Exploring the quality of skill-mastery prediction from Bayesian network learner models for smartphone-based paediatric care training in low-resource settings

Timothy Tuti
Oxford University/ KEMRI-Wellcome Trust
60 Banbury Rd, Oxford OX2 6PN, UK
timothy.tuti@kellogg.ox.ac.uk

Chris Paton
CTM&GH, Oxford University
South Parks Road, Oxford OX1 3SY, UK
chris.paton@ndm.ox.ac.uk

Niall Winters
Department of Education, Oxford University
15 Norham Gardens, Oxford OX2 6PY, UK
niall.winters@education.ox.ac.uk

ABSTRACT
There remains a lack of evidence on the implementation of Intelligent Tutoring Systems in low-resource settings, and in particular, of how adaptive instructional support can be implemented to support clinical training on smartphone devices. A core part of this challenge to determine an appropriate data and modelling approach to support adaptive instruction on mobile devices. Using data from a serious-gaming smartphone application for clinical training from Sub-Saharan Africa, this paper investigates models to support prediction of learner performance as a precursor to determining skill-mastery level. We explore Bayesian graph model configurations that predict unseen learner responses based on seen responses and investigate different combinations of these models that factor in time on task, and previous cumulative learning opportunities respectively. Our results show that a modelling approach that predicts learner performance while considering previous learning opportunities is more accurate than approaches that predicts learner performance based on time they spent on a learning task. Using time-on-task in combination with previous learning opportunities to augment prediction of learner performance produced no substantive increase on prediction accuracy compared to just using previous learning opportunities only. We discuss how our findings provide an avenue for introduction of adaptive scaffolding of feedback instruction, based on probabilistic performance thresholds informed by cumulative tries from previous attempts with the goal of helping the learner gain skill-mastery.

Author Keywords: Serious gaming, predictive accuracy, Bayesian graph models, clinical training, smartphones, neonatal care, emergency care

BACKGROUND
Sub-Saharan Africa (SSA) has a severe health workforce shortage (Anyangwe et al., 2007; Sousa A. et al., 2013), which coupled with skill imbalance, maldistribution, and lack of training opportunities, are major contributors to the poor quality of neonatal care outcomes in this region (UNICEF, 2013). Costs of face-to-face refresher trainings in the SSA region remain prohibitively high (Chaudhury et al., 2016) with efforts further constrained by the socio-economic and institutional landscape (Edgcombe et al., 2016). Mobile technologies (smartphones and tablet computers) have shown potential to address this training challenge for clinical care in SSA, given their uptake rate (around 30-50% of adult population) and pattern of usage (around 30-35% use it to access internet for information sourcing) (Silver et al., 2018). However, there is little evidence to support the implementation of learning interventions that are contextually relevant to low-income settings, that take into account health workers' initial and continuing clinical training needs and that adapt learning content in light of skill mastery and performance as they continue to develop knowledge through it (Bollinger et al., 2013; Edgcombe et al., 2016; Greenhalgh, 2001). Within low-resource contexts, investigation of learner models needed to support tailored instruction based on prediction learner's skill mastery in clinical settings is important because adaptive instructional support significantly outperforms teacher-led large-group instruction, non-adaptive computer-based instruction, and paper-based instruction (Ma, 2017) from research in high resource contexts. However, these significant positive effects of instructional support using adaptive feedback in clinical training in high-resource settings (Feyzi-Behnagh et al., 2014; Wong et al., 2015) are yet to be replicated in low-resource settings in a way that is both cognisant of the context and accessible to a vast majority of the clinical cadres (Edgcombe et al., 2016). This ignores socio-cognitive and cultural aspects of the way people learn. A typical health provider in this setting works very long hours, can hardly afford face-to-face trainings which are usually not institutionally sponsored, nor has the time or money to spend on data charges for online learning (Barteit et al., 2019; Couper et al., 2018). Their learning -which would most likely have higher learning gains from adaptive learning (Feyzi-Behnagh et al., 2014; Wong et al., 2015)- must be flexibly integrated into their context of work. In these low-resource contexts, due to these resource constraints, typical neonatal and paediatric emergency care ends up being offered by clinicians with limited or outdated training and mechanisms for reinforcing learning hardly exist contributing to the poor neonatal and paediatric care outcomes.

While there is a newfound urgency in seeking to address knowledge gaps in healthcare training in these settings using digital platforms in order to facilitate improvement of child mortality outcomes (Mwaikambo et al., 2016), presence of studies conducted in low-income resource contexts looking at adaptive learning on digital platforms like smartphones are generally scarce. From the few available studies found, adaptive learning tends to be viewed as making sure the learning content is relevant for the context, which is a limited view that does not differentiate individual progression while using these digital learning platforms (Mwaikambo & Dolphyne, 2016). Another study in this setting reports using gamification as a scaffolding environment to foster learning goal attainment where adaptivity is geared towards teaching styles, not students' learning needs (Botha et al., 2015). The existing studies in this context fail to offer evidence on how the platform metrics have been instrumental in evaluating skill mastery within the learning platforms themselves and view adaptive learning as making sure the learning content is relevant for the context. We are yet to come across any empirical studies in low-resource contexts that test tailored instruction approaches such as adaptive feedback for clinical training on digital platforms, that we could use as a basis for adaptive learning design. Subsequently, in this region, differences between learners' achievement goal orientations (such as skill mastery-intrinsic, mastery-extrinsic, performance-approach, performance-avoidance etc.)(Rawlings et al., 2017) and how that is reflected in uptake of smartphones-based learning approaches in low-income settings is largely unexplored.

There exists several educational models that are common choices for providing mastery ability estimation and diagnosing individual knowledge gaps that can guide tailored instruction approaches (Reckase, 2009; Templin et al., 2010). One modelling approach that provides greater measurement precision and model parsimony while exposing latent mastery ability and individual learning proficiency is use of Bayesian Networks (BNs) (Pearl, 2014). BNs provide a convenient and intuitive framework for specifying complex joint probability distributions that are well suited for determining what content the student has probably mastered and supporting a mechanism for adapting instructional support on digital devices (Conati, 2010; Peña-Ayala, 2014) but evidence of their use on smartphones is scarce. While BNs have been used to support adaptive feedback, there is scarce evidence for their use within smartphone devices, for clinical education, and in low-resource settings (Vanlehn et al., 2001) despite being deemed to provide adequate methods for the representation, interpretation and contextualisation of learning data (Lacave et al., 2018). In high-income settings, while the evidence gap for BNs use in clinical training persists, in the few available examples, they have been used to represent the current problem-state and all acceptable next steps (Feyzi-Behnagh et al., 2014), where they evaluate learner's ability to appraise the characteristics of the problem-state from non-exclusive options. In this context, they tend to be used to determine the degree of correctness of learner's response to learning task and scaffolding the difficulty level of learning content within clinical training (Feyzi-Behnagh et al., 2014; Wong et al., 2015). Where clinical training is on algorithmic care to be offered, BNs' directed acyclic graph (DAG) feature can be leveraged for modelling learning, where each node represents a variable (i.e. clinical step) and each arc between two variables indicates the existence of a probabilistic dependence between them. This can be used to represent the intuitive way the clinical care steps are dependent on each other and attributes that might affect the outcome of the emergency care being given (e.g. time) based on a conditional distribution (Lacave et al., 2018). In general, there is scarce evidence of how such use of BNs for evidence propagation (i.e. probability distribution over each unobserved variable, given that the value taken by some other variables is known) have been used to mitigate poor learning outcomes in clinical trainings, and of what configuration of a set of learning variables might produce the most optimal learning outcomes (Pearl, 2014). Additionally, evidence to support how such BNs might be configured in smartphones to support provision of tailored feedback in clinical training with high accuracy in low resource settings such as SSA remains scarce.

The Life-Saving Instruction for Emergencies (LIFE) project (University of Oxford, 2016) -which is the platform this research uses- is a serious games intervention designed for use with low-cost smartphones to provide training in care of very sick neonates and paediatric patients, particularly in settings like SSA. It evolves the scenario-based teaching model that is used in a traditional face to face training approach (Emergency Triage, Assessment and Treatment plus admission care, ETAT+) that has been validated (Ayieko et al., 2011; Irimu et al., 2008). 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 (Ayieko et al., 2011). More details about LIFE are provided elsewhere (Tuti et al., 2019). Key differences from previous studies introduced by LIFE's approach in exploring adaptive instruction include: (1) unlike previous studies (Feyzi-Behnagh et al., 2014), ETAT+ content being tested in LIFE is not configured to have multiple responses with varying degrees of correctness, rather it is based on how current quiz items are predictive of future quiz items where content is conjunctive with only one correct option at the end of each learning task, and (2) LIFE is exploring how to provide instructional support algorithmically based on learner performance in combination with other in-game metric(s) as opposed to exclusively using a threshold based on the number of incorrect answers provided or only considering opportunities at learning task (Veredas et al., 2014; Wong et al., 2015). This is especially important where the learning content and context are different to those in previous studies (Feyzi-Behnagh et al., 2014). The aim of this study was to explore if BNs can be used to provide adaptive learning experiences in clinical training on 'offline' smartphone devices. use learning data from LIFE - a smartphone-based clinical training application- to explore which Bayesian network learner-model configurations offered greatest precision and model parsimony in predicting student mastery of ETAT+ content as a precursor to providing adaptive tailored instructional feedback on LIFE smartphone application in efforts to develop knowledge.

METHODS

Study design, setting and participants

The analyses reported here used cross-sectional learning data of healthcare workers collected through the LIFE smartphone application. These healthcare workers were from both public and private hospitals in Kenya, varied by clinical carders such as nursing, clinical officers and medical doctors, with experience levels varying from students (who by the second year of their study are involved in frontline patient care) to consultants. They were enrolled into the study through a combination of snowballing and convenience sampling strategy. Recruitment occurred through use of peer referrals by clinicians, private professional social network accounts, regional clinical meetings, medical conferences, medical training institutions and local hospitals. The eligibility criteria for inclusion was that the participants had to be in active (or in training for) bedside-based healthcare practice. The included participants would typically be end up providing both neonatal and paediatric emergency care in this setting, due to the health workforce shortage. Their usage of LIFE followed a self-regulated learning approach. Ethical approvals did not allow for collection of demographic data within the LIFE application at the time of the pilot, therefore a breakdown of performance by participants' background characteristics is not possible.

Study intervention, variables, and data management

The intervention is LIFE Android smartphone app, which has been developed to focus on training healthcare workers to identify and manage medical emergencies, 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). LIFE provides multiple simulation training scenarios on the contextualised management of new-born resuscitation through a series of questions that elicit responses from learners in the form of multiple-choice answers or performing navigational and interactive tasks. At this pilot stage, users verified as having met inclusion criteria would be provided with a link to the application download to their own personal smartphones. The application, which has been designed to work in almost all low-end smartphones, requires no login to use, and users are identifiable using the smartphone's device unique identifier.

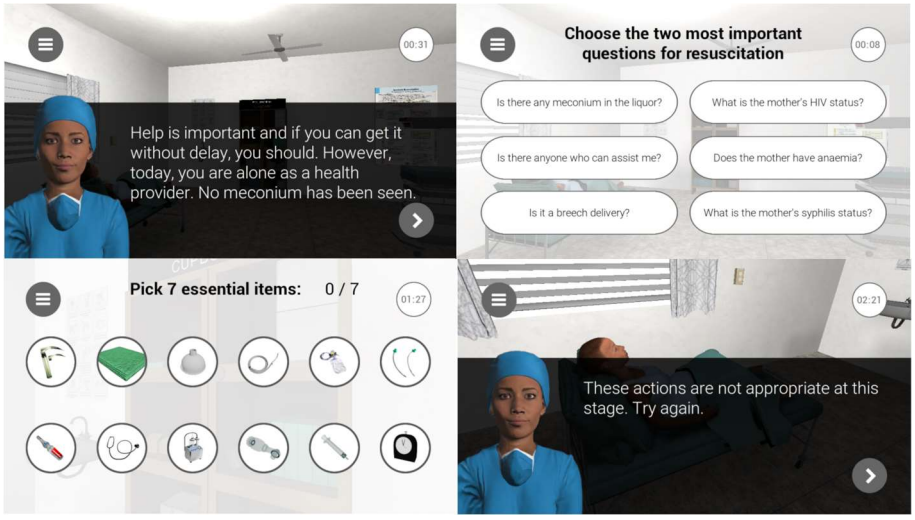

Figure 1. Examples of types of LIFE quizzes and feedback interfaces

At the beginning of each learning scenario, contextual information about the baby and the hospital setup is provided to guide the learner in decision making. Each learning task is timed by a clock, will elicit responses from the learner, and will allow the learner to proceed once they provide the correct response. In the event of an incorrect response, the learner is forced to repeat, and provided with standardised detailed feedback after the second incorrect try with a button for more information on the learning task (Figure 1). After the last learning task in the scenario, the end of a successful learning session is marked by a crying baby signalling that they are now breathing on their own. LIFE also provides performance score based on first attempt on each learning interaction at the end of each learning scenario with a detailed breakdown of the score. In the current LIFE instance, the learning scenarios include (1) Standard neonatal resuscitation, (2) Neonatal resuscitation with CPR, and (3) Neonatal resuscitation - birth with meconium. For the purposes of this study, only the standard neonatal resuscitation scenario was being piloted. Data collection was through Android-based LIFE smartphone application, which would securely transmit a copy of anonymised data to Google Firebase distributed databases. For the purposes of proposed analysis, the outcome of interest was specified as getting the answer correct on the first try. The variables of interest were time spent on learning task, number of tries per learning task, and whether feedback had been provided for each unique try per learning task. The variables selected reflect student-step variables commonly used by popular learner cognitive models to model learning rates of knowledge components by students in other (Chi et al., 2011).

Statistical methods, missing data, and sensitivity analyses

Data manipulation and statistical analyses were performed using R software(R Core Team, 2013) primarily through using the *bnlearn* package (Scutari, 2009). Variables of interest are reported using the mean and standard deviations. In efforts to identify a parsimonious model, the following Bayesian network structures were specified based on previous studies that

attribute performance learning gains on digital platforms to be associated with a combination of time-on-task (Romero et al., 2011), opportunities-per-task (Chi et al., 2011). The models were specified as fully connected directed acyclic graphs for 11 quiz items, with first having only the quiz items, the second having the quiz items and the time variable at each node, the third having quiz items and cumulative learning opportunities at each node, and the fourth including time, and learning opportunities at each node. These represent performance based on first attempt at quiz (**P**), correct on first try per quiz in current learning session given the cumulative number of previous attempts from current and previous learning sessions (**P+O**), correct on first try factoring in the time taken to attempt quiz (**P+T**), and finally correct on first try conditioned on a combination of time to attempt quiz and number of previous attempts at quiz (**P+T+O**). Time on learning task was discretised by introducing three level factors derived from quantile distribution of time variable for each quiz item. This was done to overcome *bnlearn*'s inability to allow continuous nodes (e.g. time) to be parent of discrete nodes (e.g. binary outcome correct-on-first-try). Sensitivity analysis for the discretization of time variable was done comparing a three-level and five-level ordinal discretised version, which found no difference in performance between the two forms. A key assumption based on ETAT+ content implemented in LIFE was that given the learning sequences are conjunctive and ordered based on clinical care-giving algorithms, the network would adopt a directed acyclic graph (DAG) structure to reflect this ordering by constraining the prediction from variables in the current quiz steps to only those that are after it by using a whitelist (allowable relationships in network structure) or blacklist (disallowed relationships in network structure).

Prediction accuracy for the expertly-defined inference network models was validated through employing a leave-one-out cross-validation (LOOCV) on LIFE data, where LIFE data for all but one learner would be used as the training set, and the one learner's data being used as the validation set, and this validation sequence happening for all learners in the LIFE dataset. The emphasis of the prediction, was gauging how well the models were able to predict the learner would answer the current quiz correctly, given their cumulative performance on previous quizzes, conditional on time and opportunity on learning task. Sensitivity analysis was conducted in two forms: (1) By using one hill-climbing and one constraint-based algorithm as exemplars to learn network structure and evaluate whether using a 10-fold cross-validation technique, the learned models P, P+O, P+T, P+T+O predictive accuracy is similar to those derived from LOOCV and (2) Using a model averaging approach to build a network containing only the significant arcs for P, P+O, P+T, P+T+O, and evaluate how well the expertly specified relationships in the four models are detected. This would allow to test whether a data-driven approach would predict learning performance comparable better than an expert-driven approach, and whether a hybrid approach that minimises risk of overfitting can be derived from combining the two prior approaches (Scutari, 2017). The network structures learned from the exemplar hill-climbing and constraint-based algorithms would be forced to adhere to the underpinning ETAT+ learning expectations by using the blacklist of graph relationships but would self-determine what the whitelisted graph structure would look like. Bayesian Information Criterion (BIC) score and Bayes Factor (BF) will be used to compare the expertly specified model with the averaged model, with area under the curve being used to evaluate the strength of the probabilistic relationships in the network (Claeskens et al., 2008).

RESULTS

The data reported was observed between 23rd April 2018 and 13th October 2018. Of the 187 users recorded as having downloaded and started playing the LIFE game in this period, only 77 learners (41.17%) completed a full learning session. Given the lack of difference between the complete and drop-out group, we are confident that the bias was minimal due to only using the complete dataset (Table 1).

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 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

Note:*Learners who completed at least one session: N=77, **Learners who did not complete at least on 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 for 'Complete' and 'Incomplete' using two-sample t-test for equal means.

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

From the results of LOOCV, the *Performance only* (**P**) model had moderately accuracy of 66.8% (95% CI: 50.8-82.7), which improved when time was considered (**P+T**), 71.7% (95% CI: 56.5-87), when opportunity was considered (**P+O**),

82.4% (95% CI: 70-94.8), where time and opportunity both combined with performance (**P+T+O**), 79.3% (95% CI: 66.2-92.4) (Figure 2). Across the expertly-specified models, **P+O** and **P+T+O** demonstrated the greatest improvement in model prediction accuracy from **P** model, although **P+T+O** did not offer any added accuracy over **P+O** model. This might be because conceptually, more tries take more time, therefore opportunity embeds within some explanatory power for time.

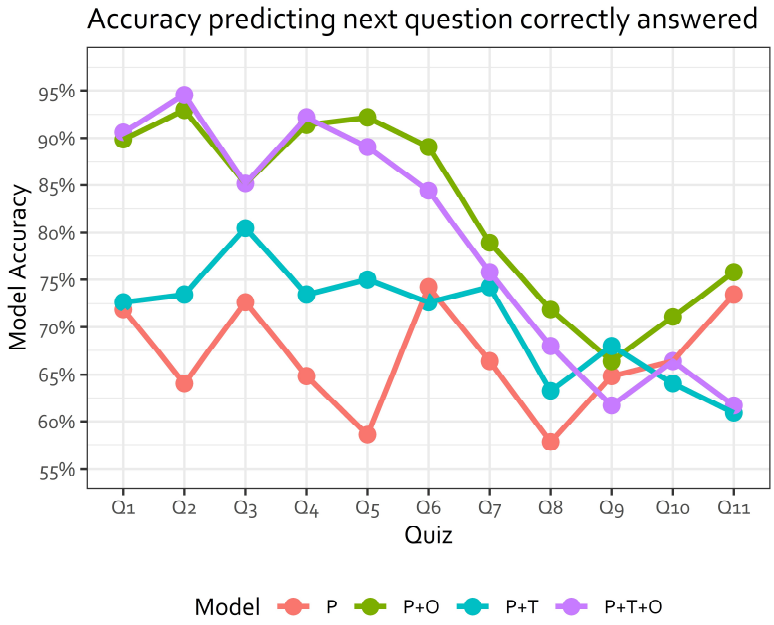

Figure 2. Expertly specified model performance

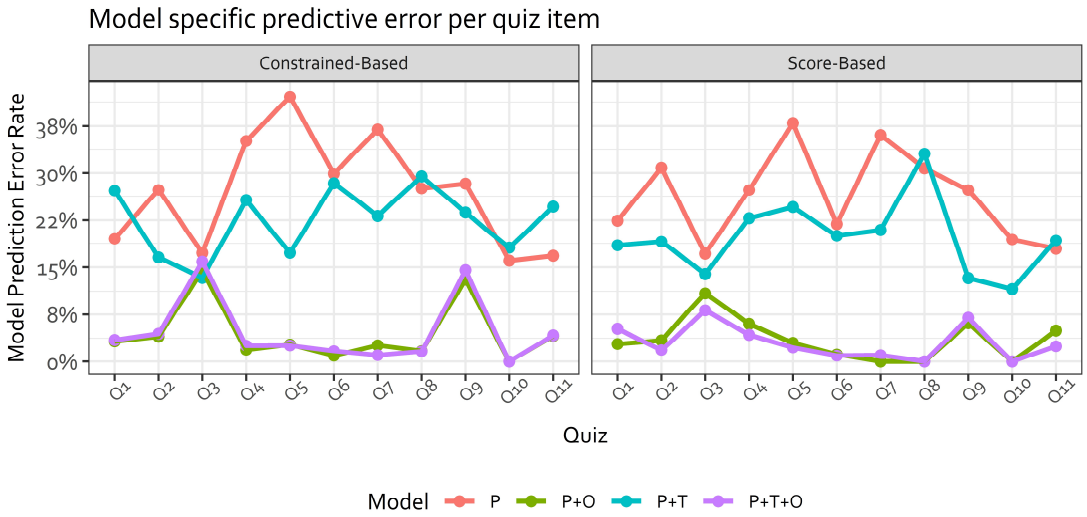

Figure 3. Model accuracy for learned Bayesian network structures using Hill-Climbing and IAMB algorithms

Further analysis by performance level revealed that the **P+O** model was more accurate in predicting very poor performers who had a final score of three or less out of eleven than any of the other models, indicative that poor performance might be highly reliant on cumulative previous tries at the quiz than time on task. The accuracy of the prediction of the models proved consistent when sensitivity of expertly specified model was compared to learning the structure of a Bayesian network using a 10-fold cross-validated hill-climbing and incremental association constraint-based exemplar algorithms (Figure 3). This demonstrates that with the given learning data from LIFE, whether through expertly-specified or data-driven learning of Bayesian structure, between time-on-task and previous opportunity on quiz, performance given previous cumulative tries is the most accurate prediction of the learner's current and future performance. A combination of both opportunity and time do not offer any substantive added prediction accuracy over just using opportunity and performance in both expertly-specified and data-driven network structure specification. Results from sensitivity analysis using model averaging approach to build a network containing only the significant arcs based on the strength of the probabilistic relationships expressed by the arcs of a Bayesian network demonstrated that models P+T and P+O performed better at

detecting individual probabilistic relationships between the quizzes in the expertly-specified models than P+T+O, and P only model (Table 2).

	P	P+T	P+O	P+T+O
Score[1]	-1068.080	-2525.140	-1962.380	-3670.480
Score[2]	-886.870	-2616.580	-2213.780	-4155.800
BF[3]	-126.070	81.830	226.810	423.830
AUC[4]	0.167	0.641	0.592	0.431

Note: Bayes Factor (BF) computed from pairwise comparison of expert-specified model versus learned model-averaged network structures. [1]***Score*** *= BIC score for expert-specified models.* [2]***Score*** *= BIC score for the model-averaged learned network models* [3]***BF*** *= Bayes Factor, quantifies the support for a model over another (1-3: Barely evident, 3-20: Positive, 20-150: Strong, >150: very strong).* [4]***AUC*** *= Area Under Curve, evaluation of how well the network detects probabilistic relationships that form individual arcs*

Table 2. Summary results from model-averaging sensitivity analysis

Given that, model averaging is meant to produce a straightforward model choice criteria and less risky predictions, **P+O** was deemed to be most informative in prediction of individual user skill mastery based on performance on two grounds: (1) **P+T+O** and **P** models had very weak strength of probabilistic relationships with an AUC of less than 0.5, (2) The Bayes factor in **P+T** (BF=81.3) and **P+O** (BF=226.8), demonstrated that the strength of evidence for use of these expertly specified models over averaged models are *strong* and *very* strong respectively in spite of better BIC scores from their equivalent averaged models. The cut-off thresholds for Bayes factor are explained in detail elsewhere (Kass et al., 1995). In summary, efforts to provide adaptive feedback within LIFE platform to assist learning progression based on predicted levels of skill mastery are most accurate where performance is conditioned on opportunity at learning task, in comparison to where performance is considered alone or where performance is conditioned on time on task. Evidence from the LIFE data analysed suggests that performance conditioned on both time on task and opportunity on task are not substantively more informative than just using performance conditioned on opportunity at learning task.

Our findings suggest that, within smartphone-based clinical training platforms for LMICs, the number of previous tries on a quiz is the most influential metric in determining cascading of instructional support, in addition to cumulative performance, not time on task. They also indicate that more student-step metrics -which tend to design complex novel learning models- might not necessarily increase accuracy of skill mastery prediction where BNs are used. Such metrics might reflect how learners tend to pace themselves on platforms like LIFE where use of learning theory -e.g. goal-achievement theory or self-determination theory etc.- might help provide context for how these metrics reflect learner-centred adaptivity and lead to better designs of skill-mastery learner models (Vandewaetere et al., 2013). Theories might also help elucidate on why opportunities and not time are more predictive of the performance of individual learners, and why a combination of both might not be as informative as using them separately. What is clear from our findings is that, a model of learning that gives pre-eminence to learning opportunity as a means for skill mastery through smartphone-based clinical training might be better suited for improving learning outcomes than one that might be time-intensive in this context for clinical training domain. In this regard, for this context, extending approaches similar to low-dose-high-frequency models of learning (Atukunda et al., 2017) by personalising instructional support (e.g. feedback content) based on opportunity at learning task in smartphone-mediated self-directed learning might be more effective in producing higher learning gains and skill mastery. This would imply that in training healthcare providers, in their current learning session, linking feedback support to the number of previous opportunities they had to attempt a learning task conditioned on their cumulative performance on previous learning tasks might precipitate learning gains. This is can be empirically evaluated in future research work in efforts to build an evidence base for adaptive learning in clinical training for low-income settings.

LIMITATIONS

While **P+O** is the model of choice, the low AUC score reported in table 2 which represents the strength of the model's probabilistic relationships is disconcerting. This implies that while the models might have accuracy in predicting step-wise learner performance, the strength of the probabilistic relationships between the quizzes and the metrics (time and task) are moderate at best. This might be due to the low numbers of observations analysed in the whole study in general but given that it is the first study of its kind looking at the utility of game metrics in prediction of skill-mastery for clinical training in low-income settings, it sheds light into a previously underexplored topic from this context. Additionally, the low predictive accuracy for poor performers –who might best benefit from adaptive support while learning- means that this learner modelling approach raises questions for how best to target those poor performers (Figure 4). While opportunity is a better metric for overall performance prediction (and does have moderately good predictive accuracy for half the quizzes of poor performers on LIFE platform), it does raise questions on whether there is another metric that might best predict poor performance within LIFE, which is not directly captured in LIFE as currently constituted.

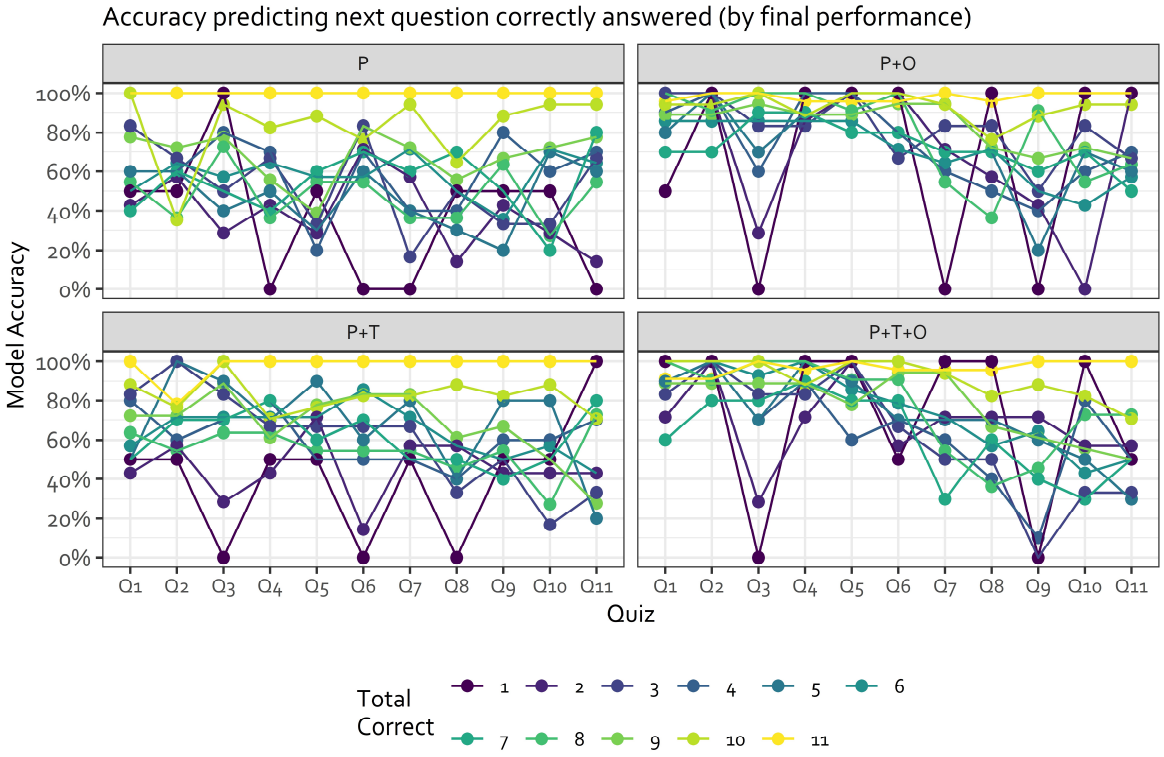

Figure 4. Predictive accuracy by level of performance.

While this study's sample is hardly generalisable, its inclusive constitution (from students to consultants, in all clinical cadres) makes it highly informative as it is the only realistic data source on how adaptive emergency care training on smartphone platforms might be designed to be delivered to health workers in low income settings. Another key challenge for future work in using BNs is in representation of multiple Knowledge Components (KCs), and the inherent dependency among them. In this study, the learning scenario used was assumed to represent one KC, but in future work, it would be necessary to ascertain if and how BNs can deal with multiple dependent KCs when modelling learners' skill-mastery.

CONCLUSIONS

Performance conditioned on opportunities at attempting learning task offer the highest increase in predictive accuracy of student skill mastery at +15.6%, compared to conditioning performance on time on learning task (+4.9%) or conditioning performance on time on learning task and opportunities at learning task (+12.5%). These findings were obtained through use of cross-validated Bayesian network prediction modelling of LIFE learning data from study participants representing all expertise levels, and in all clinical cadres. The models had modest discriminative and calibration strength in terms of probabilistic relationships used in forming the Bayesian network structure but performed much better in predicting skill mastery where LIFE game metrics were used compared to just using the traditional performance level. Our findings indicate that for low resource contexts such as SSA, extending self-directed learning approaches utilising low-dose-high-frequency models of learning on smartphones by personalising instructional support (e.g. feedback content) based on opportunity at learning task might be more effective in producing higher learning gains and skill mastery. Future work will explore this further (Tuti et al., 2019).

ACKNOWLEDGEMENTS

Funds GCRF's Intelligent Support Project awarded to NW supported this work. Additional funds from Grand Challenges' Saving Lives at Birth awarded to CP supported development and testing of the LIFE game.

REFERENCES

Anyangwe, S., et al. (2007). Inequities in the Global Health Workforce: The Greatest Impediment to Health in Sub-Saharan Africa. *International Journal of Environmental Research and Public Health, 4*(2), 93.

Atukunda, I. T., et al. (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., et al. (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

Barteit, S., et al. (2019). E-Learning for Medical Education in Sub-Saharan Africa and Low-Resource Settings. *Journal of medical Internet research, 21*(1).

Bergeron, B. (2006). Developing Serious Games. Game Development Series. Charles River Media. *Inc., Massachucetts*.
Bollinger, R., et al. (2013). Leveraging information technology to bridge the health workforce gap. *Bulletin of the World Health Organization, 91*, 890-892.
Botha, A., et al. (2015). A Teacher Tablet Toolkit to meet the challenges posed by 21st century rural teaching and learning environments. *South African Journal of Education, 35*(4).
Chaudhury, S., et al. (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.
Chi, M., et al. (2011). Instructional factors analysis: A cognitive model for multiple instructional interventions.
Claeskens, G., et al. (2008). Model selection and model averaging. *Cambridge Books*.
Conati, C. (2010). Bayesian student modeling *Advances in intelligent tutoring systems* (pp. 281-299): Springer.
Couper, I., et al. (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
Edgcombe, H., et al. (2016). Enhancing emergency care in low-income countries using mobile technology-based training tools. *Arch Dis Child*. doi:10.1136/archdischild-2016-310875
Feyzi-Behnagh, R., et al. (2014). Metacognitive scaffolds improve self-judgments of accuracy in a medical intelligent tutoring system. *Instructional science, 42*(2), 159-181.
Greenhalgh, T. (2001). Computer assisted learning in undergraduate medical education. *Bmj, 322*(7277), 40-44.
Irimu, G., et al. (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
Kass, R. E., et al. (1995). Bayes factors. *Journal of the american statistical association, 90*(430), 773-795.
Lacave, C., et al. (2018). Learning Analytics to identify dropout factors of Computer Science studies through Bayesian networks. *Behaviour & Information Technology, 37*(10-11), 993-1007.
Ma, W. (2017). *Intelligent Tutoring Systems and Learning Outcomes: Two Systematic Reviews.* Education: Faculty of Education.
Mwaikambo, L., et al. (2016). Leveraging open-source technology and adapting open eLearning content to improve the knowledge and motivation of Ghana's rural nurses. *Knowledge Management & E-Learning: An International Journal, 8*(1), 55-67.
Pearl, J. (2014). *Probabilistic reasoning in intelligent systems: networks of plausible inference*: Elsevier.
Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert systems with applications, 41*(4), 1432-1462.
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., et al. (2017). Predictive effects of temperament on motivation. *International Journal of Educational Psychology: IJEP, 6*(2), 148-182.
Reckase, M. D. (2009). Multidimensional item response theory models *Multidimensional Item Response Theory* (pp. 79-112): Springer.
Romero, M., et al. (2011). Quality of e-learners' time and learning performance beyond quantitative time-on-task. *The International Review of Research in Open and Distributed Learning, 12*(5), 125-137.
Scutari, M. (2009). Learning Bayesian networks with the bnlearn R package. *arXiv preprint arXiv:0908.3817*.
Scutari, M. (2017). Hybrid networks (mixed discrete and continuous). *Creating custom fitted Bayesian networks.* Retrieved from http://www.bnlearn.com/examples/custom/
Silver, L., et al. (2018). Internet Connectivity Seen as Having Positive Impact on Life in Sub-Saharan Africa. *Pew Research Center - Global Attitudes and Trends.* Retrieved from https://perma.cc/34DM-NS3W
Sousa A., et al. (2013). *Transforming and Scaling up Health Professional Education and Training*. Retrieved from Geneva, Switzerland
Templin, J., et al. (2010). *Diagnostic measurement: Theory, methods, and applications*: Guilford Press.
Tuti, T., et al. (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 Res Protoc, 8*(7), e13034. doi:10.2196/13034
UNICEF. (2013). Levels & Trends in Child Mortality. Report 2013. New York, USA, 2013.
University of Oxford. (2016). Life-Saving Instructions for Emergency (LIFE). Retrieved from https://oxlifeproject.org/
Vandewaetere, M., et al. (2013). Adaptivity in Educational Games: Including Player and Gameplay Characteristics. *International Journal of Higher Education, 2*(2), 106-114.
Vanlehn, K., et al. (2001). Bayesian student modeling, user interfaces and feedback: A sensitivity analysis. *International Journal of Artificial Intelligence in Education, 12*(2), 154-184.
Veredas, F. J., et al. (2014). A web-based e-learning application for wound diagnosis and treatment. *Computer methods and programs in biomedicine, 116*(3), 236-248.
Wong, V., et al. (2015). Adaptive tutorials versus web-based resources in radiology: A mixed methods comparison of efficacy and student engagement. *Academic radiology, 22*(10), 1299-1307.