, A. C. C., Niermeyer, S., English, M., Keenan, W. J., Carlo, W., . . . Lawn, J. E. (2009). Neonatal resuscitation in low-resource settings: What, who, and how to overcome challenges to scale up? *International journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics*, 107(Suppl 1), S47-S64. doi:10.1016/j.ijgo.2009.07.013
- Wang, R., DeMaria Jr, S., Goldberg, A., & Katz, D. (2016). A systematic review of serious games in training health care professionals. *Simulation in Healthcare*, 11(1), 41-51.
- Wattenberg, Martin and Viégas, Fernanda and Johnson, & Ian. (2016). How to Use t-SNE Effectively. Retrieved from <http://distill.pub/2016/misread-tsne>
- Weiner, B. (1974). *Achievement motivation and attribution theory*: General Learning Press.
- Weinstein, C. E., Acee, T. W., & Jung, J. (2011). Self-regulation and learning strategies. *New directions for teaching and learning*, 2011(126), 45-53.
- WHO. (2011). *Transformative scale up of health professional education: an effort to increase the numbers of health professionals and to strengthen their impact on population health*. Retrieved from
- WHO. (2016). Updated guideline: paediatric emergency triage, assessment and treatment: care of critically-ill children. *World Health Organization*.
- Wijk, H., Ponzer, S., Heikkilä, K., Kihlström, L., & Nordquist, J. (2019). Factors influencing effectiveness in postgraduate medical education—a qualitative study of experiences of the responsible clinical consultants. *BMC Med Educ*, 19(1), 1-9.
- Wilcox, R. R. (1990). Comparing the means of two independent groups. *Biometrical Journal*, 32(7), 771-780.
- Willcox, M., Harrison, H., Asiedu, A., Nelson, A., Gomez, P., & LeFevre, A. (2017). Incremental cost and cost-effectiveness of low-dose, high-frequency training in basic emergency obstetric and newborn care as compared to status quo: part of a cluster-randomized training intervention evaluation in Ghana. *Globalization and health*, 13(1), 88.
- Winne H. P. (2011). A Cognitive and Metacognitive Analysis of Self-Regulated Learning. In D. H. Schunk & B. Zimmerman (Eds.), *Handbook of self-regulation of learning and performance* (pp. pp 15-32). New York: Taylor & Francis.
- Winne H.P. (2011). A Cognitive and Metacognitive Analysis of Self-Regulated Learning. In D. H. Schunk & B. Zimmerman (Eds.), *Handbook of self-regulation of learning and performance* (pp. pp 15-32). New York: Taylor & Francis.
- Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. *Metacognition in educational theory and practice*, 93, 27-30.
- Winne, P. H., Nesbit, J. C., Kumar, V., Hadwin, A. F., Lajoie, S. P., Azevedo, R., & Perry, N. E. (2006). Supporting self-regulated learning with gStudy software: The Learning Kit Project. *Technology Instruction Cognition and Learning*, 3(1/2), 105.
- Winters, N., Langer, L., Nduku, P., Robson, J., O'Donovan, J., Maulik, P., . . . Nagraj, S. (2019). Using mobile technologies to support the training of community health workers in low-income and middle-income countries: mapping the evidence. *BMJ Global Health*, 4(4), e001421.
- Winters, N., Oliver, M., & Langer, L. (2017). Can mobile health training meet the challenge of 'measuring better'? *Comparative Education*, 53(1), 115-131.
- Wong, J., Baars, M., Davis, D., Van Der Zee, T., Houben, G.-J., & Paas, F. (2019). Supporting self-regulated learning in online learning environments and MOOCs: A systematic review. *International Journal of Human-Computer Interaction*, 35(4-5), 356-373.

- Wong, V., Smith, A. J., Hawkins, N. J., Kumar, R. K., Young, N., Kyaw, M., & Velan, G. M. (2015). Adaptive tutorials versus web-based resources in radiology: A mixed methods comparison of efficacy and student engagement. *Academic radiology*, 22(10), 1299-1307.
- Wood, D., Shadbolt, N., Reichgelt, H., & Wood, H. (1992). EXPLAIN: Experiments in planning and instruction.
- Xu, Z., Zhang, H., Wang, Y., Chang, X., & Liang, Y. (2010). L 1/2 regularization. *Science China Information Sciences*, 53(6), 1159-1169.
- Yudelson, M. (2016). A Tool for fitting Hidden Markov Models models at scale. Retrieved from <https://github.com/myudelson/hmm-scalable>
- Yudelson, M. (2019). *Elo, I Love You Won't You Tell Me Your K*. Paper presented at the European Conference on Technology Enhanced Learning.
- Yudelson, M. V., Koedinger, K. R., & Gordon, G. J. (2013). *Individualized bayesian knowledge tracing models*. Paper presented at the International conference on artificial intelligence in education.
- Zheng, S., Rosson, M. B., Shih, P. C., & Carroll, J. M. (2015). *Understanding student motivation, behaviors and perceptions in MOOCs*. Paper presented at the Proceedings of the 18th ACM conference on computer supported cooperative work & social computing.
- Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective *Handbook of self-regulation* (pp. 13-39): Elsevier.
- Zimmerman, B. J. (2013). Theories of self-regulated learning and academic achievement: An overview and analysis *Self-regulated learning and academic achievement* (pp. 10-45): Routledge.
- Zimmerman, B. J., & Pons, M. M. (1986). Development of a structured interview for assessing student use of self-regulated learning strategies. *American educational research journal*, 23(4), 614-628.
- Zimmerman, B. J., & Schunk, D. H. (2001). *Self-regulated learning and academic achievement: Theoretical perspectives*: Routledge.

APPENDIX

A. Appendix A – Chapter 1

A.1 Knowledge Components, Student-Step and Opportunities

Typical Intelligent Tutoring Systems (ITSs) and educational data mining evaluation studies use these terms to refer to how the learning attempts ought to be understood. The learning tasks that reflect unique "...concepts, principle, fact, skill, schema, production rule, misconception..." that the student needs to understand can be designated as Knowledge Components (i.e. KC) (Kurt VanLehn et al., 2007). This thesis overarching assumption is that in the LIFE smartphone, emergency care concepts and principles are taught using simulated scenarios within a gamified environment. Within each learning objective (Figure 1.1.2), individual *scenarios* teach more targeted concepts through presentation of *clinical steps*, which conceptualise *care algorithms* that students (who are healthcare providers) must follow when providing emergency clinical care (e.g. Figure A.1).

The nature of the knowledge domain is defined by the kind of domain a knowledge model represents and is dependent on the granularity of its representation (Sottolare et al., 2016, pg. 168-170). LIFE KCs combine representation of two types of domain knowledge: conceptual (e.g. coloured boxes in Figure A.1) and procedural knowledge (e.g. arrows in Figure A.1) and represent domain knowledge in a structural form as a series of KCs (Sottolare et al., 2016, pg. 168-170). In LIFE's case, this structural form is operationalised as the series of steps that need to be known/followed to provide emergency care as illustrated by Figure A.1.

The domain model, as illustrated by Figure A.1, can be represented as a hierarchy of the KCs (Sottolare et al., 2016, pg. 168-173). The decision of the KCs hierarchy structure, together with the level of granularity of knowledge representation depends on the intended implementation of the learning content and is determined by domain experts (Sottolare et al., 2016, pg. 168-173). For LIFE, decisions on content design were determined by global health academics and expert medical trainers involved in ETAT+ training in SSA in consultations with the relevant medical professional bodies accrediting continuous professional development.

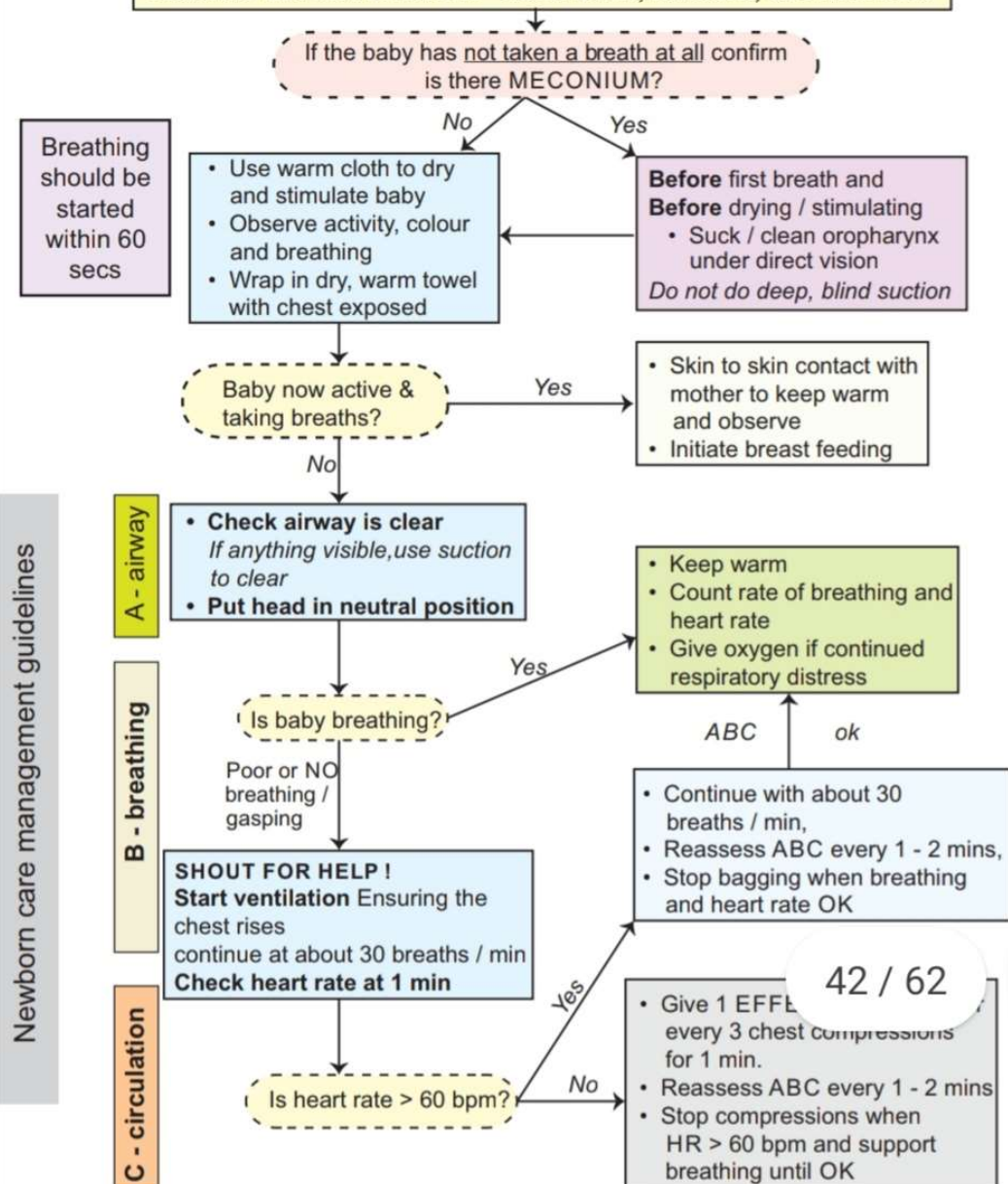
Newborn Resuscitation

For trained health workers - Be prepared

Note for all newborns:

- ✓ Practice delayed cord clamping to prevent early infant anaemia
- ✓ Clean the cord with 7.1% Chlorhexidine Digluconate (4% Chlorhexidine) once baby stable and then daily until the cord separates
- ✓ Ensure HIV risk known and give TEO & Vitamin K

PREPARE BEFORE DELIVERY - EQUIPMENT, WARMTH, GETTING HELP



38

Figure A.1: Learning content representing clinical guidelines that need follow

Multiple KCs tend to be treated as independent of each other, although they hierarchical structure can be assumed to contribute to learning performance additively or conjunctively (Corbett & Anderson, 1994; Peddycord-Liu et al., 2018). The atomic unit of knowledge linked to a student-step forms the basic unit of knowledge tracing (Chi et al., 2011). This thesis contention is that, within the LIFE platform, the atomic KC is any clinical step that requires some procedural or conceptual understanding and can be evaluated as being either correct or incorrect.

A student-step refers to each time point where the digital learning platform elicits the next step from the learner, irrespective of whether a prompt was provided or not (Chi et al., 2011). For LIFE, each attempt where LIFE elicits input from the learner which could be evaluated as being correct or incorrect is considered a student-step (Chi et al., 2011). Learning opportunities (also referred to as practice opportunities), represent the cumulative count of all the previous attempts made on the student steps on each unique KCs made by the learner (Chi et al., 2011; M. V. Yudelson et al., 2013).

B. Appendix B – Chapter 2

Results from a rapid review of learning theory used in Intelligent Tutoring Systems that were evaluated using experimental study designs between 2008 – 2018.

Table B.1: Summary of common learning theories used in the implementation of Intelligent Tutoring Systems.

Theory	Description of learning	Key concepts	Measurement imperatives	Criticisms	Notes
Self-regulatory learning theory (Zimmerman, 2013)	learning process as a flow model where each individual uses different strategies to produce an outcome that is then subjected to external feedback or revision based on an internal monitoring loop within the cognitive system	Task definition, Goal setting and planning (Forethought), Learning strategies, (Performance), Outcome monitoring and learning evaluation, Adaptations to metacognition (Self-reflection)	An optimal ITS needs to help students analyse the learning situation, set meaningful learning goals, determine which strategies to use, assess whether the strategies are effective, and evaluate their own emerging understanding of the topic.	No measure of SRL simultaneously and fully represents all targets of its model. Thus, evidence for the validity of constructs is under-represented in measurements currently used in research	The distinction between human tutor supporting metacognition vs. ITS, is that for the former, the goal is to improve learners' metacognitive behaviour, even after the metacognitive scaffolds are removed. In the latter, the goal is to improve learning while the metacognitive scaffolds are in place. In the control-value theory of achievement (Pekrun, Frenzel, Goetz, & Perry, 2007), emotions are classified as prospective (future expectancy of performance outcomes); retrospective (regarding past performance evaluations) and activity-based emotion (experienced during performance/studying).

Table B.1: Summary of common learning theories used in the implementation of Intelligent Tutoring Systems.

Theory	Description of learning	Key concepts	Measurement imperatives	Criticisms	Notes
Self-determination theory (Ryan & Deci) (Deci & Ryan, 2011)	For learning to be effective, it needs to nurture competence (of outcome and experience mastery), relatedness (of learning concepts and lived experiences), and autonomy (allowing the learner to be causal agents of their own learning progression)	Intrinsic motivation (doing something for itself, and the pleasure and satisfaction derived from participation), Extrinsic motivation (doing something as a means to an end and not for their own sake), Amotivation (no intention to act or do something).	Match Zone of Proximal Development to learner's motivational strategy Learners should experience scaffolding with appropriate motivational strategies that in turn facilitate sustained retention and effective transfer	The limitation of key concepts to just competence, autonomy, relatedness raises concerns as to how other concepts related to learning - such as on goal orientation- interact with the three. It has also been accused of confounding liking the task activity (intrinsic motivation) with achievement motivation	Intrinsic regulation is at the most self-determined end of the continuum, which is when learners engage in activities solely for the enjoyment derived from the activity itself (Deci & Ryan, 2011). Intrinsic goal framing (compared to extrinsic goal framing and no-goal framing) produce deeper engagement in learning activities, better conceptual learning, and higher persistence at learning activities
Kolb Theory (Experiential learning) (Kolb & Kolb, 2005)	Learning is the process whereby knowledge is created through the transformation of experience Do -> Reflect -> Think -> Plan: Repeat	Concrete experience (Learning by Doing), Reflective observation (Reflecting on Learning), Abstract conceptualisation (Think of the underlying conceptual knowledge being imparted), Active experimentation (Plan a test for the conceptual model identified)	Content should be designed in ways that draw on abilities from each stage of the experiential learning cycle and take the learners through the whole process in sequence.	The idea of a discrete set of learning stages does not equate to most learners' reality Learning styles are too simplistic and, whilst they fit neatly into Kolb's cycle, they fail to take account of ways of learning other than experiential	Learning as a series of processes in which the learner successively acquires a practical concrete experience, form an abstract concept by reflective observation, transform the abstract concept into a general one, apply the general concept to a new situation for verification, and develop a new individual experience and a distinctive learning style. Learning styles are a product of processing continuum (how we approach a task), and perception continuum (how we think or feel about it) and include: <i>Diverging</i> (feel and

Table B.1: Summary of common learning theories used in the implementation of Intelligent Tutoring Systems.

Theory	Description of learning	Key concepts	Measurement imperatives	Criticisms	Notes
					watch), <i>Assimilating</i> (think and watch), <i>Converging</i> (think and do), <i>Accommodating</i> (feel and do)
Felder-Silverman theory (Felder & Silverman, 1988; Graf, Viola, Leo, & Kinshuk, 2007)	Knowledge absorption is based on individual preferences and characteristics: learners have different ways they prefer to learn Learning styles are indicators of how learners perceive, interact with and respond to learning environments, including cognitive, affective and psychological behaviours (Triantafillou, Pomportsis, & Demetriadis, 2003).	Active-Reflective dimension, determines how learners prefer to process information: learn best by doing vs. think things through Sensing-Intuitive dimension determines how learners prefer to take in information: require significant memorisation of facts vs. focus on theoretical or conceptual ideas Visual-Verbal dimension determines how learners prefer information to be presented: require visual formats vs. prefer spoken/written formats Sequential-Global dimension determines how learners prefer progression in understanding: linearly ordered incremental steps vs. holistic system thinking	Modelling of learning should match learning dispositions and preferences of students Most appropriate for hypermedia courseware content design (Kuljis & Liu, 2005) Learner style designation happens once, before actual learning begins, using The Index of Learning Styles (ILS) questionnaire, developed by Felder and Soloman (Soloman & Felder, 2005)	There is no substantive evidence that learning styles exist or have any impact on learning gains, Ignores the fact that a mental strategy might be dependent on learning task, not learner's characteristic e.g. intuitive vs reflective learning tasks; Causal inference problem since narrowing of learning avenues informed by "style" might precipitate limiting understanding the content	Based on learner tendencies, although admits that that learners with a high preference for certain behaviour can also act sometimes differently.

Table B.1: Summary of common learning theories used in the implementation of Intelligent Tutoring Systems.

Theory	Description of learning	Key concepts	Measurement imperatives	Criticisms	Notes
Cognitive Load Theory (Sweller, 1994)	Learning becomes efficient once students possess sufficient knowledge to guide their own study	Intrinsic cognitive load: the idea that all instruction has an inherent difficulty associated with it Extraneous cognitive load: Under the control of instructional designers, generated by the way information is presented to learners Germane load: the load dedicated to the processing, construction and automation of conceptual schemas.	Instruction should be guided to allow working memory to focus on the construction of schemata that are required to integrate new information with the student's existing knowledge ITS should provide immediate feedback on misconceptions, as well as offering self-paced, personalized learning, based on students' interactions with the system	Lack of clarity about the cognitive load construct itself (compared to mental load, and mental effort) Difficult to validate the additivity hypothesis: That intrinsic, extraneous, and germane cognitive load all add up, and cannot exceed our working memory resources if learning is to occur. Very fuzzy distinctions between germane and extraneous cognitive load	Intrinsic load is generally thought to be immutable. It is suggested that designers ought to limit extraneous load and promote germane load. Factors which make learning unnecessarily complex or distract students from information they are trying to pay attention to, increase their cognitive load as they are processing it.

C. Appendix C – Chapter 3

C.1 Constructivist approaches to education

Constructivism requires that we turn our attention to the learners -each of whom creates their own model to explain reality- and adopt a pedagogy that balances the opportunity to assess learners' *acceptable* meaning construction and our need to construct meaning for them i.e. structure "learning" situations which channel them into evidence-based healthcare practices. This is rooted in the fundamental pillars of constructivism helpful for formulation of education interventions (George E.H., 1991) outlined below:

1. Learning as an active process in which the learner constructs meaning and systems of meaning-making systems to better give meaning to other similarly patterned situations.
2. Constructing meaning requiring cognitive effort: hands-on experience may be necessary but insufficient for learning; reflective activity should be encouraged.
3. Learning as a social activity by recognising the social aspect of learning and using interactions with environment and others as an integral aspect of learning.
4. Contextual learning: where relationship to what else we know and have experienced is given credence.
5. One needs knowledge to learn: new knowledge assimilation is dependent on having previous knowledge to build on. Efforts to train *must* reflect learner's state and provide a path into the subject for the learner based on that learner's previous knowledge.
6. Learning is not instantaneous, needs refreshing, reflective-thinking, and planning for. Reflection on learning tasks, is a product of repeated exposure and thought, where insight, is traceable to longer preparation periods.
7. Motivation is essential for learning, includes an understanding of ways in which the knowledge can be used in real-world settings, to increase learner's engagement in learning process.

C.2 The Bayesian Knowledge Tracing Model

The objective of the BKT model is to estimate probability that student has mastered a skill based on a sequence of observed responses to that skill. BKT poses this as a special case of a Hidden Markov Chain model where a different model is provided for each unique skill (Figure C.1).

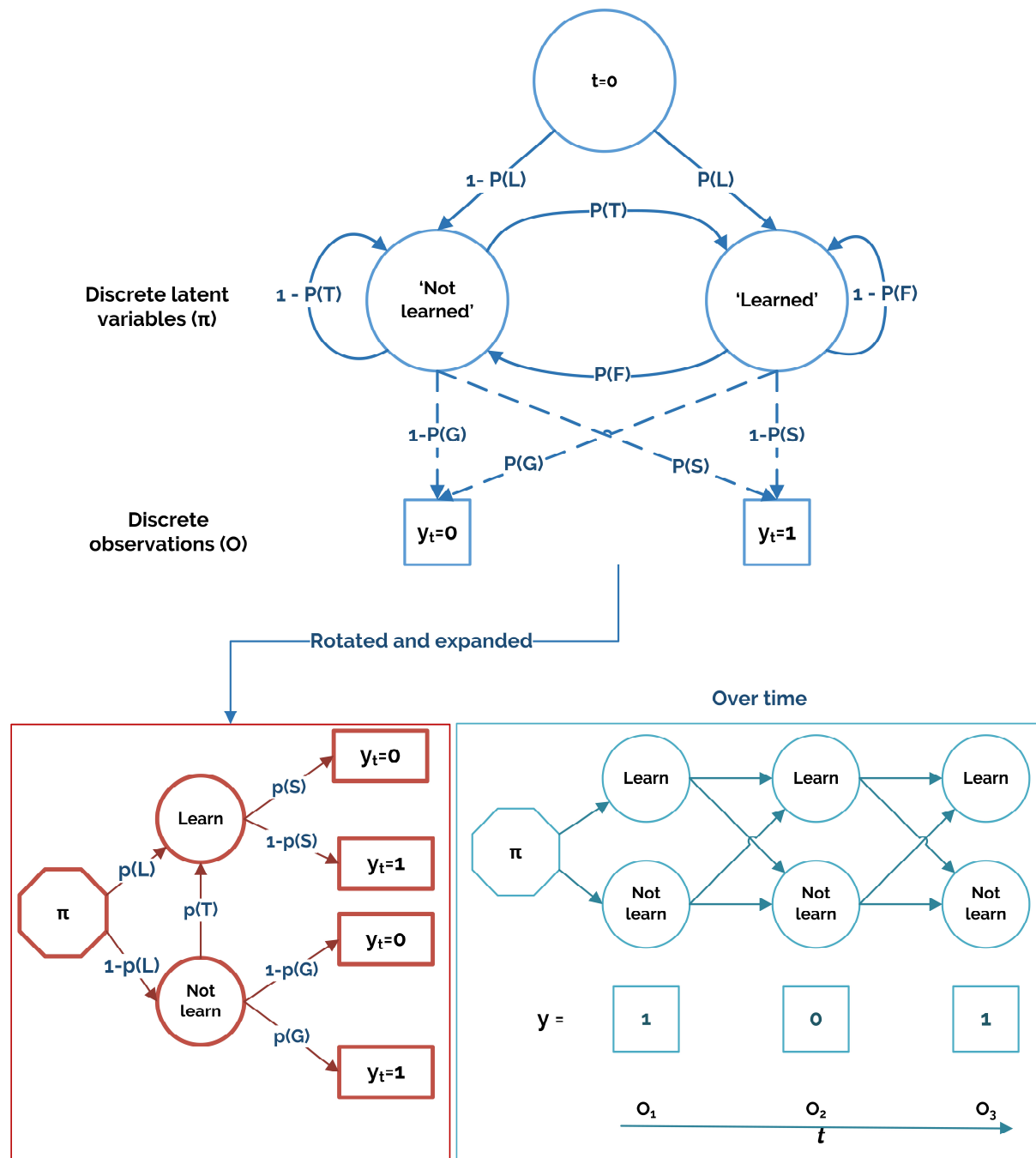


Figure C.1 Corbett and Anderson Bayesian Knowledge Tracing model

The initial probability π of student u mastering skill k is set to the **plnit** parameter

$$p(L_1)_u^k = p(L_0)^k \quad (C. 1)$$

Depending on whether the student u applied skill k correctly or incorrectly, the conditional probability is computed using the following equations:

$$p(L_{t+1}|obs = correct)_u^k = \frac{p(L_t)_u^k * (1 - p(S)^k)}{p(L_t)_u^k * (1 - p(S)^k) + (1 - p(L_t)_u^k) * p(G)^k} \quad (C. 2)$$

$$p(L_{t+1}|obs = wrong)_u^k = \frac{p(L_t)_u^k * p(S)^k}{p(L_t)_u^k * p(S)^k + (1 - p(L_t)_u^k) * (1 - p(G)^k)} \quad (C. 3)$$

The conditional probability used to update the probability of skill mastery (**learning**) according is given below.

$$p(L_{t+1})_u^k = p(L_t|obs)_u^k + (1 - p(L_{t+1}|obs)_u^k) * p(T)^k \quad (C. 4)$$

The probability of student u applying the skill k **correctly** on an upcoming practice is given below

$$p(C_{t+1})_u^k = p(L_t)_u^k * (1 - p(S)^k) + (1 - p(L_t)_u^k) * p(G)^k \quad (C. 5)$$

The approach detailed in the equations above extends the standard BKT implementation where only skill-specific parameters are used by individualising it to the learner which is predictively better (J. I. Lee & Brunskill, 2012; Pardos & Heffernan, 2010). However, this research intends to utilise skill-specific BKT model because they model parameters can be used on new study population (assuming the population is similar) and offline (individualised BKT would require every new learner to have their learning progress to be recomputed in near-real time). The process of fitting a BKT model -which is a Hidden Markov Chain Model (HMM)- is explained in detail elsewhere (Levinson, Rabiner, & Sondhi, 1983; M. V. Yudelson et al., 2013). In brief, the three fundamental problems addressed in the fitting an HMM model to trace whether the student has “learnt” the rule from Box 3.1 in Chapter 3.4 are:

- ❖ **Likelihood problem:** Given a HMM with hidden states $\lambda = (Learnt, Not\ Learnt)$ and an observation sequence $\mathbf{O}$ (such as the one in Box B.1), determine the likelihood $P(\mathbf{O} | \lambda)$.
- ❖ **Decoding problem:** Given an observation sequence $\mathbf{O}$ and a HMM with hidden states $\lambda = (Learnt, Not\ Learnt)$, discover the best hidden state sequence $\mathbf{Q}$ i.e. $P(\mathbf{Q} | \mathbf{O}, \lambda)$.
- ❖ **Learning problem:** Given an observation sequence $\mathbf{O}$ and an HMM with the hidden states $\lambda = (Learnt, Not\ Learnt)$, learn the transition and observation likelihoods, A and B from the HMM model.

The HMM model is iteratively optimised using a forward-backward algorithm, by computing initial estimates for the probabilities and using them in the next iteration to compute better estimates (M. V. Yudelson et al., 2013). The results of the iterative optimisation from fitting the HMM model is the prior (λ), transition (A), and emission matrices (B) of the following form:

$$\lambda = [p(L) \quad 1 - p(L)], \quad A = \begin{bmatrix} 1 & 0 \\ p(T) & 1 - p(T) \end{bmatrix}, \quad B = \begin{bmatrix} 1 - p(S) & p(S) \\ p(G) & 1 - p(G) \end{bmatrix} \quad (C.6)$$

The values of the equation set in C.6 correspond to the values used in equations C.2 to C.5.

D. Appendix D – Chapter 4

D.1 Ethics Approvals

D.1.1 Kenya Medical Research Institute approval



KENYA MEDICAL RESEARCH INSTITUTE

P.O. Box 54840-00200, NAIROBI, Kenya
Tel: (254) (020) 2722541, 2713349, 0722-205901, 0733-400003, Fax: (254) (020) 2720030
E-mail: director@kemri.org, info@kemri.org, Website: www.kemri.org

KEMRI/RES/7/3/1

May 22, 2018

**TO: NAOMI Muinga,
PRINCIPAL INVESTIGATOR**

**THROUGH: THE DIRECTOR, CGMR-C,
KILIFI**

Dear Madam,

RE: SERU PROTOCOL NO. 3444 (RESUBMISSION REQUEST FOR ANNUAL RENEWAL): AN EVALUATION OF A SERIOUS MOBILE GAME TO IMPROVE HEALTH WORKERS' KNOWLEDGE AND SKILLS IN EMERGENCY NEONATAL CARE IN KENYAN TEACHING HOSPITALS

Reference is made to your letter dated May 15, 2018. The KEMRI Scientific and Ethics Review Unit (SERU) secretariat acknowledges receipt of the revised application on May 17, 2018.

This is to inform you that the Expedited Review Team determined that the issue raised on the letter dated May 8, 2018 has been adequately addressed.

Consequently, the study is granted approval for continuation effective **May 22, 2018**. Please note that authorization to conduct this study will automatically expire on **May 21, 2019**. If you plan to continue with data collection or analysis beyond this date, please submit an application for continuing approval to the SERU Secretariat by **April 9, 2019**.

You are required to submit any proposed changes to this study to SERU for review and the changes should not be initiated until written approval from the SERU is received. Please note that any unanticipated problems resulting from the implementation of this study should be brought to the attention of the SERU and you should advise the SERU when the study is completed or discontinued.

You may continue with the study.

Yours faithfully,


**THE HEAD,
KEMRI SCIENTIFIC AND ETHICS REVIEW UNIT**

D.1.2 CUREC approvals

UNIVERSITY OF OXFORD DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY
Tel: +44(0)1865 274024 Fax: +44(0)1865 274027
general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

Director Professor Jo-Anne Baird



12th February 2018

Dear Timothy,

Title: Serious Gaming in Health: A theory-informed data-driven approach to learning (Evaluating an educational mobile game for teaching the resuscitation of new-borns to Kenyan health workers)

The above application (reference number ED-CIA-18-106) has been considered on behalf of the Departmental Research Ethics Committee (DREC) in accordance with the procedures laid down by the University for ethical approval of all research involving human participants.

I am pleased to inform you that, on the basis of the information provided to DREC, the proposed research has been judged as meeting appropriate ethical standards, and accordingly, approval has been granted.

If your research involves participants whose ability to give free and informed consent is in question (this includes those under 18 and vulnerable adults), then it is advisable to read the following NSPCC professional reporting requirements for cases of suspected abuse
http://www.nspcc.org.uk/Inform/research/questions/reporting_child_abuse_wda74908.html

Should there be any subsequent changes to the project which raise ethical issues not covered in the original application you should submit details to research.office@education.ox.ac.uk for consideration.

Good luck with your research study.

Yours sincerely,

Laura Molway
Member of DREC

E. Appendix E – Chapter 5

E.1 Randomisation algorithm

Figure E.1 outlines how randomisation was designed to execute on the Android Smartphones.

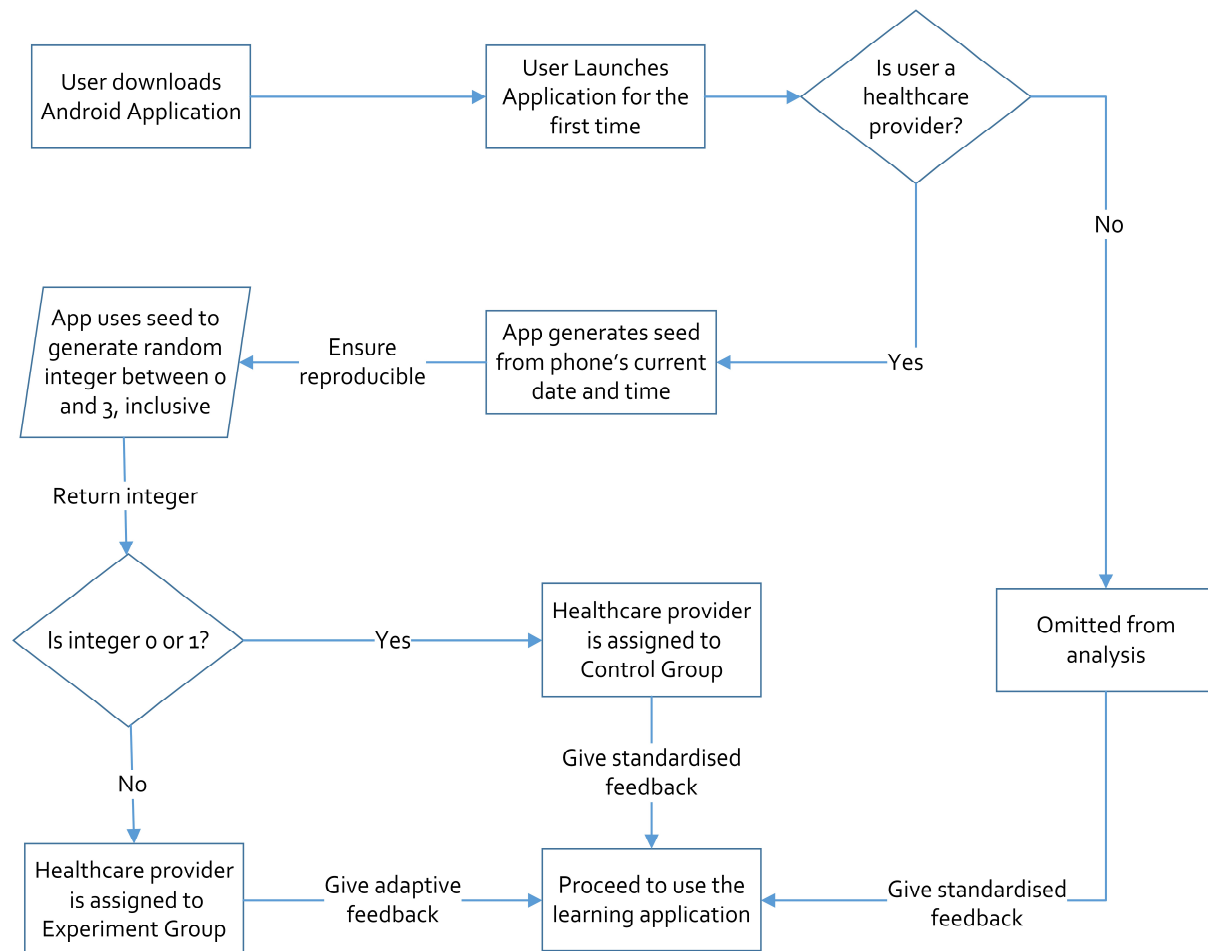


Figure E.1: Random allocation procedure

Because the random generator function (C++ **rand** function) uses a range of 4 values only i.e. [0,3], together with its unreliable performance on different computing devices (Lavavej, 2013; QuantStart, 2019), it is highly likely this was what contributed to the disparity in the allocation ratio illustrated by Figure E.2. This is compounded by a lack of an enforceable blocking mechanism based on how I implemented randomisation. This in turn made random allocation skewed towards the control group as reported in Chapter 5. However, it did not render the groups non-comparable for this experimental setup as illustrated by Tables 5.1.1 – 5.1.3 in Chapter 5.1.

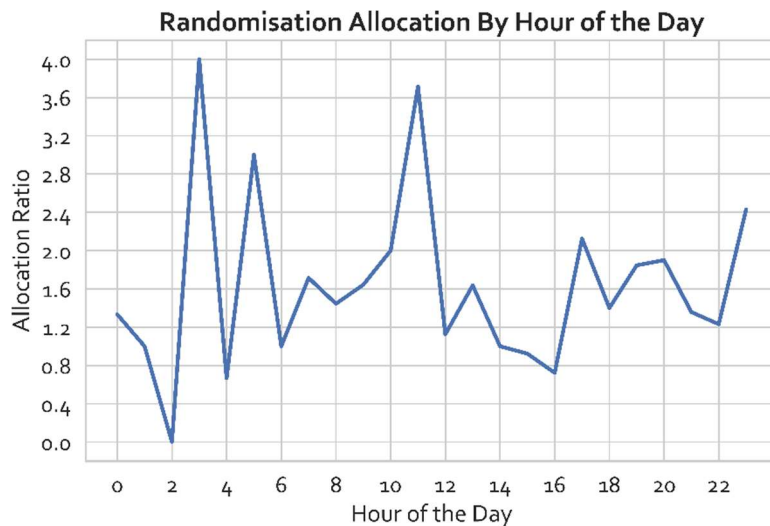


Figure E.2: Random allocation ratio based on time of day (the seed used for the generator)

	Continental Sub-Regions (Developing Countries only)	Count (N)	Percentage (%)
1	Sub-Saharan Africa*	363	69.41
2	South-Eastern Asia	44	8.41
3	Latin America and the Caribbean	38	7.27
4	Western Asia	37	7.07
5	Missing**	23	4.4
6	Eastern Asia	12	2.29
7	Northern Africa	5	0.96

Note: *299 of these are Kenya participants, representing 52% of all recruited participants. **Even though LIFE's requested Android permissions allow for collection of geographic regions of learner, there are cases where the mobile telecommunications provider did not transmit this information.

Length of spacing between learning sessions	≤1 Hour	≤1 Day	≤1 Week	≤1 Month	>1 Month	All
Control	97 (39.27)	30 (12.15)	11(4.45)	7 (2.83)	3(1.21)	148(59.92)
Experiment	66 (26.72)	17 (6.88)	8(3.24)	7 (2.83)	1(0.4)	99(40.08)
All	163 (65.99)	47 (19.03)	19(7.69)	14 (5.67)	4(1.62)	247

F. Appendix F – Chapter 6

F.1 Online Self-regulated Learning Questionnaire (OSLQ) Tool

Demographic information

Sex: Male ☐ Female ☐ Age (In years): _____

Hospital/Training centre: _____

Professional Cadre: ☐ Consultant ☐ Medical Officer ☐ Clinical Officer

Nurse ☐

Professional Level: ☐ Intern ☐ Student ☐ *Active

Experience level (in years): _____

*Any professional level where the healthcare provider is deemed not be intern or student role

Goal setting

	Very true for me	True for me	Quite true for me	Sometimes true for me	Not at all true for me
I had personal goals for the [platform] training scenario					
I had realistic performance targets for the [platform] training scenario					
Goals helped me manage my learning through [platform]					

Strategic learning

	Very true for me	True for me	Quite true for me	Sometimes true for me	Not at all true for me
When planning my learning, I adapted strategies that have worked before					
I considered alternative options <i>not provided</i> which might be better solutions [SE]					

I asked myself questions on learning content before I began learning using [platform].					
I pre-planned my study time to better accomplish my learning tasks.					

Task strategies

	Very true for me	True for me	Quite true for me	Sometimes true for me	Not at all true for me
I asked myself how what I'm learning is related to what I already know [E]					
I changed strategies when I did not meet my personal learning targets					
I had difficulty understanding new information on the [platform]'s course(s)					
While learning through the [platform], I made notes to help me understand					

Elaboration

	Very true for me	True for me	Quite true for me	Sometimes true for me	Not at all true for me
I tried to match the [platform] content with other sources of information					
Previous work experience was useful when learning through the [platform]					

Self-evaluation

	Very true for me	True for me	Quite true for me	Sometimes true for me	Not at all true for me
The score at the end reflected how well I have learned					
I have a good sense of what I have learned through the [platform]					

Help-seeking

	Very true for me	True for me	Quite true for me	Sometimes true for me	Not at all true for me
If I am having trouble learning, I prefer to figure it out by myself					

If I did not understand a task, I would prefer help to be provided before any initial attempt at it.					
I had access to peers for help when needed					
When I needed help, I asked my colleagues for more information					
The help I asked from my colleagues was different from what the [platform] provided me with.					

F.2 SRL Confirmatory Factor Analysis Model

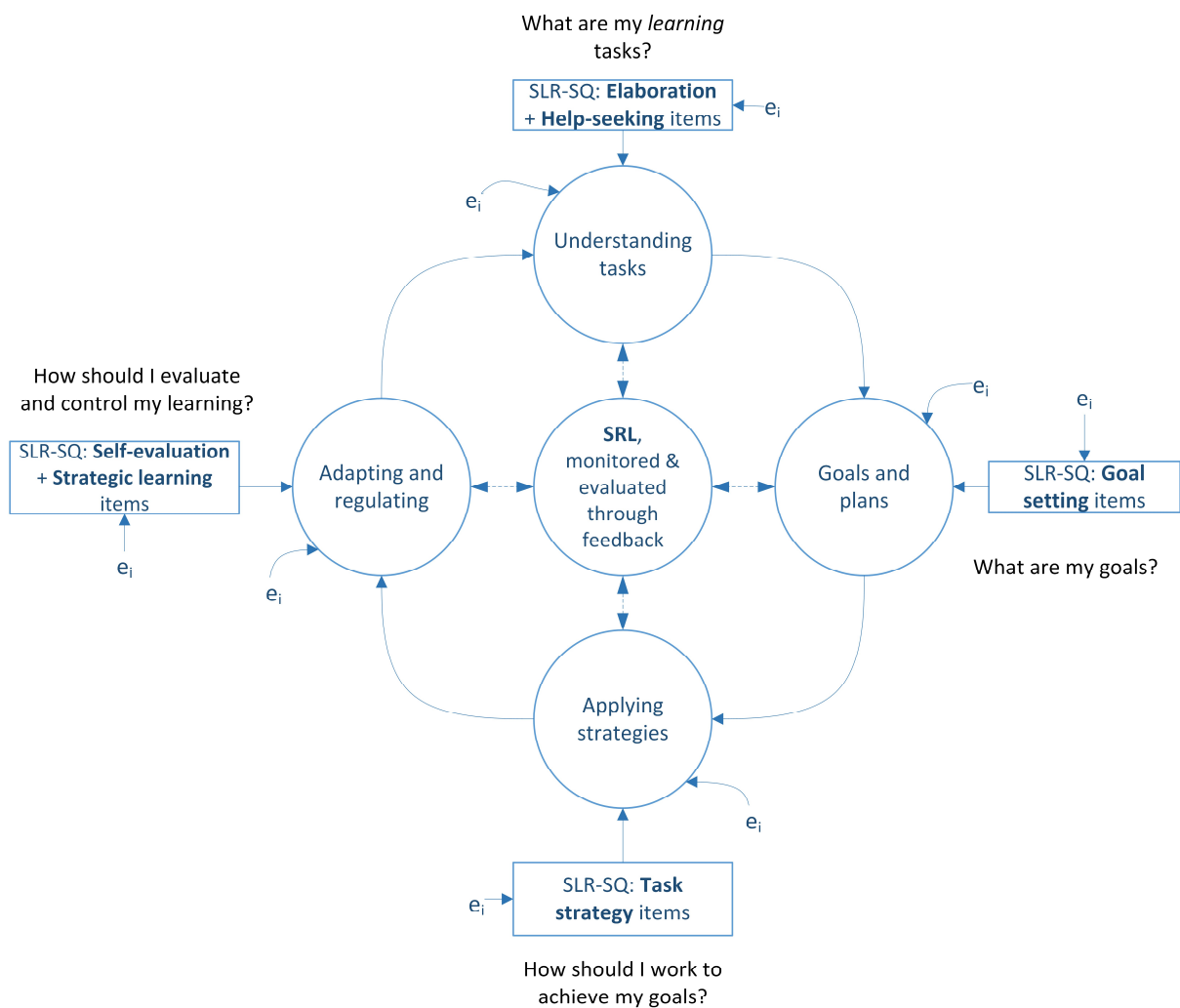


Figure F.1: SRL configuration for CFA model



Figure F.2: Three-profile model of SRL

Table F.1: SRL profiles by digital platform			
SRL \ Platform	Daktari Online N (%)	LIFE N (%)	All N (%)
Low SRL profile	22 (8.33%)	6 (2.27%)	28 (10.61%)
Average SRL profile	35 (13.26%)	26 (9.85%)	61 (23.11%)
Above Average SRL profile	11 (4.17%)	97 (36.74%)	108 (40.91%)
High SRL profile	12 (4.55%)	55 (20.83%)	67 (25.38%)
All	80 (30.3%)	184 (69.7%)	264

G. Appendix G – Chapter 7

G.1 A primer into cross-validation

Cross-validation (CV) is a machine learning technique used to evaluate model performance, which uses a re-sampling procedure to evaluate a model, an especially useful feature if there is limited data (Kohavi, 1995). In typical execution of CV, data is split into two portions, with one of the portions used to train the model, and the other portion for evaluating model's predictive performance either as a validating dataset or a test-set. This thesis used 10-fold CV in Chapter 7 and 9 with the summary explanation of this process provided in the steps below.

1. In a k-fold cross-validation, the dataset is split into k equal portions, where k is a positive integer and is usually either 5 or 10. Each subset of the data is referred to as a *fold* (Figure G.1, G.2),
2. The model is fit using k-1 folds and the k-fold can be used to:
 - a. *Tune model hyperparameters* (Chapter 8)(Figure G.2), where the model performance metrics are then evaluated from using a different test dataset, then recorded (Shah, 2017), OR
 - b. Evaluate model performance where the model performance metrics for each iteration (Figure G.1) or *split* (Figure G.2) are then recorded (Chapter 7).
3. This process is repeated until every fold has been used as a validating dataset.
4. Model performance is then evaluated as both the average of the model performance metrics across all the folds, and predictive performance on the test dataset (Figure G.2).



Figure G.1: Illustration of cross-validation

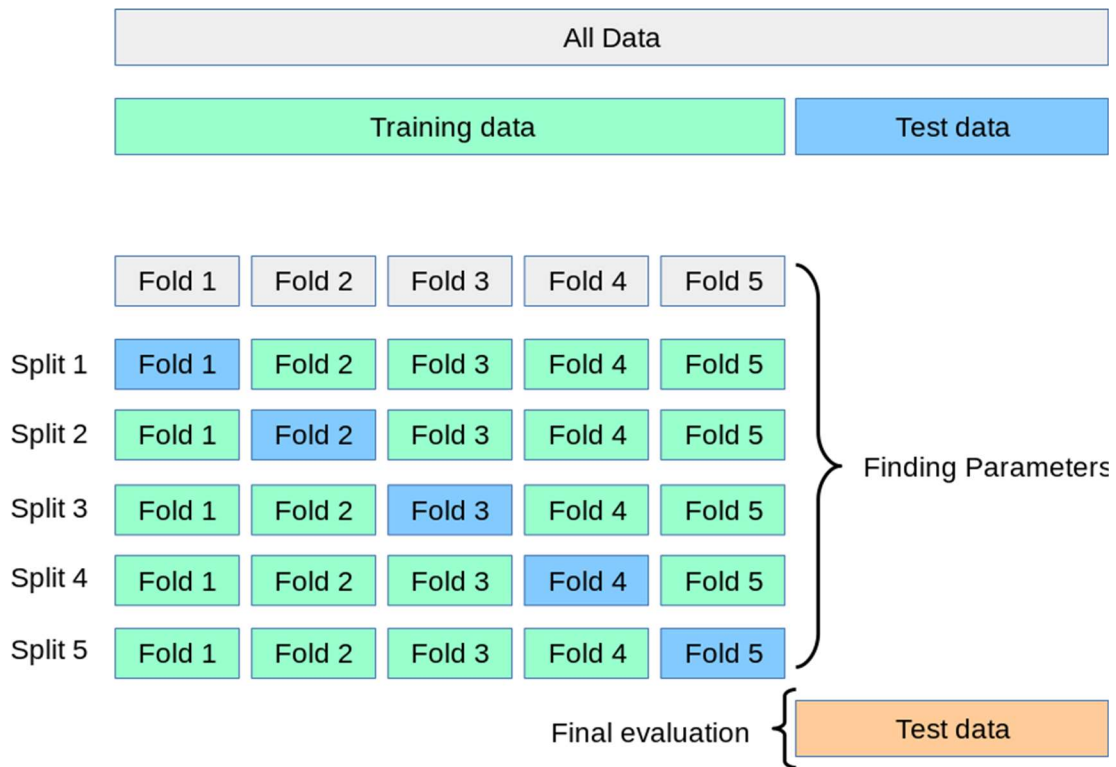


Figure G.2: Different configuration of cross-validation

H. Appendix H – Chapter 8

Table H.1: Learning Spacing By Running Learner Performance

	None	<= 1 Hour	<= 1 Day	<= 1 Week	<= 1 Month	> 1 Month	All
Score (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
0	147 (9.74%)	27 (1.79%)	12 (0.8%)	9 (0.6%)	9 (0.6%)	6 (0.4%)	210 (13.92%)
10	24 (1.59%)	0 (0%)	1 (0.07%)	1 (0.07%)	2 (0.13%)	1 (0.07%)	29 (1.92%)
20	38 (2.52%)	7 (0.46%)	5 (0.33%)	2 (0.13%)	1 (0.07%)	0 (0%)	53 (3.51%)
30	46 (3.05%)	7 (0.46%)	5 (0.33%)	2 (0.13%)	2 (0.13%)	2 (0.13%)	64 (4.24%)
40	62 (4.11%)	16 (1.06%)	1 (0.07%)	4 (0.27%)	1 (0.07%)	2 (0.13%)	86 (5.7%)
50	78 (5.17%)	55 (3.64%)	18 (1.19%)	18 (1.19%)	7 (0.46%)	7 (0.46%)	183 (12.13%)
60	54 (3.58%)	23 (1.52%)	15 (0.99%)	14 (0.93%)	6 (0.4%)	2 (0.13%)	114 (7.55%)
70	32 (2.12%)	36 (2.39%)	15 (0.99%)	14 (0.93%)	12 (0.8%)	2 (0.13%)	111 (7.36%)
80	32 (2.12%)	72 (4.77%)	28 (1.86%)	14 (0.93%)	10 (0.66%)	6 (0.4%)	162 (10.74%)
90	20 (1.33%)	74 (4.9%)	31 (2.05%)	20 (1.33%)	6 (0.4%)	4 (0.27%)	155 (10.27%)
100	39 (2.58%)	203 (13.45%)	52 (3.45%)	22 (1.46%)	18 (1.19%)	8 (0.53%)	342 (22.66%)
All (N)	572	520	183	120	74	40	1509

Note: Column-wise sums correspond to the last column of Table 7.1.

Table H.2: Multivariate MANCOVA analysis

	Value	F	df1	df2	p-value
Group	0.991653	2.788857	3	994	0.04
Cadre	0.902755	8.641079	12	2630.168	< 0.001
Group:Cadre	0.958626	3.528598	12	2630.168	< 0.001
Lag	0.724932	125.7209	3	994	< 0.001
Experience	0.984617	5.176449	3	994	0.001

Table H.3: Univariate MANCOVA analysis

Term	Dependent Variable	Sum of Squares	Degrees of Freedom	Mean Square	F-Statistic	p-value
Group	Mastery	19.1827	1	19.1827	1.8036	0.179
	Dropouts	17.0959	1	17.0959	2.2628	0.133
	Wheelspin	47.9066	1	47.9066	7.6226	0.005
Cadre	Mastery	303.3242	4	75.8310	7.1298	< 0.001
	Dropouts	472.0955	4	118.0239	15.6216	< 0.001
	Wheelspin	267.8660	4	66.9665	10.6553	< 0.001
Group:Cadre	Mastery	170.2738	4	42.5684	4.0024	0.003
	Dropouts	86.1221	4	21.5305	2.8498	0.023
	Wheelspin	179.8726	4	44.9681	7.1551	< 0.001
Lag	Mastery	1437.8014	1	1437.8014	135.1852	< 0.001
	Dropouts	23.3483	1	23.3483	3.0904	0.079
	Wheelspin	1454.6340	1	1454.6340	231.4526	< 0.001
Experience	Mastery	122.5994	1	122.5994	11.5271	< 0.001
	Dropouts	82.8193	1	82.8193	10.9619	< 0.001
	Wheelspin	4.4851	1	4.4851	0.7137	0.398

I. Appendix I – Chapter 9

I.1 Nominal Group Technique Data Collection Procedure

Details on how to conduct Nominal Group Technique (NGT) are detailed elsewhere (McMillan et al., 2016). Due to concerns about exposure to serious gaming and digital learning technologies, limited time going through the LIFE app and language barriers, NGT main question - “*How might LIFE™ be better designed to support and motivate personalized learning in Kenyan context?*” - was broken down into modular questions using visual aides to help guide the NGT. In brief, the phases listed below explain how NGT was carried out.

1. *Nominal or silent phase.* During this phase individuals considered their responses to **the question** privately and write down their responses without discussion. Interaction was discouraged to prevent individuals dominating the process of idea generation. The role of the facilitator at this stage was to introduce the task and to focus the group upon the question without influencing the responses individuals made.
2. *Item generation phase.* The ideas generated by individuals were then shared within the group. This was done through a process which involved the leader asking each member in turn to present an idea in a **succinct** phrase. All ideas were recorded verbatim on a flip chart. This procedure continued without discussion until no new ideas were forthcoming. Participants were allowed to pass when they had no items to present but could contribute again in future rounds if an idea occurred to them subsequently.
3. *Discussion and clarification phase.* This step in the process enabled the participants to ask questions, to seek clarification and to indicate agreement or disagreement with any items on the list. Each item was discussed sequentially, and participants were able to combine items with similar meanings or to eliminate those items which were duplicated. The facilitators role during this phase was to lead the discussion and to ensure that the influence of the more vocal or powerful members of the group was minimised.

4. *Voting phase.* Each participant was asked to select ten items on the list that were most important to them and to rank these in order of priority, assigning a value of 1 to the most important item and a value of 10 to the least important item. This is done independently and privately. The facilitator then collected the ranking forms and tallied the results in order to determine the group's priorities.

I.2 Guide For the Brief Open-Ended Semi-Structured Interview (Stage 2)

- Open with introductions
- Go through informed consent process
- LIFE game play

GLOBAL AIM: How might LIFE be better designed to support and motivate personalised learning of new and continuing healthcare courses in Kenya.

1. What about real-life frontline experience in a Kenyan hospital practice does this LIFE scenario not address?

2. In your experience, given varying learning capabilities of clinician students, how would you advise LIFE to better individually support each learner?

3. What would you advise LIFE to include to better encourage self-assessment and reflective evaluation within student clinicians?

4. Within LIFE what do you think should be done to ensure accurate understanding of the material at a personal level?

I.3 Guide For the Open-Ended Semi-Structured Interview (Stage 3)

- Open with introductions
- Go through informed consent process

1. **How many times did you end up playing LIFE?**
2. **[IN GENERAL]**, how has your experience going through ETAT+ material through LIFE smartphone application been?
 - a. **Why do you think it took you #N tries to achieve a score of 100%?**
 - b. **Do you think your experience had any role in how you performed? Why?**
3. **Does the content you've learnt through LIFE relate to the training you are currently receiving?**
4. What made ETAT+ content on the LIFE smartphone application **[DIS]**engaging for you and why?
5. **[IN DETAIL]**, how did you find the feedback provided through LIFE? Why?
 - a. **If you didn't get a task correct, how would you prefer help to be provided?**
6. Did you use the LIFE application more than once?
 - a. If yes,
 - i. What made you use LIFE for learning again and why?
 - ii. How did your learning strategy change in the subsequent play and why?
 - iii. What would make using LIFE more rewarding for you?
 - b. If no,
 - i. What made you not repeatedly use of LIFE?
 - ii. Finish this sentence: "*If only LIFE was ... I would use it more*"
7. What would motivate you to re-use LIFE to learn **during your own free time?**
8. Given the chance to earn *Continuous Professional Development* (CPD) points, would you use LIFE differently? Why **[NOT]**?

9. If you knew you were going to be asked to provide neonatal resuscitation care based on your score on LIFE, how would you use LIFE app differently?

10. Did learning through LIFE have any impact on your beliefs about your capabilities to deliver emergency care? Why?

11. Anything more to add?