

# Building A Learner Model for a Smartphone-Based Clinical Training Intervention in a Low-Income Context: A Pilot Study

Timothy Tuti<sup>1\*,3</sup>[0000-0002-7915-3004], Chris Paton<sup>2</sup>[0000-0001-6952-9621],  
Mike English<sup>2,3</sup>[0000-0002-7427-0826], Niall Winters<sup>1</sup>[0000-0001-8597-2914]

<sup>1</sup> Department of Education, University of Oxford, 15 Norham Gardens, Oxford OX2 6PY, UK

<sup>2</sup> Centre for Tropical Medicine and Global Health, University of Oxford, Oxford OX1 3SY, UK

<sup>3</sup> KEMRI-Wellcome Trust Research Programme, Nairobi 00100, Kenya

\*timothy.tuti@kellogg.ox.ac.uk

## Abstract

Research is lacking on developing adaptive learning applications for training health workers in low-resource settings making student modelling approaches supporting individualised learning to remain largely unexplored. This study targeted a clinical training intervention using smartphones in a low-resource context to explore if clinicians' performance patterns can be differentiated into distinctive groups based on an inferred proficiency level using cluster analysis. We also explored the applicability of Knowledge-Component (KC) cognitive learning models - Additive and Performance Factor Models (AFMs, PFMs) - in describing these patterns and their accuracy in predicting performance. The intervention provides simulation training on contextualised management of new-born resuscitation through a series of learning interactions that elicit responses through multiple-choice answers and interactive tasks. AFMs and PFMs were used to explore the impact of previous exposure to KCs within the learning intervention on learner performance. We demonstrate that effectiveness of low-dose-high-frequency training might be linked to successful attempts in previous learning sessions. Additionally, there exists intermediate and expert cadres of health workers who would benefit more from cascading-challenge scenarios. From these results, we propose a preliminary cognitive learning model as a basis for adaptive instructional support on smartphones for clinical training in low-resource settings.

**Keywords:** Serious gaming, predictive accuracy, clinical training, smartphones, neonatal care, emergency care, Sub-Saharan Africa, Performance Factor Models, Additive Factor Models

## 1. Background

Sub-Saharan Africa (SSA) produces over 24% of the global disease burden but only has 3% of the global health workforce [1, 2]. This severe workforce shortage, coupled with health workforce skill imbalance and maldistribution, and lack of training opportunities are major contributors to the poor quality of neonatal care outcomes in the region [3]. Mobile technologies (smartphones and tablet computers) have shown potential to address this learning challenge in SSA, given their uptake rate (around 30-

50% of adult populations) and pattern of usage (around 30-35% use it to access internet for information sourcing) [4, 5]. A typical health worker in this setting works very long hours, would find it hard to pay personally for face-to-face training, is likely unable to spend much time or money on learning online, and the institution they work for would usually also be constrained financially from funding further training [6, 7]. Consequently, their learning must be flexibly integrated into very busy working lives and mechanisms for reinforcing learning must be strengthened. There is little evidence from low-resource contexts such as Africa of learning interventions that are cognisant of this context, that take into account individual health workers' initial and continuing clinical training needs, and that deliver tailored learning content, feedback and resources in light of skill mastery and performance as they continue to develop knowledge through it (i.e. adaptive learning) [4, 8, 9].

Such learning adaptations are common in the Intelligent Tutoring Systems (ITSs) 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 [10]. 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 [11]. The learning tasks that produce these data points represent unique Knowledge Components (KC) which reflect learning "...concepts, principle, fact, skill, schema, production rule, misconception..." [12]. The most common student modelling approaches for KCs are Additive Factor Models (AFMs) [13], and Performance Factor Models (PFMs) [14], and detailed explanations of these are provided in the next section. Appreciating that reinforcement of KCs 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) [15, 16]. Evidence of the successful implementation of student-modelling approaches on digital platforms in clinical training in order to facilitate adaptive learning and improve learning outcomes is scarce [17], and virtually non-existent for emergency care training in low income settings [18]. In high income settings, despite the important role of smartphones in facilitating personalised learning, there is still a lack of research investigating mobile-based ITSs [19]. These are the gaps that ITS are yet to systematically address. Additionally, differences between learners' achievement goal orientations (such as skill mastery-intrinsic, mastery-extrinsic, performance-approach, performance-avoidance etc.) [20] and how that is reflected in the uptake of smartphones-based learning approaches in low-income settings is largely unexplored. 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 [21]. Such metrics would arguably help inform

the successful implementation of digital training platforms to bridge the skills gap in clinical care provided in low-income settings where low-cost highly accessible training opportunities are hard to come by.

### 1.1. Additive Factor Models (AFMs) And Performance Factor Models (PFMs)

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. Skill-mastery has commonly been conceptualised as the odds of learners completing a task correctly as a linear function of the prior opportunities they had for the learning task, conditioned on skill difficulty (i.e. AFMs). It has also been commonly conceptualised as the odds of learners completing a task correctly after taking the correctness of learners' responses into account based upon previous performance features such as the number of previous successful and unsuccessful practices (i.e. PFM). Additive Factor Models (AFM) are used to evaluate conjunctive skills in learning data. Its **additive** nature is due to a linear combination of skill parameters determining  $p_{ij}$  described in the equation below:

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

**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  $j$ ,  $\theta_i$  is the coefficient for proficiency of student  $i$ ,  $\beta_k$  is coefficient for difficulty of the knowledge component or skill  $k$ ,  $Q_{kj}$  is the Q-matrix cell for step  $j$  using skill  $k$ ,  $\gamma_k$  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 [11]. 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 [13]. On the other hand, Performance Factor Models (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 [22]. It estimates the different effects of practicing learning opportunities.

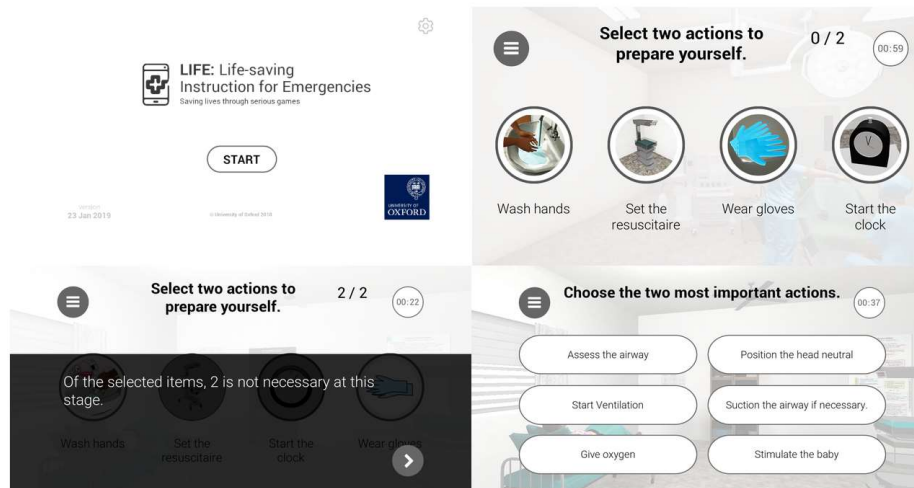
$$\text{logit}(p_{ijt}) = \theta_i + \beta_j + \alpha_j S_{ijt} + \rho_j F_{ijt} \quad (2)$$

**Where:**  $i$  represents student  $i$ ,  $j$  represents step  $j$ ,  $\theta_i$  is the coefficient for proficiency of student  $i$ ,  $\beta_j$  is coefficient for difficulty of the step  $j$ ,  $X_{ijt}$  is binary correct/incorrect outcome for student  $i$  at step  $j$  on trial  $t$ ,  $S_{ijt}$  is the count of previous success up to trial  $t$ ,  $F_{ijt}$  is 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 [23].

## 1.2 The intervention

The Life-Saving Instruction for Emergencies (LIFE) project [24] -which is the platform this research uses- is a serious games platform intended for use with low-cost smartphones to provide training in the care of very sick neonates, particularly in low resource settings with the hope 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 [25, 26]. Consequently, it follows a specific ordering of clinical care-giving algorithms with each learning task being timed. The learner starts a scenario which provides some background information to the learning task, and on each learning task, must provide input either through multiple choice questions, selection of items necessary for the learning task, or performing on-screen interactive tasks (e.g. navigating to equipment, switching on machines etc.) (Figure 1).



**Fig 1.** Selected screenshots of LIFE application

On each incorrect attempt by the learner, standardised feedback is provided with the option of more information and the learner must repeat until they successfully respond to the question before being allowed by the smartphone application to proceed. The end of the scenario is signalled by a crying baby indicating that the baby is now breathing, with a breakdown of scores by quiz provided. The scenario model that is used is one that replicates Emergency Triage, Assessment and Treatment plus admission care i.e. ETAT+ face-to-face training approach training that is validated [27, 28]. The ETAT+ content it adapts has already been used to train over 5,000 healthcare workers and 2,000

medical students across Eastern and Southern Africa, and now East Asia [27, 28]. LIFE is meant to be accessible at scale by healthcare providers and able to function off-line on low-end smartphone devices and provide self-regulated training opportunities akin to continuous professional development at almost no cost. We don't know have much evidence about adaptive learning in this context and using these types of interventions.

The aim of this study was to analyse data from a mixed cohort of LIFE users to: (1) Explore existence of learning patterns indicative of individual differences between players; (2) Compare AFM and PFM directly on a learning dataset derived from clinical training on smartphone devices to evaluate their predictive accuracy of learners' performance in a low-income context; (3) Propose a preliminary cognitive model of learning as a basis for adaptive student-step instructional support on smartphone devices for a low-income setting based on the observed behaviours in (1) and (2). This was done to generate a working model for how adaptive learning might work as a basis of an ITSs model.

## **2 Methods**

### **2.1 Study design, setting and participants**

This study was a retrospective observational study [29] of healthcare providers from both public and private hospitals in Kenya, in clinical cadres such as nurses, clinical officers and medical doctors, with experience levels varying from students to consultants. Participants were enrolled into the study through a combination of snowballing and convenience sampling strategy. Recruitment occurred through use of peer referrals among clinicians, private professional social network accounts, regional clinical meetings, medical conferences, medical training institutions and local hospitals. In total, 187 participants were recruited. The eligibility criteria for inclusion were that the participants had to be either in training for, or active in, clinical care. Therefore, the participants included those with experience in offering clinical care.

### **2.2 Study Variables, And Data Management**

The LIFE version used in this study provides 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 platform provides performance score feedback based on the learner's first attempt at each learning interaction. Data collection was through Android-based LIFE smartphone application, which would securely transmit a copy of anonymised data to a Google Firebase distributed database. For the purposes of the proposed analysis, the outcome of interest was specified as getting each answer correct on the first try. The variables of interest were time spent on learning task, number of previous tries (i.e. opportunities) per learning task, and whether hints had been provided for each unique try per learning task.

### 2.3 Statistical Methods, Missing Data, And Sensitivity Analyses

Data manipulation and statistical analyses were performed using R software's *glmer* and *TraMineR* packages [30-32]. Variables of interest are reported using their mean value and standard deviation. Longest Common Subsequence (LCS) cluster analysis [31] was used to explore whether there existed differentiable student learning trajectories from the sequence of their performance on LIFE content on first try. LCS was used to ensure that the conjunctive nature of LIFE content (steps to resuscitate a neonate in distress) was factored into how learning trajectories are derived. Time on task, hint usage and previous opportunities on the learning task were used as illustrative variables to explore how the derived clusters vary by these features. In the second phase of analyses, Additive Factor Model (AFM) and Performance Factor Model (PFM) were used to construct a cognitive model based upon the learning behaviours and performance data observed from LIFE, and to explore the ability of these models to make predictions on the patterns of learning by adult students using LIFE platform in low-income context and for clinical training. Finally, from the analyses we propose a model for implementation of adaptive personalised learning on smartphone devices for clinical training in low-resource settings. For the purposes of analysis, users with missing data (i.e. incomplete learning session on LIFE) were omitted. However, to evaluate whether excluding observations with missing data would bias the results, analysis of the difference in means for the features of interest between the complete and missing data group was conducted. The sensitivity of the prediction accuracy for AFM and PFM is reported as both the average classification accuracy from 10-fold cross-validation [33] of the dataset, and the area under curve (AUC) computed from training the models on 70% of the data and testing their performance on remaining 30% (test dataset). This was done to assess how well these models can distinguish between unseen learner performances on the seen ETAT+ content delivered through LIFE and evaluate if these models were overfitting the learning data.

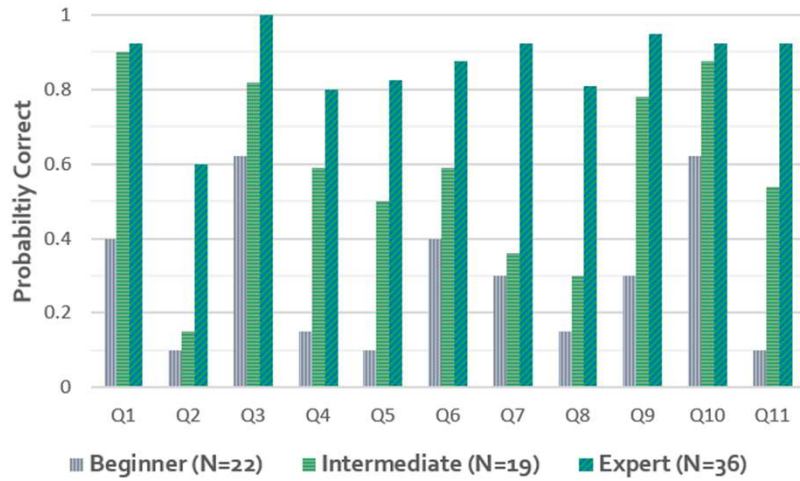
## 3 Results

The data reported were observed between 23<sup>rd</sup> April 2018 and 13<sup>th</sup> October 2018. Of the 187 users recorded as having downloaded and started playing the LIFE game in this period, 77 learners (41.17%) completed a full learning session. Table 1 describes all learners who attempted to use LIFE, divided into those who had a complete learning session and those who did not complete a learning session. Due to inability to collect demographic data within the LIFE application at the time of data collection, a detailed breakdown of participants' backgrounds is not possible. LIFE training session data from non-completers of the game was not included in subsequent analyses as we are confident of minimal bias due to only using the complete dataset in the subsequent analysis given that we demonstrated non-significant differences in indicators of interest between completers and non-completers (Table 1). Only data from learners who completed a session were used in subsequent analyses. The use of the dataset with complete learning sessions was to ensure that analysis was reflective of the sequentially conjunctive nature of LIFE content (ordered steps to resuscitate a neonate in distress) and learner's actual performance across all quiz items.

**Table 1.** Summary statistics of pilot data from LIFE game play

Indicator	Complete*		Incomplete**		P-Value <sup>‡</sup>
	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 i.e. Opportunities	2.18	2.72	2.62	3.4	0.328
Average performance (%) ***	55.66	28.08	49.02	31.42	0.132
<b>Note:</b> <i>*Learners who completed at least one session: N=77,</i> <i>**Learners who did not complete at least on learning session: N=110.</i> <i>***Average performance based on number of quizzes attempted.</i> <i>‡From evaluating if there is a difference in the mean of the values between 'Complete' and 'Incomplete' groups</i>					

From the cluster analysis, learner performance could be categorised into three distinctive groups based on inferred proficiency level to reflect individual differences (Figure 2). The identification of the three clusters was guided by the Point Biserial Correlation, Average Silhouette Width, and Calinski-Harabasz indices [34]. From Figure 2, quiz two -which was about the selection of equipment necessary for resuscitation- appeared problematic for all learners, with beginners performing poorly across all quizzes, while learners in the intermediate cluster struggled in quizzes between five and eight. Learners in expert category had exemplary performance that improved with subsequent quizzes. This also suggests the need to classify the ‘difficulty’ of the questions as well as the learners. From the variables in Table 1, there was no substantive difference in the odds of the time taken to complete a learning task between the ‘beginner’ and ‘expert’ proficiency cluster (Table 2). A possible explanation for this might be that beginners might be guessing a lot and experts know so both appear to be quick. However, the relevance of time spent on learning task for the ‘intermediate’ proficiency cluster was almost twice the odds of the other two. As expected, provision of feedback on incorrect attempts significantly predicted membership to the ‘beginner’ group unlike the other proficiency categories, with learners in this category having almost twice the odds of being provided with this type of feedback compared to the other groups (Table 2).



**Fig 2:** Distinctive clusters of learning trajectories as defined by performance. Y axis represents proportion of answers that are correct.

The effect of previous opportunities at attempting the quiz was significant across all proficiency groups, with ‘expert’ proficiency group associated with a better use of these opportunities than the other proficiency groups. Based upon the observed behaviours reported in Table 2, we sought to apply common cognitive models (AFM and PFM) and evaluate their ability to explain student performance in this setting.

**Table 2.** How learner proficiency clusters vary by learning metrics

	Beginner		Intermediate		Expert	
<i>Predictors</i>	<i>Odds Ratios</i>	<i>95% CI</i>	<i>Odds Ratios</i>	<i>95% CI</i>	<i>Odds Ratios</i>	<i>95% CI</i>
Task Time	0.67***	0.53–0.85	1.94***	1.53–2.46	0.69**	0.52–0.90
Opportunity	0.90**	0.83–0.97	0.90**	0.84–0.97	1.17***	1.10–1.25
Feedback	1.89*	1.02–3.50	0.95	0.50–1.78	1.16	0.53–2.51
Learners, N (%)	22 (28.6%)		19 (24.7%)		36 (46.8%)	
Nagelkerke’s R <sup>2</sup>	0.214		0.065		0.226	
<i>Note: *** = p-value ≤ 0.001, ** = p-value ≤ 0.01, * =p-value ≤ 0.05</i>						

The outcome of interest was getting the answer correct on first try. The independent variables were a combination of the KCs (quizzes), opportunities and cluster membership. The exploratory hypothesis was that membership in these learner clusters (beginner, intermediate, and expert) would be associated with progressively better learning rates and behaviours for utilisation of time on task, feedback and when skill difficulty and opportunities at knowledge components are considered. From the results



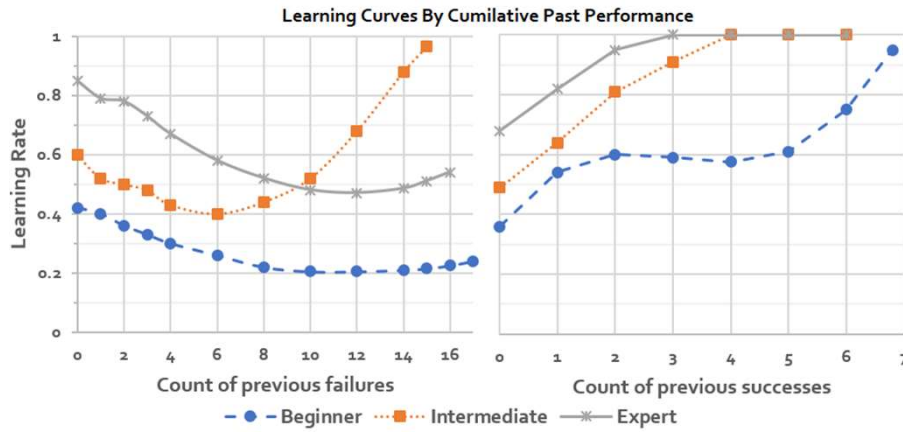
of the AFM and PFM student modelling analyses, in general, as expected, (1) the odds of a learner completing a learning task correctly increased based on the proficiency clusters, from ‘beginner’ to ‘expert’ clusters, and (2) the time on learning task had a significant positive effect on the odds a learner in ‘intermediate’ cluster completing a learning task correctly given prior opportunities they had had at the learning task and conditioned on skill difficulty. However, the use of feedback on incorrect attempts had a significant positive effect for increasing the odds of completing the learning task in the ‘intermediate’ group when the number of prior opportunities at the learning task were considered.

**Table 3:** Results from student modelling based on learning opportunity modes

	AFM Prediction		PFM Prediction	
Predictors	<i>Estimates</i>	<i>95% CI</i>	<i>Estimates</i>	<i>95% CI</i>
Task Time	-0.04***	-0.06 – -0.03	-0.05***	-0.07 – -0.03
Feedback (Ref: No Feedback)	-0.04**	-0.07 – -0.02	-0.04	-0.07 – 0.00
Ref: Beginner				
Intermediate	0.14***	0.11 – 0.17	0.16***	0.12 – 0.20
Expert	0.35***	0.32 – 0.37	0.39***	0.35 – 0.43
Task Time (Ref: Beginner)				
Task Time (Intermediate)	0.03 **	0.01 – 0.05	0.04**	0.01 – 0.07
Task Time (Expert)	-0.01	-0.03 – 0.01	0.01	-0.02 – 0.04
Feedback (Ref: Beginner)				
Feedback (Intermediate)	0.05*	0.01 – 0.09	0.04	-0.02 – 0.10
Feedback (Expert)	0.02	-0.02 – 0.06	0.06*	0.01 – 0.11
AIC/BIC	2023.6/2159		1844.3/2044.7	
Nagelkerke’s R <sup>2</sup>	0.467		0.384	
Note: *** = $p\text{-value} \leq 0.001$ , ** = $p\text{-value} \leq 0.01$ , * = $p\text{-value} \leq 0.05$ , $N = 1663$				

It is also shown to have a positive effect on the odds of learning outcome in the ‘expert’ group, when previous opportunities were broken down into successful and unsuccessful attempts on learning task (table 3). Overall, across all proficiency groups, previous successes were associated with a better rate of a progress in gaining experience or new skills i.e. learning curves. However, there was an unexpected activation of a significant

positive association from previous failures in the intermediate proficiency group, who continued to attempt learning tasks despite initial decline in learning rate, which in the long run – was associated with increasingly better learning outcomes (Figure 3).



**Fig 3.** LIFE learning curves from performance factor model

The evaluation of how accurate these cognitive models are in constructing and predicting the learning behaviours and performance data observed from LIFE was tested on 30% of the data with 70% being used as training set. From a 10-fold cross-validated model evaluation on the training set, AFM had an accuracy of 68% while PFM had an accuracy of 71%. When the models were tested on the 30% of the data set not used in training them, AFM had an AUC score of 77% while PFM had a score of 81%. In both instances, PFM outperformed AFM in accurately predicting and distinguishing the performance from ‘unseen’ students on LIFE knowledge components. While the R-squared values reported paint AFM to be better than PFM for modelling LIFE learning data, the accuracy and AUC scores indicate that the AFMs might be overfitting the data more than PFM, given relatively weaker performance in prediction of learning data from ‘unseen’ students on ‘seen’ steps. Overall, it would appear that PFM model has best model performance (Table 3), minimises bias of overfitting compared to AFM, and explains learner behaviour relatively well.

## 4 Discussion

### 4.1 Summary of Findings

The aim of this study was to explore users’ learning patterns from a smartphone based clinical training intervention in low-income settings and explore which student modelling approaches are best representative of the learning performance and intervention use behaviour. This was done to provide a basis for proposing a cognitive model of adaptive learning on smartphone devices for a low-income setting based on the observed learning behaviours. From analyses, based on patterns of performance on LIFE content, three proficiency learner groups were uncovered: beginner, intermediate and expert. The time spent on learning tasks between the beginner and expert groups was similar, with intermediate proficiency group spending almost twice as much time

on each knowledge component compared to the other groups (Table 2). Despite initial failures, learners in the ‘intermediate’ proficiency group demonstrated positive learning gains in the long run with each subsequent opportunity on learning task (Figure 2). This might be indicative that the experts know the content, beginners guess and learners in the intermediate group try to think. In general, previous successes were most influential in producing higher learning gains with previous failures having the opposite effect on ‘beginner’ and expert groups. While ‘beginner’ group used feedback hints approximately 50% more than the other groups (Table 2), it was the ‘expert’ group that was able to capitalise on usage of use of hints for higher learning gains (Table 3). For LIFE content delivered through smartphone, the predictive accuracy of 80% for performance factor models was moderately good [35] and is comparatively better than use of additive factor models. This would make it more appropriate in constructing cognitive model of learners who use LIFE for clinical training.

#### **4.2 Relation To Other Studies**

The emphasis of emergency care training in low resource settings to use low-dose high-frequency (LDHF) training is not new but has been recently introduced in SSA [36]. Our findings support this approach by demonstrating how learning gains are most improved where the opportunity to learn is high. This is commonly done to facilitate gains in knowledge and skills within health workers [16]. However, this study goes further by demonstrating that in general, gains in the effectiveness of high-frequency low-dose training might be linked to successful attempts in previous training sessions. Additionally, we found within health workers, a cluster of learners (intermediate and experts) who would arguably benefit more from challenging scenarios which require more time in reflection. How different clinical training intervention use metacognitive scaffolds to improve knowledge gains is not new[37-39], but hardly present in research from low resource settings [18]. Moreover, while using adaptive learning demonstrates significantly better knowledge gains than alternatives [37-39], the current LDHF training models –which are not adaptive to individual learner needs- are still the most commonly implemented models of learning, usually face-to-face, and at a very high cost [15, 16, 36]. This study further explores how the use of smartphone devices to deliver clinical training using short simulation-based learning activities, can begin to accommodate self-regulated learning over time, which have been show to optimise learning in similar settings [40].

#### **4.3 Implications of Findings**

From our findings, in low resource settings, LDHF scenario-based clinical training conditioned on skill difficulty and learner proficiency, might produce higher cumulative learning gains where previous opportunities at the learning task are successful. Additionally, such education interventions might need to accommodate learners who prefer to struggle, who take their time in making attempts, who purposefully underutilise feedback, preferring repeated unguided attempts. While this might not be true for all learners, using smartphone devices to offer LDHF clinical training is yet to adopt ways to pin-point differentiated learning preferences that might guide better instructional design. Given the limited data used for these analyses, additional qualitative work will be conducted to validate the findings. 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 achievement goals which are arguably not linked to getting high performance in the shortest time possible but rather, taking time to reflect on the learning concept(s). This however, might not be true for all learners. From our smartphone intervention, a cognitive model for clinical training in low-resource settings using smartphone devices might be better served if it encourages repeated practice, while allowing learners to take control of how long they prefer to struggle on knowledge components, with feedback as a way out. From our findings, this would allow for learners who are activated to learn in presence of both past failures and successes respectively, while offering more support to those in most need, such as beginners.

#### **4.4 Limitations**

While AFMs and PFM models had reasonably moderate performance on LIFE data, the relatively low accuracy of cross-validation values on figure is disconcerting. This might be due to the low numbers of observations analysed in the whole study in general, making it challenging to provide more accurate estimates. However, given that the data collected is from a pilot -arguably unique- study looking at the utility of digital learning metrics in prediction of skill-mastery for clinical training in low-income settings, it sheds light into a previously underexplored topic. This limitation can be addressed at a later stage as we continue to generate data to support the evidence base of these kinds of interventions. Further qualitative studies will be conducted to support interpretation of findings. While this study's sample is hardly generalisable, its inclusive constitution (from students to consultants, in all clinical cadres) makes it highly informative as a realistic data source on developing cognitive models for adaptive emergency care training on smartphone platforms delivered to health workers in low income settings. We are yet to find a comparable student-step data source (and studies) for this subject in this context.

### **5 Conclusions**

In this study, we analyse the smartphone-based learning patterns for a clinical training intervention in low-income settings and explore which cognitive approaches are best representative of the learning performance and intervention use behaviour. Overall, this research found that in scenario-based learning approaches can extend low-dose high-frequency training approaches offered through smartphone devices, by targeting differentiated learner groups whose learning rates significantly vary. While in general, they all share positive learning gains from previous successes at learning tasks, among them, are those whose use of time on learning tasks in combination of presence of past failures, produced positive learning gains. Additionally, hints through feedback are utilised more by those with 'low' proficiency but produce significantly higher learning gains in those with higher proficiency. Future work will explore the comparative effectiveness on learning outcomes, of differentiated feedback conditioned on proficiency level in low-dose high-frequency training approaches delivered through smartphone devices.

## 6 Acknowledgements

Funds from Economic and Social Research Council (ESRC) awarded to TT through Oxford University, with additional funds from GCRF's Intelligent Support Project and Saving Lives at Birth supporting this work. The funders had no role in drafting or submitting this manuscript.

## References

1. Anyangwe, S. and C. Mtonga, *Inequities in the Global Health Workforce: The Greatest Impediment to Health in Sub-Saharan Africa*. International Journal of Environmental Research and Public Health, 2007. **4**(2): p. 93.
2. Sousa A. and Flores G., *Transforming and Scaling up Health Professional Education and Training*, in *Policy Brief on Financing Education of Health Professionals*. 2013, WHO: Geneva, Switzerland
3. UNICEF, *Levels & Trends in Child Mortality. Report 2013*. New York, USA, 2013. 2013.
4. Edgcombe, H., C. Paton, and M. English, *Enhancing emergency care in low-income countries using mobile technology-based training tools*. Arch Dis Child, 2016.
5. Silver, L. and C. Johnson. *Internet Connectivity Seen as Having Positive Impact on Life in Sub-Saharan Africa*. Pew Research Center - Global Attitudes and Trends 2018 [cited 2018 18th Dec]; Available from: <http://www.pewglobal.org/2018/10/09/majorities-in-sub-saharan-africa-own-mobile-phones-but-smartphone-adoption-is-modest/>.
6. Couper, I., et al., *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, 2018. **18**(1): p. 553.
7. Barteit, S., et al., *E-Learning for Medical Education in Sub-Saharan Africa and Low-Resource Settings*. Journal of medical Internet research, 2019. **21**(1).
8. Bollinger, R., et al., *Leveraging information technology to bridge the health workforce gap*. Bulletin of the World Health Organization, 2013. **91**: p. 890-892.
9. Greenhalgh, T., *Computer assisted learning in undergraduate medical education*. Bmj, 2001. **322**(7277): p. 40-44.
10. Vanlehn, K., *The behavior of tutoring systems*. International journal of artificial intelligence in education, 2006. **16**(3): p. 227-265.
11. Chi, M., et al., *Instructional factors analysis: A cognitive model for multiple instructional interventions*. 2011.
12. VanLehn, K., P. Jordan, and D. Litman. *Developing pedagogically effective tutorial dialogue tactics: Experiments and a testbed*. in *Workshop on Speech and Language Technology in Education*. 2007.
13. Cen, H., K. Koedinger, and B. Junker. *Comparing two IRT models for conjunctive skills*. in *International Conference on Intelligent Tutoring Systems*. 2008. Springer.
14. Pavlik Jr, P.I., H. Cen, and K.R. Koedinger, *Performance Factors Analysis--A New Alternative to Knowledge Tracing*. Online Submission, 2009.
15. Chaudhury, S., et al., *Cost analysis of large-scale implementation of the 'Helping Babies Breathe' newborn resuscitation-training program in Tanzania*. BMC Health Services Research, 2016. **16**(1): p. 681.
16. Willcox, M., et al., *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, 2017. **13**(1): p. 88.
17. Fontaine, G., et al., *Effectiveness of adaptive e-learning environments on knowledge, competence, and behavior in health professionals and students: protocol for a systematic review and meta-analysis*. JMIR research protocols, 2017. **6**(7).

18. Opiyo, N. and M. English, *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, 2015(5): p. 1.
19. Mousavinasab, E., et al., *Intelligent tutoring systems: a systematic review of characteristics, applications, and evaluation methods*. Interactive Learning Environments, 2018: p. 1-22.
20. Rawlings, A.M., A. Tapola, and M. Niemivirta, *Predictive effects of temperament on motivation*. International Journal of Educational Psychology: IJEP, 2017. **6**(2): p. 148-182.
21. Vandewaetere, M., et al., *Adaptivity in Educational Games: Including Player and Gameplay Characteristics*. International Journal of Higher Education, 2013. **2**(2): p. 106-114.
22. Galyardt, A. and I. Goldin. *Recent-performance factors analysis*. in *Educational Data Mining 2014*. 2014.
23. Gong, Y., J.E. Beck, and N.T. Heffernan, *How to construct more accurate student models: Comparing and optimizing knowledge tracing and performance factor analysis*. International Journal of Artificial Intelligence in Education, 2011. **21**(1-2): p. 27-46.
24. University of Oxford. *Life-Saving Instructions for Emergency (LIFE)*. 2016 [cited 2018 18th Dec]; Available from: <https://oxlifeproject.org/>.
25. Bergeron, B., *Developing Serious Games*. Game Development Series. Charles River Media. Inc., Massachusetts, 2006.
26. Wang, R., et al., *A systematic review of serious games in training health care professionals*. Simulation in Healthcare, 2016. **11**(1): p. 41-51.
27. Ayieko, P., et al., *A Multifaceted Intervention to Implement Guidelines and Improve Admission Paediatric Care in Kenyan District Hospitals: A Cluster Randomised Trial*. PLOS Medicine, 2011. **8**(4): p. e1001018.
28. Irimu, G., et al., *Developing and Introducing Evidence Based Clinical Practice Guidelines for Serious Illness in Kenya*. Archives of disease in childhood, 2008. **93**(9): p. 799-804.
29. Mann, C., *Observational research methods. Research design II: cohort, cross sectional, and case-control studies*. Emergency medicine journal, 2003. **20**(1): p. 54-60.
30. R Core Team, *R: A Language and Environment for Statistical Computing*. 2013, R Foundation for Statistical Computing: Vienna, Austria.
31. Gabadinho, A., et al., *Analyzing and visualizing state sequences in R with TraMineR*. Journal of Statistical Software, 2011. **40**(4): p. 1-37.
32. Bates, D., et al., *Fitting linear mixed-effects models using lme4*. arXiv preprint arXiv:1406.5823, 2014.
33. Kohavi, R. *A study of cross-validation and bootstrap for accuracy estimation and model selection*. in *Ijcai*. 1995. Montreal, Canada.
34. Desgraupes, B., *Clustering indices*. University of Paris Ouest-Lab Modal'X, 2013. **1**: p. 34.
35. Khajah, M., R.V. Lindsey, and M.C. Mozer, *How deep is knowledge tracing?* arXiv preprint arXiv:1604.02416, 2016.
36. Atukunda, I.T. and G.A. Conecker, *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, 2017. **5**: p. S12.
37. Feyzi-Behnagh, R., et al., *Metacognitive scaffolds improve self-judgments of accuracy in a medical intelligent tutoring system*. Instructional science, 2014. **42**(2): p. 159-181.
38. Veredas, F.J., et al., *A web-based e-learning application for wound diagnosis and treatment*. Computer methods and programs in biomedicine, 2014. **116**(3): p. 236-248.
39. Wong, V., et al., *Adaptive tutorials versus web-based resources in radiology: A mixed methods comparison of efficacy and student engagement*. Academic radiology, 2015. **22**(10): p. 1299-1307.
40. van Houten-Schat, M.A., et al., *Self-regulated learning in the clinical context: a systematic review*. Medical Education, 2018. **52**(10): p. 1008-1015.