Computers and Education: Artificial Intelligence 3 (2022) 100106

Contents lists available at ScienceDirect

Computers and Education: Artificial Intelligence

journal homepage: www.sciencedirect.com/journal/computers-and-education-artificial-intelligence

Accelerating online learning: Machine learning insights into the importance of cumulative experience, independence, and country setting

Nora A. McIntyre *

Southampton Education School, University of Southampton, UK

ARTICLE INFO

Keywords:
Machine learning
Online learning
COVID-19
Country comparison
Developing expertise

ABSTRACT

Cumulative experience is important for developing expertise through in-person learning, along with country setting and gender, but evidence is limited the role of these features in online learning. Yet, COVID-19 has catalysed the centrality of online learning, such that the efficacy of online learning is now highly relevant. Although the Pandemic triggered a surge of self-report and literature review research on stakeholder perceptions of online learning, less educational research has used big data to understand online learning. Therefore, the present research mined online learning data to identify features that are important for developing expertise in online learning. Data mining of 54,842,787 initial data points from one online learning platform was conducted by partnering theory with data in model development. Following examination of a theory-led machine learning model, a data-led approach was taken to reach a final model. The linear regression model was regularised with the Lasso penalty to enable data-driven feature selection. Twenty-six features were selected to form an extreme gradient boosting model that underwent hyper-parameter tuning. All cross-validation adopted the grid search approach. The final model was used to derive Shapley values for feature importance. As expected, cumulative experience, country differences, low-and-middle-income country status, and COVID-19 were important features for developing expertise through online learning. The data-led model development resulted in additional insights not examined in the initial, theory-led model: namely, the importance of meta-cognition and independent learner behaviour. Surprisingly, no male advantage was found in the potential for expertise development through online learning.

1. Introduction

Cumulative experience is an integral part of developing expertise (de Bruin et al., 2007). With increasing deliberate practice, learners are increasingly likely to master subject-related knowledge and skill, as demonstrated by their learning outcomes (Sternberg, 2001a). Online learning is an environment that has gained prominence for developing expertise, including that in mathematics. However, there is strong potential for between-country differences in the extent to which learners can develop expertise through advanced and demanding educational technology such as online learning (Srite & Karahanna, 2006). This is particularly true when online learning must take place individually, in the home, and without teacher presence — as was the often case during COVID-19 (Dhawan, 2020). Yet, it may be that the pandemic has catalysed learners' opportunity for online mathematics learning such that the cumulative experience prompted marked improvements in mathematical expertise, especially in the context of the particular online platform (OECD, 2020). Nevertheless, a female disadvantage can be expected, especially in view of home-related pressures placed upon girls more than boys, especially in low-and-middle-income countries (Liu et al., 2020). Accordingly, it was the goal of the present research to examine the role of cumulative experience, country, COVID-19, and gender in the potential for developing expertise through online learning.

1.1. *Cumulative experience as deliberate practice for developing expertise*

Cumulative experience[1] is the total experience of an individual learner in one domain and is one dimension of learner expertise (de Bruin et al., 2007). Sternberg (1999) contends that indicators of academic ability to be measures of developing expertise, which require

* Southampton Education School, Her postal address is Building 32, University Rd, Highfield, SO17 1BJ, Southampton, UK.
E-mail address: N.McIntyre@soton.ac.uk.

[1] Although the literature discusses cumulative experience that can be harnessed via group learning (e.g., Reagans et al., 2005), the individual's cumulative experience is the focus of the present research.

https://doi.org/10.1016/j.caeai.2022.100106
Received 6 June 2022; Received in revised form 31 October 2022; Accepted 1 November 2022
Available online 10 November 2022
2666-920X/

deliberate practice that is reflective and focused (see also Ericsson et al., 1993). In other words, performance tests are merely snapshot indicators of time-bound learning progress rather than absolute indicators of giftedness or ability (Sternberg, 2001a). Therefore, to become experts (Sternberg, 2001b; see also Ericsson & Lehmann, 1996), learners need to engage in deliberate practice (Ackerman, 1988; Kulasegaram et al., 2013): that is, planfully designed activities that are responsive to the learner's progress, whilst the learner is encouraged to have intent on improvement (Ericsson, 2018). Academic expertise is developed through cumulative experience of such activities.

Skill-related expertise is distinct from academic expertise (i.e., general intelligence; Cianciolo et al., 2006, pp. 613–632). It is task-specific and contextually situated (Chi et al., 2014; Sternberg, 2001a). The ecological view of expertise emphasises the importance of tacit knowledge, which are all-essential to the successful manifestation of explicit knowledge: as expertise develops, time- and place-specific knowledge are increasingly applied to a specific online environment with improving results (Cianciolo & Sternberg, 2018). Thus, through cumulative experience on each online platform, online learners develop expertise that is specific to the context and skills relevant to specific online learning environments, which indirectly expresses itself via cumulative improvement in learning outcomes.

Cumulative experience with individual online learning has been shown to make a significant impact on subject knowledge test performance (Amarius & Fredriksson, 2021), but this was limited to an investigation on second-language learning. Much like other recent students on online learning, this research involved a small sample (n = 87) and learner-reported online learning experiences (see also Yates et al., 2021). In fact, most studies have focused on the Higher Education learner population (e.g., Al-Salman & Haider, 2021; Biswas et al., 2020; Jiang et al., 2021; Schlenz et al., 2020). The present research focuses on the importance of cumulative experience, in particular, for developing expertise online using a much bigger sample (n = 54,842,787) of school-aged learners via objectively recorded online learning experience from platform analytics.

1.2. Country setting and online learner expertise

Country is another important factor in learners' potential to develop expertise through online learning. Learners' expectations of technology have shown between-country variation. Whereas East Asians (i.e., Koreans) use social media to consolidate and sustain relationships, North Americans use it to seek new connections (Kim et al., 2011). In fact, cultural dispositions have been found to predict technological readiness, with countries high in uncertainty avoidance (e.g., Thailand) showing greater apprehension towards technology adoption, in contrast to those that are low in uncertainty avoidance (e.g., UK) who explore innovations more readily (De Mooij & Hofstede, 2011). Additionally, feminine cultures (e.g., Thailand and Kenya) are more likely than masculine cultures (e.g., UK) to rely upon subjective norms to broach technological innovations. Likewise, cultures low in power distance (e.g., UK) are more likely to adopt technology than those high in power distance (e.g., Kenya and Thailand; Srite & Karahanna, 2006). Moreover, collectivist cultures (e.g., Kenya and Thailand) tend to adopt or abandon technology use depending on social norms, in contrast to individualistic cultures (e.g., UK) whose technology use relates to individual characteristics such as age, gender, and experience, as well as social norms (Venkatesh & Zhang, 2010; see Hofstede, 1986 for cited country profiles). The importance of cultural dispositions features in motivations that underlie online learning, too. Learners showed greater need for understanding of task requirements (prior to task completion) with increased uncertainty avoidance, as displayed by Chinese learners when compared with Malaysian learners. They would be more interested in pursuing their own growth and development during online learning when individualism is high as it is among Malaysian learners, as compared with when collectivism is high as it is among Chinese learners (Bing & Ai-Ping, 2008).

Additionally, income differences between countries result in access-related inequalities. As there are specific countries in the Global North that spearhead advancements at the digital frontier, so are there specific countries with distinctly limited access to technology due to systemic constraints as well as within-country inequalities (OECD, 2019). Indeed, online learning environments such as intelligent tutoring systems (such as the one presently investigated) are situated within an ecological system wherein country-level factors dictate learners' chances of accessing online learning opportunities reliably, or at all (Nye, 2015). Reasons for between-country disparities include differences in the agenda and priorities of national gatekeepers, the local infrastructure, and the local culture such as school technological support, teachers' technological readiness, as well as home resources and restrictions. Thus, a disadvantage can be expected among low-and-middle-income countries (LMICs) in expertise development through online learning.

Although the cited research has considered country as a factor in online learning, the present research makes a significant contribution via a three-way country comparison with use of big data whilst accounting for cumulative experience within one platform — and doing so using objective measures rather than learner-reported perceptions (Launois et al., 2019). Additionally, the present research accounts for the potential impact of between-country income differences by taking these into consideration — and doing so through a large-scale quantitative lens for perhaps the first time rather than adopting a widely used qualitative and subjective (e.g., King et al., 2018), or single country (Adam, 2020), perspective.

1.3. The 'covid effect' on developing online learning expertise

The COVID-19 pandemic occurred during the life of this research and was expected to have had a dramatic effect on the amount of online learning that students engaged with. Online learning saw an unprecedented and unforeseeable surge. However, Covid combined with country to predict important between-country differences in readiness for online education. Some countries responded to the Pandemic by promptly setting up new online educational platforms (e.g., China and Singapore), whereas others used open educational resources or focused on using content already available that could be delivered online via existing digital platforms. Some countries put in place teacher professional development for online education (e.g., Italy, OECD, 2020), and every country was distinct in the duration of school closures (OECD, 2021).

Moreover, since COVID-19, online learning resources have increased or decreased in their potential as educational supports, depending on learners' infrastructure. Although supports for online learning was reported as a challenge regardless of country income, learners in high income countries could apply for relevant support, such as in the UK where laptops or alternative printed resources were distributed (OECD, 2020). Connectivity has been a pertinent factor among learners in high income countries, and even more so in low-and-middle-income countries (Dhawan, 2020; see also Aboagye et al., 2021; Baticulon et al., 2021; Chang & Fang, 2020; as cited in Mseleku, 2020). Covid almost undoubtedly increased online learning in countries where infrastructure met some threshold of strength and, in turn, prompted an improvement of learners' readiness for online learning. However, Covid equally decreased online learning where infrastructure was weak such that readiness for online learning was not touched, let alone improved at all. Accordingly, it was deemed essential for this global event to be included and reported in the present analysis.

Although the Pandemic prompted numerous within-pandemic review (e.g., Dhawan, 2020) and perception (e.g., Yates et al., 2021) studies on forced online learning through emergency school closure, little-to-no research has investigated the 'Covid effect' via longitudinal, objective data that accounts for change over time. This is another significant contribution of the present research.

1.4. The gender effect on online mathematics learning

Gender differences are well established in mathematics learning. Although girls have typically been found to achieve more highly than boys (for a meta-analysis, see Voyer & Voyer, 2014), they have also shown a greater tendency towards math anxiety than boys (Casad et al., 2015). Bring income inequalities into the equation and the female disadvantage in math-related learning becomes stark, both in high-income countries (Catsambis, 1994) and in LMICs (e.g., R. Liu et al., 2020) — not least because of culture-specific views of girls' responsibilities in the home on top of the greater strain faced more generally by parents in LMICs (Putnick & Bornstein, 2016), although there are between-country differences in gender norms across LMICs.

Similar patterns have been found in relation to educational technology, with a female advantage in learner engagement reported for game based learning — along with a call for this pattern in game based learning to be harnessed to address the wider gender disparities in Science Technology Engineering and Mathemathics (STEM) education (Khan et al., 2017). Correspondingly, girls have been found to spend longer on, and to benefit more academically from, technology supported learning (Reychav & McHaney, 2017). Furthermore, gender differences have been reported with regard to preferences in and the use of online environments (Lin et al., 2017). However, as with in-person environments, boys tend to display greater academic confidence as well as satisfaction with the digital (Kilbourne & Weeks, 1997; Reychav & McHaney, 2017) and online (Bruns, 2010; Guiller & Durndell, 2006) learning environment.

Relatedly, gender differences in engagement with online learning among LMICs were highlighted by the Pandemic: girls were found to persist with distance education notably (though not significantly) more than boys in a number of LMICs including Bangladesh and Jordan, although the opposite gender pattern was found in Ethiopia (Jones et al., 2021). Nevertheless, the general consensus is that a female disadvantage prevails in terms of educational access and attendance, due to long-standing cultural factors around the world where, when there are domestic burdens to be shared by children, it is normally the girls who must take these on (i.e., parental, teacher, community, curriculum; Jafree, 2021): a pattern that has been observed in online learning (e.g., FutureLearn, Tronstad, 2021). The author is only aware of theoretical (Mathrani et al., 2021) and review (Gnanadass & Sanders, 2018) studies on gender disparities in online learning, such that the present research makes a unique contribution by examining potential gender differences found in online learning analytics.

1.5. Machine learning for online learning research

Data science has fast become relevant with the prominence of online learning. Because such data was collected without an original intention for research, and because of the number of data points typically confronted (initial n = 54,842,787 in the present study), it should be treated as big data. In order to make the most of potential insights from big data, data mining can be employed to uncover unexpected insights and relationships that exist in the data (Ratner, 2011). Machine learning is particularly relevant for handling real-world data which do not necessarily meet strict statistical assumptions: data from online learning platforms are one such example. More than ever before, a vast array of machine learning analytic solutions are now available to handle the complexities of big data from the real-world, to optimise such data, and to provide interpretable insights for applied research, especially when heavily informed by domain (or theoretical) expertise: a detailed repertoire of these capabilities and resources is demonstrated in the present research.

1.6. Research objectives

The present study pioneered online learning research by examining the importance of cumulative experience in online learning as objectively measured by learning analytics. It further pioneered by examining features reflecting social inequalities: namely, country culture and income, as well as gender. Moreover, the present research harnessed the affordances of two traditions in data science (Chen et al., 2020; Rosé et al., 2019): the theory-led and the data-led approach to data science. To do this, model development started with features derived from theory before selecting taking an entirely data-driven approach to feature selection, and only doing so when theory-led features did not provide enough explanation. Thus, rich insights were anticipated by implementing this dual approach to online learning data.

The theory-led approach was taken by selecting themes and related features (i.e., predictors) that are likely to play an important role in online learning: these were *cumulative experience*, *country setting*, *COVID-19*, and learners' *gender*. Accordingly, the theory-led hypotheses were as follows:

Hypothesis 1. Cumulative experience will develop expertise in online learning within one particular platform and, as such, will predict increasing learning outcomes.

Hypothesis 2. Given country differences in the infrastructural, academic, and socioemotional supports for online learning, the country setting will be found to relate to online learning outcomes.

Hypothesis 3. The online learning necessitated by the emergency school closures during COVID-19 are likely to predict a spike in online learning outcomes.

Hypothesis 4. Given the tendency for boys to receive stronger educational support than girls, a male advantage was expected in learning outcomes.

Hypothesis 5. There was an additional research expectation for data-led feature selection to identify unanticipated features that have importance in predicting online learning outcomes.

2. Method

The full research process is visualised in Appendix 1.

Secondary data analysis was conducted on data from an online learning environment by Maths-Whizz called Whizz Education.[2] Whizz Education is an online intelligent tutoring system covering 22 age-appropriate topic areas in Mathematics (see Supplementary Material 1), which break down into 1222 lessons, each with its own learning objectives (Supplementary Material 2). Additionally, a sample dataset is available in Supplementary Material 3.

Log files shared from this platform spanned the years 2016–2020 inclusive. Each data point from this platform represented one completed lesson, which involved exercise and a test. For each lesson datapoint, an anonymised pupil ID was provided, and each pupil was linked with an anonymised school ID. Although the laptop was data science ready (see

[2] To provide further information about the platform, it was used by partnering schools as an additional learning resource for students, though not normally as part of metrics for students' academic performance in schools. Accordingly, students will at times have learned the topics in class before attempting the questions on the platform. However, students of partnering schools are also able to access the platform at home and home access to the platform will have applied to all uses of the platform during the Pandemic's school closures. During platform use, it was not mandatory for students to complete all the questions in order to progress onto subsequent exercises or lessons: they simply had to attain sufficient marks and exceed a threshold (i.e., pass) in order to attempt the next exercise or lesson — hence the feature, *run mode*, which relates to whether the student has jumped forward or if the 'tutor' had jumped them backwards to revise previous materials. Further details of the platform, including examples of the lessons, can be found via www.whizz.com/maths-games.

Table 1
Feature dictionary, with all features initially included in analysis (i.e., Phase 2 model development).

	Feature name	Feature explanation	Feature engineering process
1	*topicId*	Topic identifier (22 topics in total)	None; from log file
2	*mathLevel*	Academic difficulty of the *lesson* (not level of the child). The academic difficulty was framed in terms of the academic age targeted by a lesson and was divided by quarters of a year (i.e., one year divided into 0.25, 0.50, 0.75, and 1.00)	None; from log file
3	*exerciseId*	Within each quarter, exercises were sequenced in order of difficulty and ranged from 100 to 1000 (i.e., 100, 200, 300, etc.), incrementing at intervals of 100 within each quarter then resetting at the next quarter.	None; from log file
4	*stackDepth*	The feature, stackDepth, related to the lesson's mode, with the default value being stackDepth = 1 to signify progression; if a learner failed a default, progression lesson, they would regress to a simpler exercise in to a lesson mode with stackDepth = 2; failing that, the learner would be regressed further to even simpler exercise at stackDepth = 3. If the learner passed the stackDepth = 3 exercise and test, they would move back to complete the exercise and test at stackDepth = 2 then, if they pass that test, return to the lesson at stackDepth = 1.	None; from log file
5	*timeTaken*	How long the learner took to progress from beginning of lesson to the end, including the exercises and test.	None; from log file
6	*questionTime*	How long the learner took to complete the exercise questions.	None; from log file
7	*tutorialTime*	How long the learner took to complete the tutorial as a whole.	None; from log file
8	*totalQuestions*	The number of questions that the learner attempted in that lesson.	None; from log file
9	*lesson_type*	The default progression tutor exercise, regression tutor exercise, replay exercise, tutor test.	None; from log file
10	*total_help*	The number of times help was sought by the learner.	None; from log file
11	*replay*	A summary feature indicating whether the lesson was a standard, progression one, or whether the learner was repeating the lesson for whatever reason.	None; from log file
12	*markedYear*	*2016, 2017, 2018, etc.*	Computed from log file variable, *marked* (e.g., 30/01/2020 07:40)
13	*markedMonth*	*January = 1, February = 2, etc.*	Computed from log file variable, *marked* (e.g., 30/01/2020 07:40)
14	*markedWeek*	*1 to 52 for each calendar year*	Computed from log file variable, *marked* (e.g., 30/01/2020 07:40)
15	*since_covid*	*1 = 2020; 0 = 2016 to 2019*	Computed from log file variable, *marked* (e.g., 30/01/2020 07:40)
16	*Male*	Dummy variable (or one-hot coding)	Dummy generated from *gender* (original variable)
17	*Female*	Dummy variable (or one-hot coding)	Dummy generated from *gender* (original variable)
18	*play_count*	The total number of lessons completed by each learner	Computed from log file variable, *anonymised_pupil_id* (e.g., 88873931)
19	*indiv_pupil_t*	The cumulative lesson count for each learner	Computed from log file variable, *anonymised_pupil_id*
20	*birthYear*	Year of birth	Computed from log file variable, *date_of_birth* (e.g., 01/01/2006)
21	*birthMonth*	Month of birth	Computed from log file variable, *date_of_birth* (e.g., 01/01/2006)
22	*pupil_ageQuart*	The learner's age in quarters. That is, year + quarter, e.g., 12.25 for 12 years and a quarter; births between January and March were quarter = 0, births between April and June were quarter = 0.25, etc.	Computed from log file variable, *date_of_birth* (e.g., 01/01/2006)
23	*mathAbility*	Learner academic age. For example, a learner with pupil_ageQuart = 12.25 years who is attempting a lesson with mathLevel = 9.25 will be showing the mathAbility of +3 years.	Computation: *pupil_ageQuart - mathLevel*
24	*Kenya*	Dummy variable (or one-hot coding). 1 = Kenya, 0 = UK or Thailand	Computed using *Kenya* data file as reference
25	*UK*	Dummy variable (or one-hot coding). 1 = UK, 0 = Kenya or Thailand	Computed using *UK* data file as reference
26	*LMIC*	Dummy variable (or one-hot coding). 1 = LMIC (Kenya or Thailand), 0 = HIC (UK)	Computed from *Kenya* and *Thailand*
27	*InCountryDep*	1 (least deprived) to 3 (most deprived) using country-specific deprivation codes as applied at school level. Missing data were replaced by the sample-level mean (i.e., 2.08) and rounded to the nearest integer (i.e., 2). Additional notes: The UK deprivation status was calculated using the Index of Multiple Deprivation 2019 (IMD2019, Penney, 2019). This is a decile index which was split into three bins using the Pandas cut function. The Kenyan deprivation status came as a three-level feature, with urban being the most well-resourced, rural as middling, and hardship as the least well-resourced learners. The Thai deprivation stati came as a three-level feature: private or independent schools were rated to be the most well-resourced, followed by provincial public schools, and rural public schools as the least well-resourced.	Computed from log file variable, *deprivation*

Apparatus for details), a random subsample of n = 5000 (seed = 1) still needed to be taken from the whole sample of n = 54,842,787, in order to enable the computational processing demands of the large dataset and advanced analyses implemented in this study. Note that the wisdom in random sub-sampling for data mining is recognised in the field (Attewell & Monaghan, 2015; Bouckaert & Frank, 2004; King & Resick, 2014; Ratner, 2011; Sculley & Pasanek, 2008).

2.1. Participants

The Maths-Whizz Tutor by Whizz Education is designed for and accessed by learners aged between 5 and 13 years. The platform is available for use in schools and in the home: during the COVID-19 pandemic, data was collected solely from the home context for a significant period, whereas it was collected from both schools and homes before onset of the pandemic. Within the random subsample of n = 5000, n = 2581 were male and n = 2418 were female.

The platform is available to multiple countries, three of which were sampled in the present analyses: namely, Kenya, Thailand, and the UK. Thus, two low-and-middle-income countries (LMIC; i.e., Kenya and Thailand) and one high-income country (the UK) were sampled, enabling between-culture (Kenya vs. Thailand vs. the UK) and between-income-status (Kenya and Thailand vs. the UK) comparisons. Within the random subsample of n = 5000, n = 1755 data points were from Kenya, n = 1128 were from Thailand, and n = 2117 were from the UK.

This data was not publicly available: rather, a partnership was

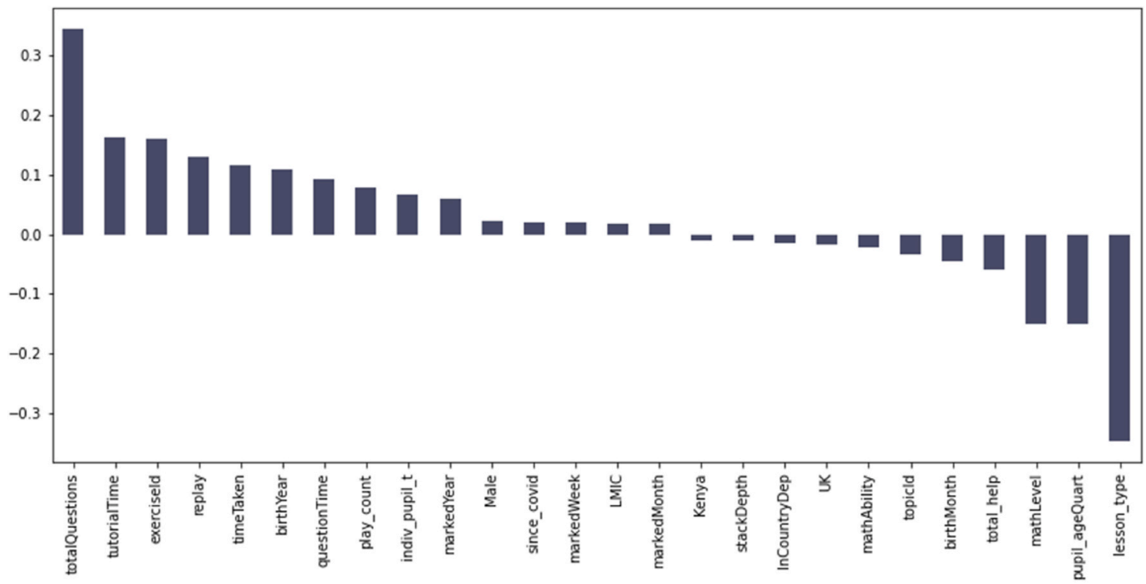

Fig. 1. Correlations between potential features and learning outcome (*lesson_mark*). Transformed data are represented here.

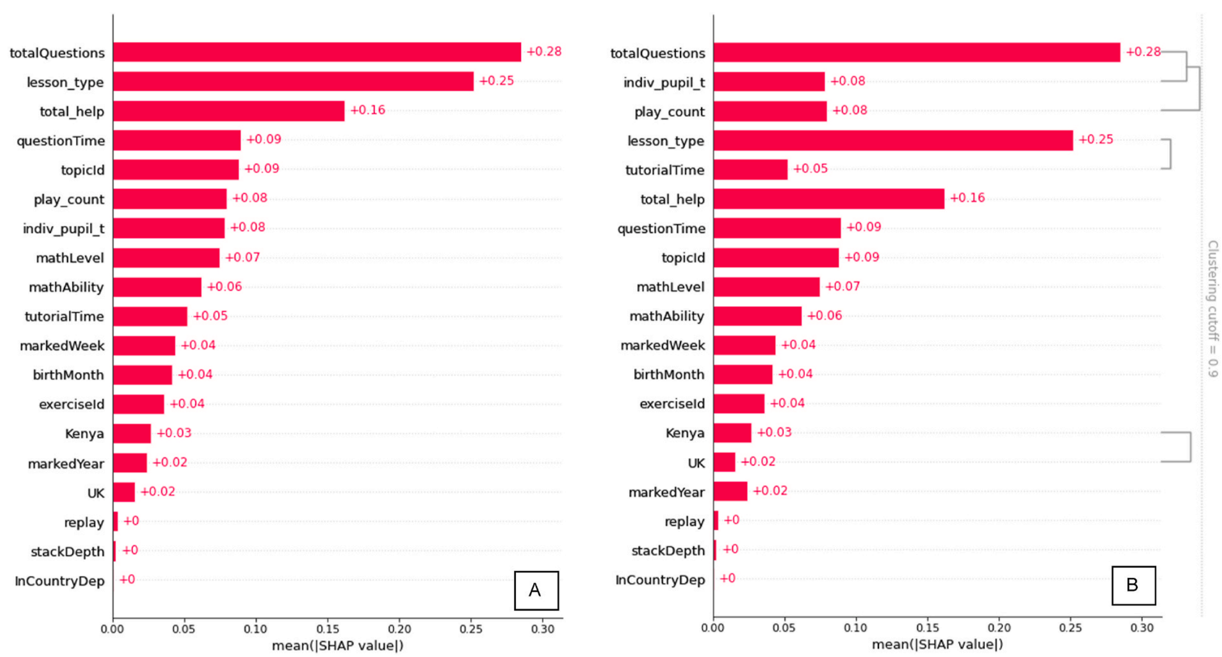

Fig. 2. Bar plot of feature importance of features in the final model, using mean absolute Shapley values. Panel A shows the features ordered from the most important to the least, in the final model. Panel B shows the features are generally ordered in the same way, but with clustering where features are related to each other. Transformed data are represented here.

formed between the present author and Maths-Whizz to conduct the present analyses, with support from Maths-Whizz colleagues to obtain the necessary data for relevant feature engineering. Thus, by agreement to collaborate and following conversation with the company CEO, Maths-Whizz gave consent to participation in this research. Additionally, since this is analysis of secondary data in which no living persons were identifiable, no institutional ethics approval was required.

2.2. Analytic tools

Data science laptops were used for this analysis: MSI Stealth, NVIDIA RTX 3060 GPU, 16 GB RAM, 2.60 GHz, 500 GB SSD (laptop 1); Dell Precision 7560, NVIDIA RTX A5000, 32 GB RAM, 4.80 GHz, 1 TB SSD (laptop 2). On it, Jupyter Notebook was used during most of the data pre-processing and feature engineering as the Python environment of choice before Google Colab was locally hosted for further pre-processing and all of the analyses reported in this article. The notebook for the present data analyses is available upon request from the author.

The Python libraries used for the present analyses include Vaex as the platform for handling the full dataset which was stored and manipulated in hdf5 format. Some pre-processing of the whole dataset was applied to the whole dataset in Vaex and as hdf5 files for speed (Breddels & Veljanoski, 2018), but more major data manipulation (e.g., data joining, conditional selection for generating new features) was performed in Pandas (McKinney, 2010). During feature generation, Numpy (Harris et al., 2020) was used in parallel with both Vaex and Pandas. Data was

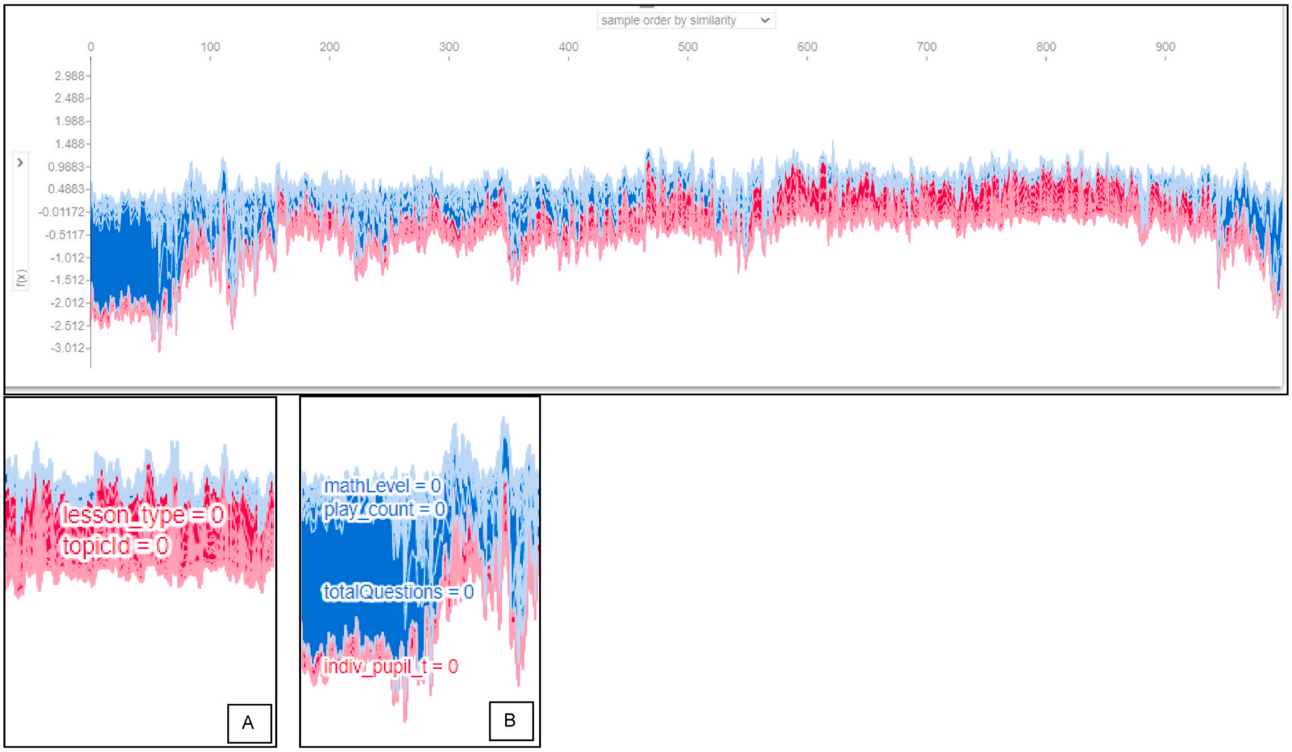

Fig. 3. Collective force plot showing the overall effect of all features included in the final model, using absolute mean Shapley values. As the graph progresses to the right, effects of the most important features for each individual learner are shown. Features that push the prediction higher (to the right) are shown in red, and those pushing the prediction lower are in blue. The x-axis shows participant number, ordered by similarity for this plot. Panel A gives a snapshot of the features that generally reduce *lesson_mark*; Panel B shows a snapshot of features that increase *lesson_mark*. Transformed data are represented here. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

converted to DMatrix format via sklearn (Pedregosa et al., 2011) for memory efficiency and to enable training speed during extreme gradient boosting in XGBoost (Chen & Guestrin, 2016) which was run after linear and regularised regression (see Analysis and Results for details). The sklearn (Pedregosa et al., 2011) library was used to run linear regression, elastic net cross-validation, regression with elastic net penalty, lasso cross-validation, and regression with lasso penalty. Shapley values and related visualisations were obtained through shap.Explainer method from the SHAP library (Lundberg & Lee, 2017).

2.3. Measures

The outcome variable was *lesson_mark* which had a maximum score of 100 for each lesson. This has a maximum of 100 for each lesson. To formula for *lesson_mark* is as follows: lesson mark = 100 x score/total number of questions attempted for this lesson. The questions may have been part of lesson exercises, or lesson tests.

2.3.1. Features

All the features available for selection in this analysis are listed in Table 1: all of these were initially included in Phase 2 for data-led feature selection and model development, whereas only *cumulative experience, country, gender,* and *since_covid* were included in Phase 1 for the theory-led feature selection.

More specifically, In Phase 1, the analytic model was theory-led and based on the author's domain expertise, the established literature for identification of the most important constructs, and the most robust measures in my data as representatives of the most relevant constructs identified from initial data analysis when theoretically significant features were noted. Thus, through a theory-led perspective on the initial data exploration, priority was given to theoretical significance and analytic parsimony. Accordingly, the features were *cumulative experience* (*indiv_pupil_t*[3]), *country* (*Kenya* dummy, *UK* dummy; the Thailand dummy was not needed in analytic model since Kenya = 0 and UK = 0 means Thailand = 1), *since_covid*, and *gender* (Male dummy; the Female dummy was not needed since Male = 0 means Female = 1). The outcome variable was *lesson_mark* throughout model development.

Next, in Phase 2, the data-led approach to model development began with data-led feature selection. The feature selection methods are reported in full, in Appendix 2. To begin with the variables available for feature selection excluded variables that represented the same construct as the target variable, *lesson_mark*. Therefore, *indiv_pupil_t* was excluded. Also excluded were string variables such as marked and date of birth, which served as the bases of engineered features such as *marked_year* and *pupil_ageQuart*. The variables, Country and gender, were made redundant by use of the dummy variables relating to each: namely, *Kenya* and *UK* replaced country and *Male* replaced gender in the analytic model. Thus, 26 variables were available for feature selection during Phase 2, the data-led model development. These were: *topicId, mathLevel, exerciseId, stackDepth, timeTaken, questionTime, tutorialTime,*

[3] With parsimony being a priority in the theory-led model, one feature was chosen for each construct. The feature, *indiv_pupil_t*, was chosen to represent cumulative experience instead of *play_count* for the theory-led model because *indiv_pupil_t* captures learners' change over time and, as such, better represents intra-individual variation than *play_count* which consists of the same value for each of the learner's data points.

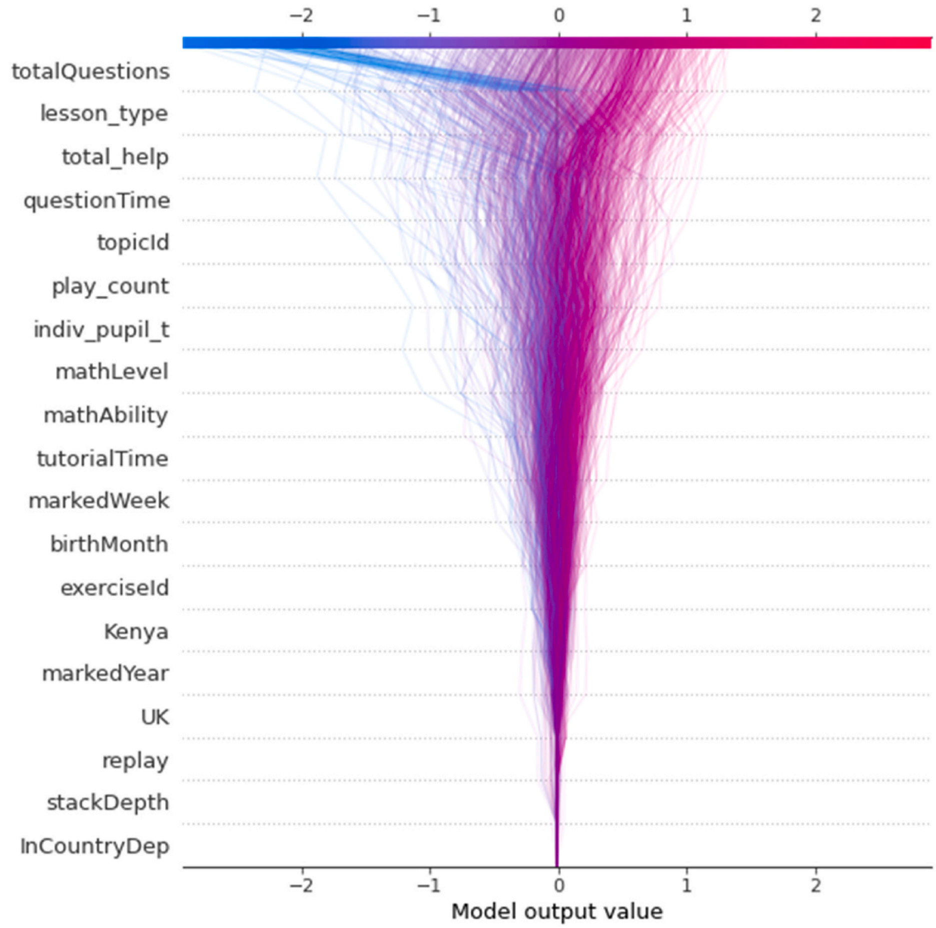

Fig. 4. Decision plot of feature importance, using mean absolute Shapley values. The model output value is the learning outcome (*lesson_mark*). Features that push the prediction higher (to the right) are shown in red, and those pushing the prediction lower are in blue. The fainter a line, the fewer learners it represents. Transformed data are represented here. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

totalQuestions, lesson_type, total_help, markedYear, markedMonth, markedWeek, Male, mathAbility, pupil_ageQuart, birthYear, birthMonth, replay, Kenya, UK, LMIC, indiv_pupil_t, play_count, since_covid, InCountryDep. Again, the outcome variable was *lesson_mark*. Correlations between learning outcomes (*lesson_mark*) and the 26 potential features are shown in Fig. 1.

2.3.2. Outcomes

To assess the value of each feature in explaining the outcome (*lesson_mark*), feature importance analysis was conducted. Feature importance is crucially different from correlations in that, rather than seeking to understand how much of the total variance one predictor explains, feature importance analysis assesses each feature (or predictor) in relation to all the other features included in the final model for explaining the outcome. To report analytic insights from the final model, Shapley values were computed for interpretable feature importance, followed by Shapley interaction values in order to understand between-feature relationships (Aas et al., 2021; Rodríguez-Pérez & Bajorath, 2020). In doing so, analytic outcomes arising from the 'black box' of extreme gradient boosting can be mapped onto substantive concepts and enable theoretical contribution from the present research.

2.3.2.1. Explainable machine learning. Shapley values are one form of feature attribution methods in explainable machine learning. Attribution methods are feature-by-feature evaluations of the feature's contribution to an outcome variable. Note that the explanations do not claim to provide causal attributions, but rather contrastive ones—ones which the most important, select causes of an event relative to other possible causes (or features) available. Thus, interpretable machine learning analysis would not ask 'how' something happened, but rather 'why' it did. Shapley values enable this.

"The goal of SHAP is to explain the prediction of an instance x by computing the contribution of each feature to the prediction. The SHAP explanation method computes Shapley values from coalitional game theory. The feature values of a data instance act as players in a coalition. Shapley values tell us how to fairly distribute the 'payout' (i.e., the prediction) among the features" (Molnar, 2021, Section 9.6.1.). To assess this, each feature in the present analysis was assessed by four axioms: *marginal contribution*, or how much is lost by removing the individual feature from the 'game' (or prediction, Axiom 1); *interchangeable value*, the extent to which the features brings the same amount as other features in which case these features share the same value (or 'reward', Axiom 2); *dummy players*, the extent to which the individual feature contributes nothing, in which case these features are attributed to have zero value (Axiom 3); *multiple parts*, the complexity of the 'game' (or prediction) is taken into consideration and value is attributed according to and across the complexity (e.g., value is not attributed in the same way every time and is reviewed regularly for potential adjustments, Axiom 4). Thus, Shapley values indicate "how to be smart in competitive situations, and how to be fair in cooperative situations" (Noah's Ark 2.0, 2020).

In data science, the Shapley value assumes that every feature contributes to the prediction to some degree. The full prediction by all included features is equally distributed to every feature in the model. Consequently, the Shapley value for a feature is not only an indication of how much an outcome value changes when the feature is added. It also

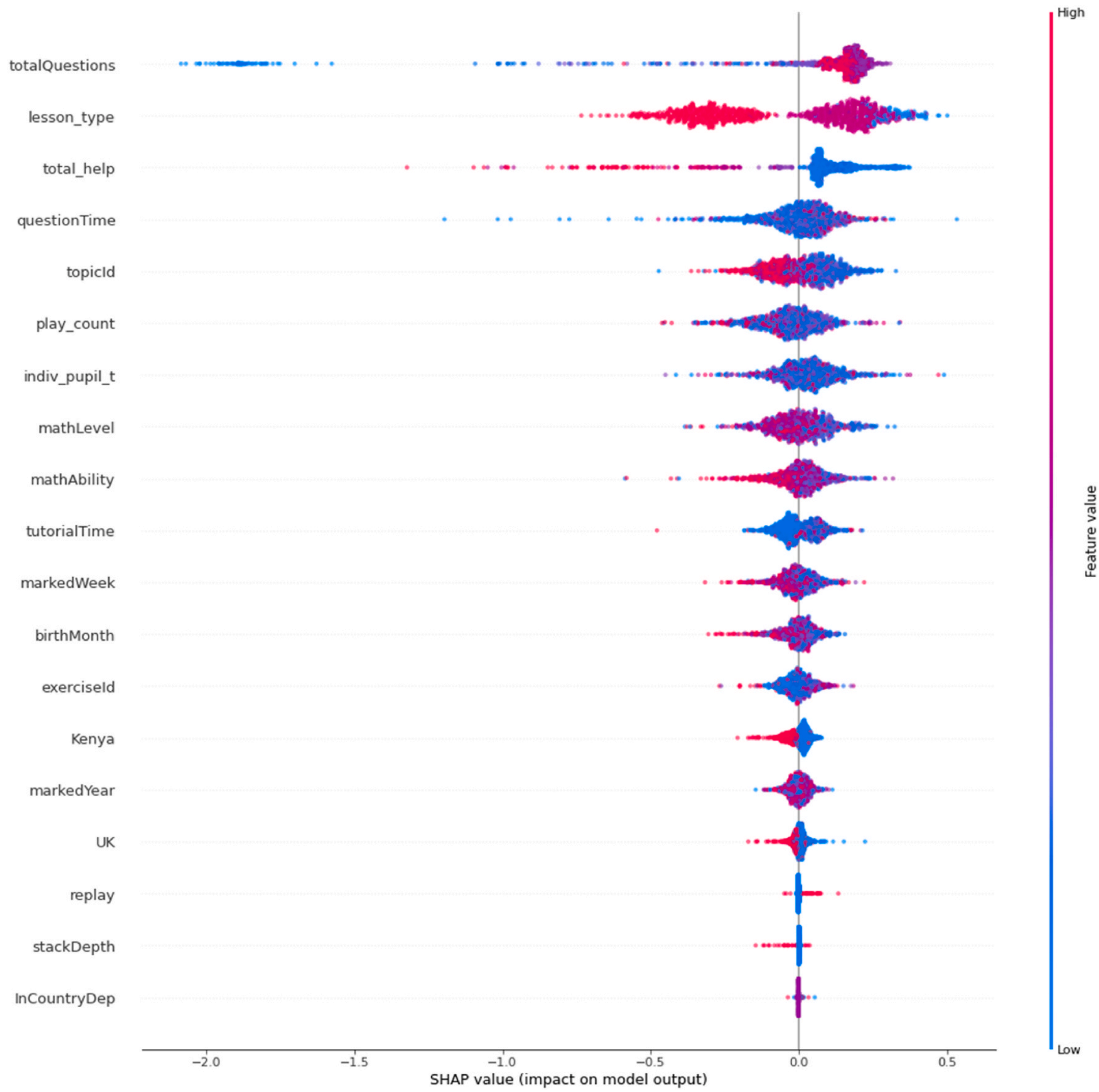

Fig. 5. Summary plot of feature importance in final model, using mean absolute Shapley values. Features that push the prediction higher (to the right) are shown in red, and those pushing the prediction lower are in blue. Transformed data are represented here. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

indicates *how this feature changes all features' average contributions* during the process of the model being built feature-by-feature until it reaches the output of the final model (i.e., the model being interpreted). The actual model prediction, with the current feature included, is compared with these features' mean prediction of the outcome variable. In this way, the Shapley value explains the difference between the single feature's prediction and the *global* average prediction of the outcome value: that is, in relation to all possible coalitions (or all possible groups of features).

One advantage of Shapley values is that they offer uniquely human explanations of models when compared with alternative model interpretation metrics such as Local Interpretable Model-agnostic Explanations (LIME) which target the difference between the prediction and a *local average* prediction: that is, for the specific instance of the prediction in question (Aas et al., 2021). In contrast to such local attribution methods, Shapley Values provide explanations that extend beyond the current, local context and prediction, but reaches globally across the data and across the combinations of features in the model.

Furthermore, absolute Shapley values were used in this research. This is in accordance with simulation outcomes comparing performances of the original Shapley value with the absolute Shapley value for classic machine learning analyses, including regression. The absolute Shapley value outperforms, when it comes to feature importance evaluation (Liu, 2020).

Feature clusters are reported alongside individual feature importance analysis or feature interaction analysis as appropriate. For this, hierarchical clustering is used, whereby features with distance = 0 are redundant (i.e., can replace the other[s] and the model still attains comparable performance [i.e., accuracy]) and those with distance = 1 are independent of each other. The feature clustering bar plot in Fig. 2 (Panel B) shows clustering at distance = 0.9 (Lundberg, 2018). The chosen clustering threshold is near-independent. Therefore, clustering was used to support interpretation of the final model, rather than for model development.

2.3.3. *Performance evaluation*

Performance evaluation during model development was carried out using the root mean squared error (RMSE) and Adjusted R^2 values. In contrast to other metrics (e.g., accuracy, f-scores, AUC – ROC), these measures are appropriate to regression models. The RMSE is the square

Table 2
Data-led model development, containing all features initially included in the data-led model for model development. Coefficients that emerged from the regularised regression with the Lasso penalty when predicting math learning (lesson_mark), in descending order of coefficient size.

	Feature	Coefficient
1	lesson_type	−0.29575
2	totalQuestions	0.240236
3	total_help	−0.2224
4	exerciseId	0.127881
5	mathLevel	−0.12053
6	questionTime	0.082319
7	replay	−0.08203
8	stackDepth	−0.05057
9	tutorialTime	0.047677
10	mathAbility	−0.04253
11	indiv_pupil_t	0.041615
12	Kenya	−0.03517
13	markedYear	0.02604
14	birthMonth	−0.02411
15	play_count	0.015164
16	UK	−0.01038
17	markedWeek	0.009218
18	topicId	−0.00719
19	InCountryDep	−0.00297
20	birthYear	0
21	pupil_ageQuart	0
22	LMIC	0
23	markedMonth	0
24	timeTaken	0
25	since_covid	0
26	Male	0

Table 3
Shapley values for all the features in the final model for predicting *lesson_mark*.

	M	SD	min	max
totalQuestions	0.284809	0.417596	8.84E-05	2.085259
lesson_type	0.251696	0.119971	0.003476	0.7339
total_help	0.162096	0.167613	0.003197	1.322545
questionTime	0.089857	0.101445	4.98E-05	1.195895
topicId	0.088402	0.061777	0.000525	0.472555
play_count	0.079801	0.072169	7.46E-05	0.461696
indiv_pupil_t	0.078404	0.068885	0.000171	0.490393
mathLevel	0.074632	0.061395	0.000112	0.384514
mathAbility	0.062186	0.063527	4.90E-05	0.586204
tutorialTime	0.052018	0.038008	1.17E-05	0.478691
markedWeek	0.044057	0.040454	1.46E-05	0.316812
birthMonth	0.041914	0.04144	0.000146	0.305351
exerciseId	0.035904	0.031779	2.79E-05	0.267847
Kenya	0.026703	0.022954	6.44E-05	0.206519
markedYear	0.024229	0.02035	1.10E-05	0.145523
UK	0.015709	0.019257	2.11E-06	0.223571
replay	0.003474	0.009204	1.36E-06	0.133414
stackDepth	0.002671	0.00997	1.16E-06	0.145441
InCountryDep	0.000987	0.002744	6.55E-06	0.05387

root of the average of the squared difference between the target value and the value predicted by the regression model, whilst the Adjusted R^2 is the squared percentage of variation described the regression line. The RMSE was chosen rather than the MSE, in order for the performance measure to align with that of the response variable, thus improving interpretability of the performance value. The Adjusted R^2 was chosen rather than the raw R^2, in account for the number of parameters and the sample size, thus resolving the impact of the sample size on the performance value. The use of these metrics during model development is fully document in Appendix 2.

2.4. Analysis

The outcome variable was *lesson_mark* throughout model development. Out of the full sample of n = 54,842,787, a random sample of n = 5,000 (random seed = 1) was used in analyses.

Prior to model development, a baseline model was set up with the outcome variable, *lesson_mark*, being predicted by its mean. Following this, model development commenced. During both theory-led (Phase 1) and data-led (Phase 2) model development, simple linear regression models were run first. These were then regularised to adjust for non-linear features and distributions: grid search cross-validation (rather than randomised search cross-validation; Worcester, 2019) was used first with elastic net penalty then with the Lasso penalty as appropriate.[4] Subsequently, extreme gradient boosting (a.k.a. XGBoost, Chen & Guestrin, 2016) was employed to maximise the computational resources available for peak speed and model performance (i.e., predictive performance). XGBoost models underwent automated hyperparameter tuning via grid search cross-validation (Worcester, 2019), followed by final manual hyperparameter tuning.

The outcome of the model development was the final analytic model discussed in the Results below. Its full development record is reported in Appendix 2.

3. Results

The analytic outcomes from the final model are now reported. Overall patterns of the final model are reported first, with plots presented in order of within-feature granularity. Analytic outcomes are then organised by features. The features that were identified for analysis through theory are reported first (i.e., features from the theory-led model; Hypotheses 1 to 4). The features that emerged as most important from data-led development are then reported upon (Hypothesis 5). Finally, feature interactions according to SHAP interaction values are examined. At times, figures will show subsamples of individual feature importance: these are then further subsampled from in the narrative, with particular focus on conceptual contribution to the field from the final model in this analysis.

3.1. Overview of analytic outcomes

The collective force plot for the model (Fig. 3) shows that, among most learners, features in the final analytic model contribute to the increase and improvement of learning outcomes (*lesson_mark*), although some decrease learning outcomes. Panel A suggests that the features potentially contributing to the decrease of learning outcomes (lesson_mark) include *mathLevel, play_count,* and *totalQuestions*. Meanwhile, Panel B shows *lesson_type* and *topicId* contribute to the increase of learning outcomes, as does *indiv_pupil_t* according to Panel A. The subsequent analysis will shed more light on individual analyses.

The decision plot (Fig. 4) provides an overview of feature importance in the final model. It shows the distribution of feature importance and allows some between-feature comparison in terms of importance level and within-feature importance variability. Typically, the most important features are found to also be the most heterogeneous in terms of within-feature importance variability. These are *totalQuestions*, *lesson_type, total_help, questionTime,* and *topicId.*

Similarly, the summary plot (Fig. 5) shows the features in order of importance, but it provides greater granularity regarding the within-feature distribution of feature importance. Indeed, some of the most important features, *totalQuestions*, *total_help,* and *questionTime* show the greatest dispersion. Additionally, some *lesson_types* distinctly push the prediction of learning outcomes higher (in red) whereas others distinctly push it lower (in blue). The narrative below continues with further discussion of *lesson_type* in its effect on learning outcomes.

[4] In the data-led model development (Phase 2), feature selection was carried at this stage when the cross-validation with Lasso penalty was taking place. Features emerging with coefficient values that were above the absolute zero were brought into the next stage of the model development.

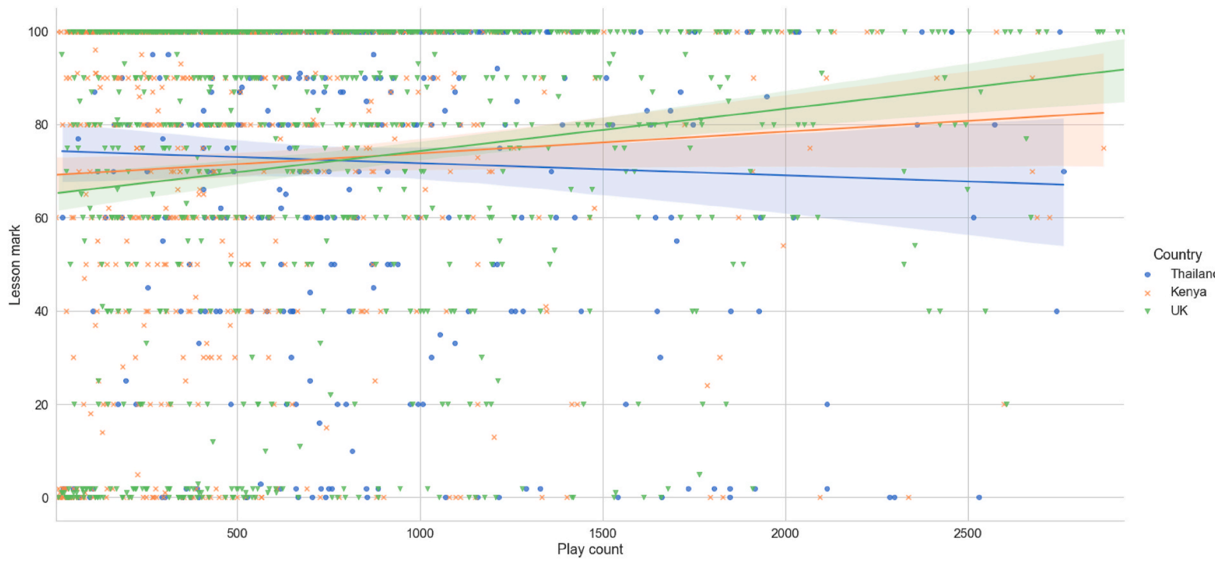

Fig. 6. Scatter plot showing how country (UK, Kenya, and Thailand) was related to learning outcomes (*lesson_mark*). Untransformed data are represented here.

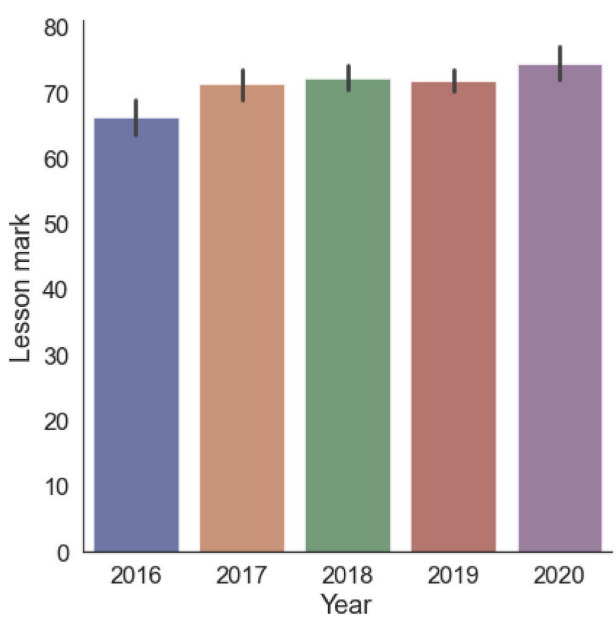

Fig. 7. Changes in online learning outcomes as the years (*markedYear*) progressed.

3.2. Individual feature analysis

Features emerging from the theoretical framework were those included in the theory-led model and will be examined first, followed by the most important features to emerge from the final model which was data-led. The means (M) and standard deviations (s.d.) reported are for the absolute Shapley values to emerge from the final, data-led XGBoost model.

3.2.1. Theory-led features

Of the theory-led features, cumulative experience as measured by *indiv_pupil_t* (i.e., the cumulative lesson count for each learner) was included in the final model (Table 2, feature 11) from the cross-validation with Lasso penalty. It was found to have the feature importance of absolute Shap M = .08 (s.d. = 0.07, Table 3). A feature related to *indiv_pupil_t* was also selected for the final model: namely, *play_count* (i.e., summed platform experience), which was also found to be one of the most important features for predicting learning outcomes (*lesson_mark*, M = 0.08, s.d. = 0.07). Moreover, hierarchical clustering revealed these two measures to cluster together, along with *totalQuestions*[5] (total number of questions attempted during that lesson). Thus, the hypothesised importance of cumulative experience and related constructs in online learning found support in the final model.

Country had been hypothesised to play an important role in predicting learning outcomes. Indeed, an overall country effect was supported by the feature selection process, when country variables qualified to be included for the data-led model (Table 2): namely, *Kenya, UK,* and *InCountryDep*. Furthermore, two country features, *UK* and *Kenya,* were found to cluster together, giving credence to the overall importance of country setting in the potential for learners to benefit from online learning, as well as the importance of LMIC status (Fig. 2B). The within-country deprivation level (*InCountryDep*) emerged to have the absolute Shapley value M = 0.00099, s.d. = 0.0027. The surprise, however, is in the positive prediction of both country dummies (Kenya, M = 0.02, s.d. = 0.02; UK M = 0.02, s.d. = 0.02), suggesting an overall Thai disadvantage which is supported by Fig. 6 (*lesson_mark* ~ *country* scatterplot) in which Thai learners appear to make the lowest learning gains out of the three countries sampled in the present analyses. It is notable that LMIC status does not feature in the final model. Altogether, the effect of country on online learning outcomes seems to be at within- and between-country level rather than at between-continent level.

The anticipated Covid effect was not seen in the final model via the primary feature, *since_covid*, which did not qualify to be selected as a feature in the final model (Table 2, feature 25). However, *markedYear* was a related feature for the same concept and it did qualify in feature selection either (Table 2, feature 13), emerging with some importance (M = 0.02, s.d. = 0.02), lending some support to the hypothesised importance of Covid in predicting online learning outcomes (Fig. 7).

The expected gender effect was not found (Table 2, feature 26).

3.2.2. Data-led features

Based on the final, data-led model, the most important features include *totalQuestions* (i.e., the number of questions attempt during the lesson; M = 0.28, s.d. = 0.42), *lesson_type* (M = 0.25, s.d. = 0.12), and *total_help* (M = 0.16, s.d. = 0.17; Fig. 8).

[5] Note that the feature, *play_count*, was not found to cluster with *indiv_pupil_t* and *totalQuestions* until the distance was set at 0.9.

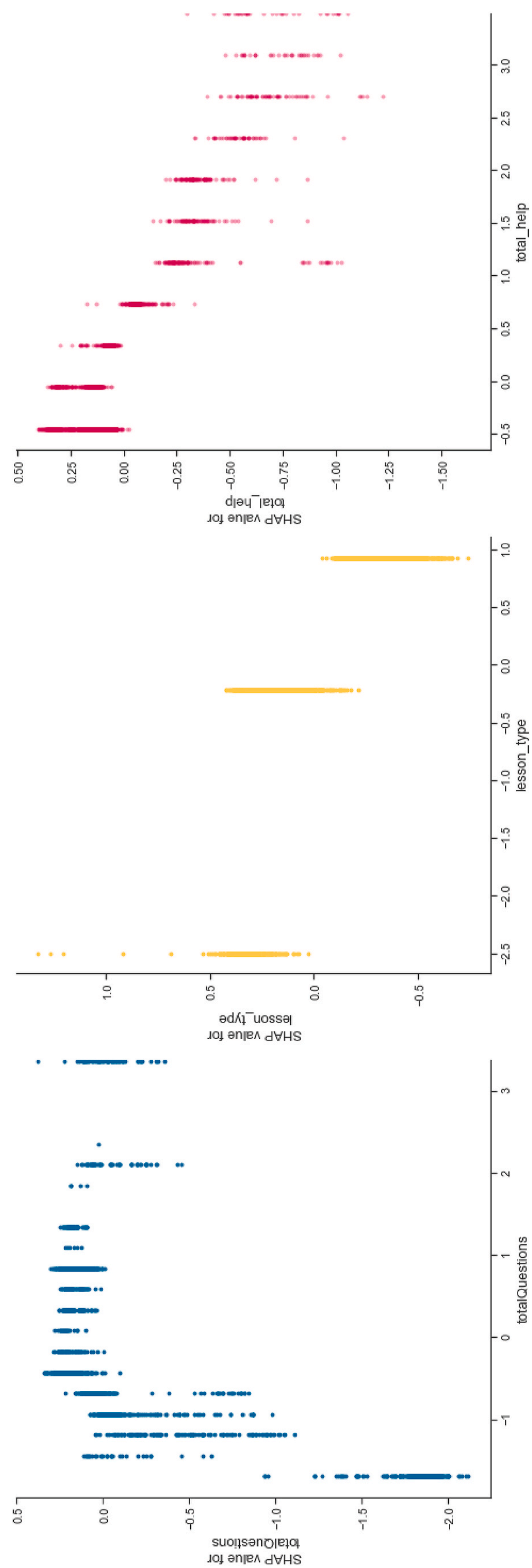

Fig. 8. Top three most important features in predicting *lesson_mark*. Transformed data are represented here.

Fig. 9 shows the relationship between learning outcomes and *total Question*'s, alluding to the importance of within-lesson engagement.

Fig. 10 shows the relationship between *lesson_type* and learning outcomes: *replay* lessons led to the best learning outcomes, followed by *tutor_pb* (i.e., tutor playback, or jumping backwards through tutor exercises), then *skip_lesson* which resulted in the lowest lesson marks (see also Fig. 8b). The feature, *total_help* (M = .16, s.d. = 0.17) is third most important and seems to have an optimal number of help requests before help-seeking ceases to support learning outcomes (Fig. 8c). In particular, it appears that students can ask for too much help, even from the first help requests onwards (Fig. 8c).

3.3. Feature interactions

Table 4 shows the top 70 feature interaction pairs. Fig. 11 shows the top 20 interactions. However, the remainder of the Results section will discuss the features that interacted with the theory-led features in the final model (that is, the three that found support from the final model), followed by the top six interacting feature pairs, which are also those with Shap interaction values (φ) of .05 and above (see Table 4).

3.3.1. Interactants with theory-led features

Upon inspecting the top 70 interacting feature pairs (Table 4), the importance of cumulative experience (*indiv_pupil_t*) and summed platform experience (*play_count*) could be found to form 12 interacting features pairs with $\varphi \geqslant 0.03$. These are shown in Fig. 12. Among these, cumulative experience (*indiv_pupil_t*) particularly combines with *questionTime* to push learning outcomes higher ($\varphi = 0.04$). Cumulative experience (*indiv_pupil_t*) also combines with higher levels of math difficulty (*mathLevel*), where greater experience combines with math difficulty to push learning outcomes higher ($\varphi = 0.03$). Additionally, summed platform experience (*play_count*) can be observed to push learning outcomes lower with increasing *mathAbility*, suggesting that with increasing *mathAbility*, less platform experience is required ($\varphi = 0.03$).

From Table 4, country could be found to predict learning outcomes with three potential interactants (Fig. 13). Learners being situated in *Kenya* were more likely than those outside of Kenya to predict decreasing learning outcomes when combined with summed platform experience (*play_count*, $\varphi = 0.02$). On the other hand, learners in the *UK* display a propensity to increase learning outcomes as *questionTime* increased ($\varphi = 0.01$). This alludes to a LMIC-related disadvantage in online learning.

The feature, *markedYear*, represented the Covid effect when the *since_covid* feature failed to be selected for the final model. The 'Covid effect' is the effect of emergency school closure resulting in the necessity of engaging with online learning. When combined with increasing *questionTime* ($\varphi = 0.02$), *indiv_pupil_t* ($\varphi = 0.01$), and *play_count* ($\varphi = 0.01$), increasing *markedYear* (i.e., closer to or during Covid) was linked with decreasing learning outcomes (Fig. 14). Thus, the 'beneficial' effect of Covid on online learning outcomes received some opposition when combined with measures related to engagement (*questionTime*) and cumulative experience (*indiv_pupil_t, play_count*).

3.3.2. Data-led feature interactions

To turn to the data-led perspective, the top six interacting features are shown in Fig. 15. The feature, *lesson_type,* emerged from the data-led analyses to have particular importance in online learning outcomes: it combined with fewer *total_helps,* increasing *totalQuestions*, and shorter *questionTime* to predict improving learning outcomes. The features, *lesson_type* and *tutorialTime* had also been found to cluster together (Fig. 2B), suggesting potential interaction between these features in predicting *lesson_mark*, although the feature pair's Shap interaction value did not support an interactive relationship per se.

Two interactions especially warrant discussion here. As *totalQuestions* increased, learning outcomes decreased: that is, until *total_help*

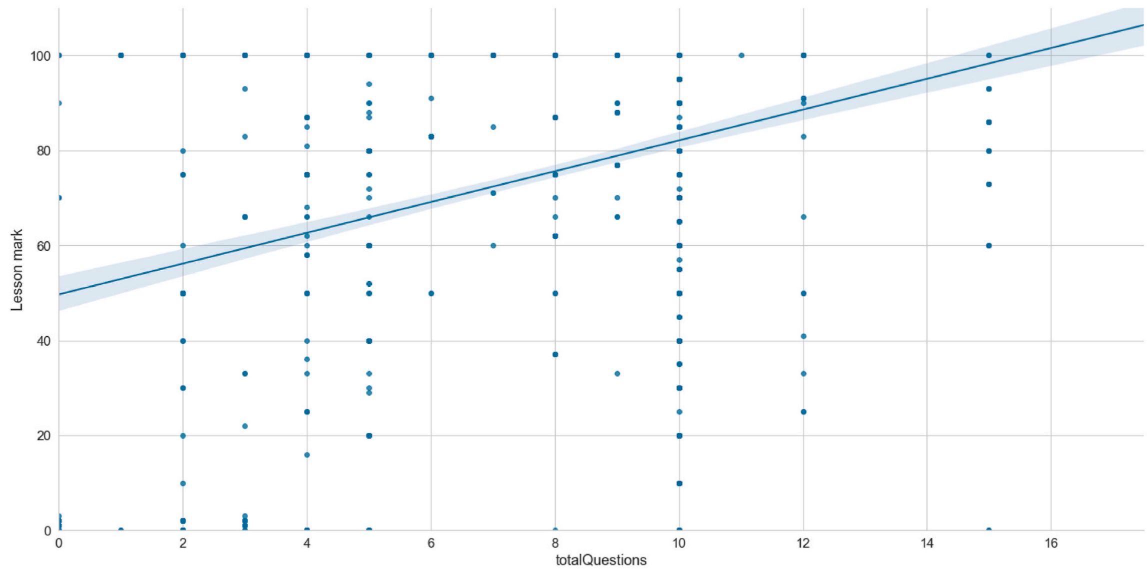

Fig. 9. Scatter plot showing how totalQuestions was related to learning outcomes (lesson mark). Untransformed data are represented here.

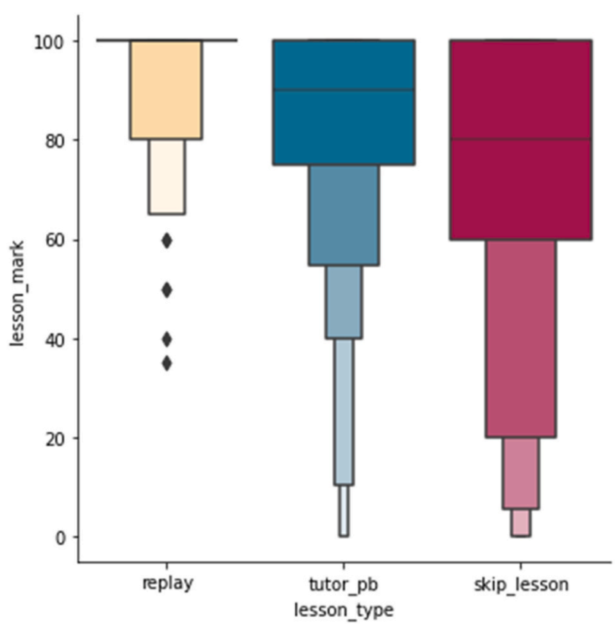

Fig. 10. Lesson type was a significant feature in predicting students' learning outcomes (leson_mark), according to mean absolute Shapley values. This figure breaks down the lesson types that featured in a random subsample (n = 1000) of the current data set. Note that tutor_pb refers to tutor playback: that is, jumping backwards through tutor exercises. Untransformed data are represented here.

occurred, which saw improving learning outcomes (φ = .18). Shorter *questionTime* combined with shorter *timeTaken* to predict decreasing learning outcomes; in contrast, learning outcomes improved as both features increased. These had also been found to cluster together (Fig. 2B), supporting potential interaction between these features in predicting *lesson_mark*.

4. Discussion

The present machine learning analysis combined the theory-led approach to data science with the data-led approach. For this, features were derived from theory only during initial model development. When proven necessary, a data-led approach was employed in feature selection during subsequent model development. The narrative for the final model's analytic outcomes then started with the theory-led features (Hypotheses 1 to 4) before turning to features that have emerged to be important from the data-led model development (Hypothesis 5). By bringing both traditions together in the present data mining, rich insights have emerged to largely support the theory-led hypotheses regarding online learning, but also to supplement these with some unexpected feature patterns from the data-led perspective.

4.1. *Developing expertise through cumulative online learning experience (Hypothesis 1)*

Deliberate practice is a core dimension to developing expertise. The importance of cumulative experience (*indiv_pupil_t* and *play_count*) in online learning found support from the present research. The more lessons learners attempted, the better their learning outcomes were over time. This aligns with prior research, wherein skill-related expertise has been observed to develop in a context-specific manner (e.g., Sternberg, 2001a): the more deliberate practice in the 'context' of one online learning environment, therefore, the better the learner's effectiveness in navigating and benefitting from that environment. Thus, cumulative experience in online learning has been shown to resemble deliberate practice which, in turn, develops expertise. Moreover, the present online environment provides instructor feedback. Tutor-initiated playback and replay (i.e., *lesson_type*) led in their importance to online learning outcomes, lending further support for cumulative exposure to online learning environments as deliberate practice for developing expertise (i.e., learning outcomes, Fig. 16).

Indeed, effective deliberate practice typically involves the support from resources such as teachers or tutors (Ericsson et al., 1993). Another component of deliberate practice was given prominence in the final, data-led model: namely, the number of questions attempted in a lesson (*totalQuestions*). Correspondingly, skipping lessons (*lesson_type*) was associated with decreased learning gains. Altogether, "engagement in activities specifically designed to improve performance in a domain" (Meinz & Hambrick, 2010, p.914) is yet another expression of deliberate practice and has been found in this study to have importance in developing expertise. Thus, the online learning platform featured in this study does not involve mindless practice (Ericsson, 2018). Rather, cumulative experience with online learning in this study has been seen to resemble and reap the benefits of deliberate practice for developing expertise.

Table 4
The top 70 SHAP interaction values.

	Feature	Shap interaction value	cum_diff
1	total_help * lesson_type	0.1	
2	totalQuestions * lesson_type	0.07	−0.04
3	questionTime * lesson_type	0.06	0
4	play_count * indiv_pupil_t	0.06	0
5	total_help * totalQuestions	0.06	0
6	mathAbility * lesson_type	0.05	−0.01
7	questionTime * mathLevel	0.04	−0.01
8	indiv_pupil_t * questionTime	0.04	0
9	indiv_pupil_t * totalQuestions	0.04	0
10	indiv_pupil_t * lesson_type	0.04	0
11	topicId * lesson_type	0.03	0
12	birthMonth * mathAbility	0.03	0
13	mathLevel * lesson_type	0.03	0
14	indiv_pupil_t * mathAbility	0.03	0
15	questionTime * totalQuestions	0.03	0
16	topicId * mathLevel	0.03	0
17	indiv_pupil_t * mathLevel	0.03	0
18	topicId * questionTime	0.03	0
19	tutorialTime * totalQuestions	0.03	0
20	play_count * mathAbility	0.03	0
21	play_count * mathLevel	0.03	0
22	markedWeek * questionTime	0.03	0
23	topicId * play_count	0.03	0
24	mathAbility * questionTime	0.03	0
25	play_count * questionTime	0.03	0
26	mathAbility * mathLevel	0.03	0
27	markedWeek * play_count	0.03	0
28	topicId * indiv_pupil_t	0.03	0
29	topicId * totalQuestions	0.03	0
30	play_count * totalQuestions	0.03	0
31	questionTime * exerciseId	0.03	0
32	tutorialTime * questionTime	0.03	0
33	questionTime * total_help	0.03	0
34	topicId * mathAbility	0.02	0
35	topicId * markedWeek	0.02	0
36	topicId * birthMonth	0.02	0
37	topicId * exerciseId	0.02	0
38	birthMonth * questionTime	0.02	0
39	play_count * exerciseId	0.02	0
40	markedWeek * indiv_pupil_t	0.02	0
41	markedWeek * mathAbility	0.02	0
42	markedWeek * birthMonth	0.02	0
43	markedWeek * mathLevel	0.02	0
44	play_count * birthMonth	0.02	0
45	mathLevel * totalQuestions	0.02	0
46	birthMonth * indiv_pupil_t	0.02	0
47	birthMonth * lesson_type	0.02	0
48	play_count * lesson_type	0.02	0
49	tutorialTime * total_help	0.02	0
50	markedYear * questionTime	0.02	0
51	play_count * Kenya	0.02	0
52	mathLevel * exerciseId	0.02	0
53	mathLevel * total_help	0.02	0
54	mathAbility * tutorialTime	0.02	0
55	markedYear * indiv_pupil_t	0.01	0
56	indiv_pupil_t * exerciseId	0.01	0
57	play_count * tutorialTime	0.01	0
58	birthMonth * markedYear	0.01	0
59	topicId * total_help	0.01	0
60	UK * questionTime	0.01	0
61	play_count * markedYear	0.01	0
62	birthMonth * mathLevel	0.01	0
63	mathAbility * exerciseId	0.01	0
64	mathAbility * totalQuestions	0.01	0
65	topicId * tutorialTime	0.01	0
66	tutorialTime * lesson_type	0.01	0
67	tutorialTime * mathLevel	0.01	0
68	topicId * Kenya	0.01	0
69	tutorialTime * exerciseId	0.01	0
70	indiv_pupil_t * tutorialTime	0.01	0

Contrary to previous findings which have posited help-seeking as a positive and adaptive behaviour that is associated with learner confidence (e.g., Nelson & Fyfe, 2019; Undorf et al., 2021), the present research suggests that help-seeking may only be constructive to a very limited extent since, after one help-seeking attempt, further help-seeking predicted declines in learning. However, greater light can be shed on the mechanisms of adaptive help-seeking behaviour for developing expertise (i.e., improving learning outcomes) by future clustering of learners represented by the present data in accordance with their socio-cognitive profile such as those in the cited studies. Sequential analyses can also be conducted on other online learning data to ascertain the events that lead to adaptive versus maladaptive help-seeking. For now, the present analyses suggest that independence and perseverance are the winning traits for developing expertise in online learning, which echoes the importance of independence in reading (e.g., Babbs & Moe, 1983) — a process closely related to online learning (Coiro, 2011; Leu et al., 2018; Zhang & Duke, 2008).

Moreover, independence is an indicator of meta-cognitive capability (Anderson, 2012), and the present analyses thus lend support to the importance of metacognition for developing expertise in online learning. Indeed, meta-cognition has been highlighted as a central dimension of expertise development alongside deliberate practice, both generally and for online learning. Sternberg (2001b) contended that meta-cognition demonstrably improves with cumulative experience with expertise being developed, in the context where the task is being performed. Meta-cognition is particularly important for engagement with online learning (Lang, 2021; Stevens, 2020).

Altogether, the final analytic model would suggest that expertise development via online learning relies on cumulative experience on the platform that resembles deliberate practice. On the present online learning platform, deliberate practice involves engagement with the tailored activities, reflection through the questions, and feedback from the online 'tutor' who both indicates whether questions were handled correctly as well as whether the learner is ready to progress onto the next stage of learning. Accompanying such cumulative experience, there must be meta-cognition which, in the present analyses, involves independence and perseverance as opposed to continued and unhindered help-seeking. This framework is summarised in Fig. 16.

4.2. *Country factors in online learning outcomes (Hypothesis 2)*

Country was predicted to have importance in the efficacy of online learning. The present analyses found support for between-country differences in the potential of developing expertise through online learning. In particular, there seemed to be a Thai disadvantage in online learning, wherein cumulative experience with online learning results in the least expertise development. Therefore, potential culture-specific differences in online learning are worthy of consideration. Cultural differences in social orientation have been documented to correlate with the degree to which children can sustain their own learning, or self-regulate. So, learners from interdependent and collectivist cultures were found to exceed those from independent, individualistic cultures in self-regulation as measured by thought suppression (Seeley & Gardner, 2003). Yet, it is the learners in Thailand rather than those in less collectivist cultural settings that have been shown to struggle with independent online learning: the Thai disadvantage might be explained by the interpersonal core of collectivist self-regulation, in which the interpersonal reference point is crucial in order for collectivist learners to exercise self-regulation (Trommsdorff, 2010). Motivation is another between-country factor: the collectivist online learner's motivations for engaging in online learning is more often extrinsic, compelled by others, in contrast to the individualistic online learner's motivation who undertakes online learning for oneself rather than for others (Lim, 2004), which provides further insight into the present Thai disadvantage in the expertise development through online learning. Furthermore, Thailand has been observed not only to be a collectivist cultural setting but also to

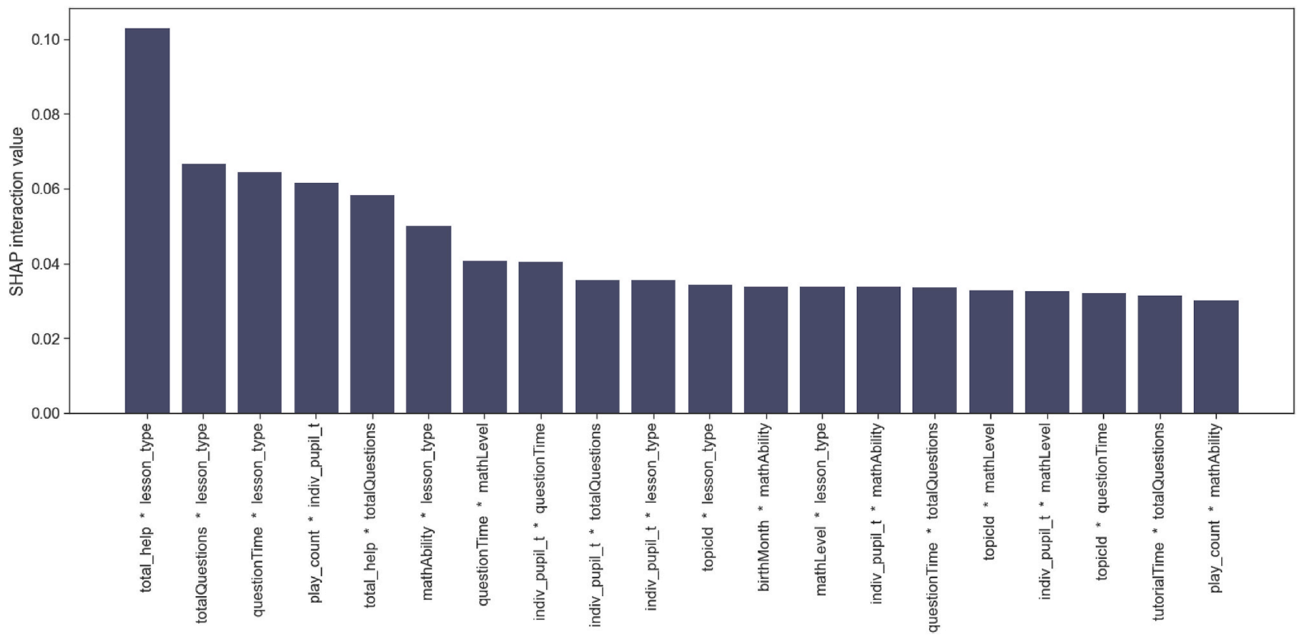

Fig. 11. SHAP interaction values for predicting learning outcomes (*lesson_mark*), from the strongest interaction to the weakest. Only the top 20 interactions are shown here. Transformed data are represented here.

have high power distance (Hofstede, 1986) — a characteristic that has been used to explain Thailand's exceptionally low English language proficiency (Young, 2021), which itself may be further compounding learners' lesser gains made from online learning (Education First, 2021).

In addition, there was an expectation that LMIC status would explain country differences in expertise development through online learning. This expectation was challenged when the LMIC feature was not selected for the final (data-led) analytic model. However, the Kenyan disadvantage in benefitting from cumulative online learning experience found in the interaction analyses, the Thai disadvantage found in the individual feature analyses, as well as the UK advantage in making learning gains through cumulative online learning experience, together, offer support to the expected LMIC differences in developing expertise through online learning. Unlike low-tech educational resources such as [BLINDED FOR REVIEW][PRESENT AUTHORS], the technologically demanding nature of online learning poses stark challenges to LMICs in which "power shortage, weak internet connectivity, low internet coverage, and low computer literacy" (Pasha et al., 2016, p. 225) dominate — especially in individual households. Indeed, the strains and considerations are comprehensive, if educational technology is to serve as well in LMICs as in high-income countries (HICs): these include infrastructure, class group integration, student technological capability, student engagement, and the data collected for platform design and analytics (Kaye & Ehren, 2021).

4.3. The effect of COVID-19 on developing expertise online (Hypothesis 3)

Learners had been expected to enjoy increased expertise development through online learning when the COVID-19 pandemic prompted universal online learning through emergency school closure. This was not found to be the case from the most direct measure of the 'Covid effect' available (*since_covid*). One reason for this null finding could have been that only five months were sampled from the start of the pandemic (i.e., from March to July 2020 in the countries sampled by this study) within the since_covid feature. Had more time been sampled from the start of the pandemic, then the potential for developing expertise through online learning may have been more detectable as learners establish their digital readiness for online learning which was limited among many learners, especially those in LMICs such as Kenya and Thailand (Ahmadon et al., 2020, pp. 1–4; Akuratiya & Meddage, 2021).

Nevertheless, some support for the importance of COVID-19 was found as online learning expertise development improved over the years (*markedYear*). Since the Pandemic took place only in the final periods of the data analysed, advancement in learning outcomes over the years can be inferred as being due to the Pandemic. However, it is acknowledged that improvements may also be explained by the general, universal technological advancement and adoption (Lin & Johnson, 2021; Umar et al., 2017; Yeung et al., 2021).

4.4. Gender differences in online expertise development (Hypothesis 4)

The anticipated gender difference in online expertise development was not found. In fact, the relevant feature was ranked the lowest in the cross-validation process for feature selection. This may be a reflection of a change in the tide in terms of gender inequalities, or the persisting inequalities have been unsuccessfully captured in the particular model reported here. A future analysis with other outcome variables would add important insight into areas in which gender inequalities persist, such as educational access (Shilling, 1991), especially in LMICs (Jewitt & Ryley, 2014).

4.5. Limitations

The present measure of learner expertise is not comprehensive with multiple dimensions (e.g., Palmer et al., 2005 for teacher expertise) but only consisted of one single dimension, hence the use of cumulative expertise as the primary term. However, future research would benefit from exploring a comprehensive measure of learner expertise in predicting benefits from individualised learning online. Components of learner expertise include subject performance (Kalyuga, 2006), prior knowledge (Lee & Kalyuga, 2011), and their working memory (Kalyuga, 2009). Accordingly, the more limited a learner's working memory — or their learner expertise — the sooner they encounter issues around cognitive load (Kalyuga, 2008; Yeung, 1999). Whereas the present research focuses on cumulative experience and takes experience as an indicator of expertise (see also de Bruin et al., 2007), extant literature typically takes learner performance as an indicator of learner expertise.

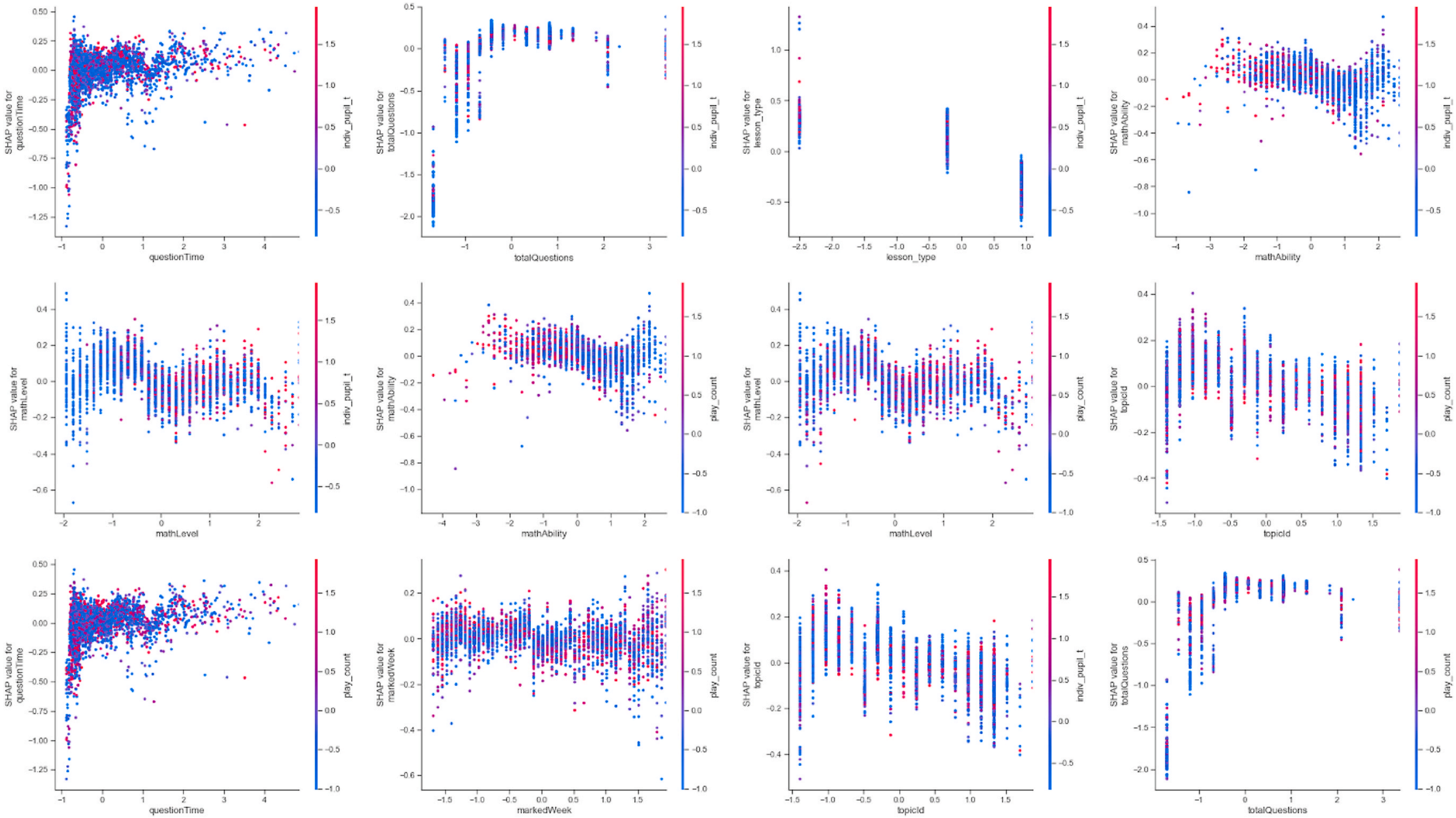

Fig. 12. Dependence plots of the importance of the features that interact with cumulative experience (either *indiv_pupil_t* or *play_count*) in predicting learning outcomes (*lesson_mark*), according to mean absolute Shapley values. Transformed data are represented here.

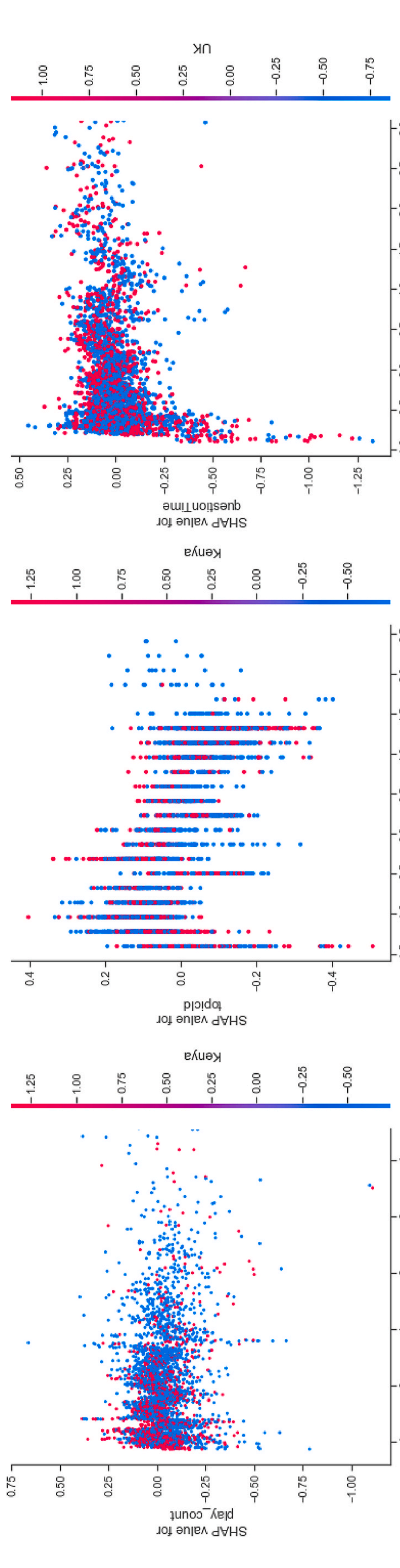

Fig. 13. Dependence plots of the importance of the features that interact with Country (either UK or Kenya) in predicting learning outcomes (lesson mark), according to mean absolute Shapley values. Transformed data are represented here.

Thus, the conceptualisation of learner expertise is limited to date, with insufficient consideration for meta-cognitive qualities and socio-emotional skills (Lang, 2021) are integral to successful learning, including with the design of educational technology (Chiu, 2018). Subsequent research needs to better encompass the nuances and complexities in learner expertise in the context of educational technology.

It is not known whether the platform-specific learning outcomes in the present analysis generalise beyond the online environment. However, subject-specific expertise is domain-general and is likely to transfer across contexts such that the expertise development observed in the presentation is likely to transfer to other contexts, online or in-person. The domain-general nature of subject expertise has found support from Hwang and Shin (2019), who found syntactic knowledge to transfer from a within-language context to a between-language context. Similarly, the between-context transferability of subject knowledge has widespread support, including from math education research. For example, algebraic problem-solving processes have been found to transfer between subjects, from Mathematics learning to Physics (Bassok & Holyoak, 1989). Moreover, within each, Mathematics and Physics, problem-solving has been found to transfer between contexts (Rebello et al., 2017). In particular, the transfer subject expertise improves as the learning environment increases in resemblance to the target context (Herrington et al., 2003; Nemanich et al., 2009; Nokes, 2009) and in interactivity, or feedback (Anderson, 2004; Boling et al., 2012). So, students' subject-related expertise in mathematics, gained through cumulative experience in the online learning environment, can be expected to transfer out of the online learning environment and be applied in another context.

Such cross-context knowledge transfer has also been found with online learning. Cumulative experience in class-based online learning has been shown to increase the implicit learning of another language (e.g., fluency, Omura, 2021). Similarly, cumulative experience in learning one foreign language seems to transfer to another foreign language (Hwang & Shin, 2019). Cumulative experience in online learning for Engineering students has also found subject expertise (or knowledge) to improve in other, offline contexts (Alkhalil et al., 2021, pp. 396–399). As such, it is possible that cumulative experience in math learning on one online context (or platform) can transfer to another context, digital or otherwise.

It should be acknowledged that the present research does not, on its own, explain any single concept or association. This is due to the limitations of a highly quantitative and aggregative nature of the present approach, not to mention the novelty of machine learning analysis in the field of educational science. Rather, the present research must stand alongside related research with the multiplicity of research approaches for a comprehensive understanding of the concepts discussed. Moreover, within the quantitative research paradigm, sequential analyses should follow the present research with temporal properties taken into account: then, stronger insights can be derived into potential pathways of causality.

4.6. Future directions

Longitudinal analysis can be conducted on this data, in future. Analysis from the time series family of statistical techniques would resolve potential differences in how the importance of cumulative experience, country, gender, and COVID-19 develop over time in predicting expertise development (or learning outcomes). Related, it will be insightful to track and analyse developments in the impact of COVID-19 in the importance of cumulative expertise, country, and gender, in expertise development.

More technically, although absolute Shapley values were selected for their outperformance of the original Shapley value, this metric carries with it shortcomings to account for in future work. Although the absolute Shapley value used in this study carries pioneering strengths, especially the strong theory-based algorithm and interpretability

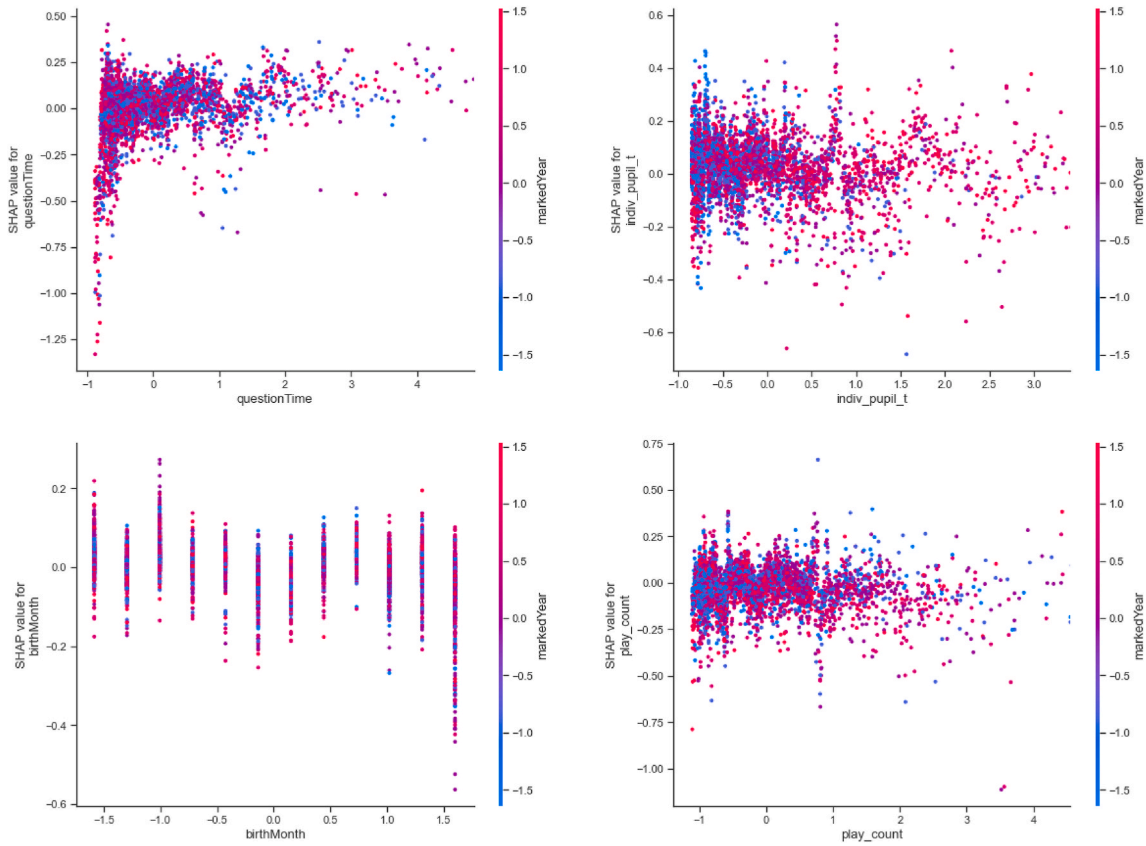

Fig. 14. Dependence plots of the importance of the features that interact with Covid, as measured by *markedYear,* in predicting learning outcomes (*lesson_mark*), according to mean absolute Shapley values. Transformed data are represented here.

relative to non-Shapley metrics for feature importance, increasing attention is being given to the limitation of this and preceding Shapley values (e.g., zero-based and original Shapley values) for the presence of multicollinearity and dependence among the features included in any final model. Alternatives to the absolute Shapley value should be explored for the present research question, including the Kernal Shap (Aas et al., 2021) and the multi-collinearity corrected Shapley values (Basu & Maji, 2022).

4.8. Implications and conclusions

Through a partnership between theory- and data-led model development and analysis, machine learning was applied to big data from one online learning platform. Developing expertise was found to be possible through online learning, with particular resemblance of deliberate practice (Hypothesis 1): cumulative experience in online learning yielded improving learning outcomes over time; the 'tutor' responses available from the sampled online learning platform served as instructor feedback. The importance of meta-cognitive capabilities in developing expertise was also supported: learners' ability to operate as independent rather than help-seeking learners was encouraged, in order to attain higher learning outcomes. This finding extends the distinction between adaptive and maladaptive help-seeking from in-person learning (Newman, 2002) to online learning: namely, that effective self-regulation underlies adaptive help-seeking wherein the learner seeks to develop as an independent learner, rather than simply to obtain the correct answer.

Online learning was found to have differing potential for developing expertise depending on learners' country settings (Hypothesis 2). A Thai disadvantage was found, potentially due to Thai learners' need for social references in order to self-regulate successfully. These country differences echo existing calls for country-specific appropriateness in online learning platform design, given the significant between-country differences that exist with regard to social dynamics for learning (Liu, 2017; Ogan et al., 2015). The final model also pointed to a LMIC disadvantage in developing expertise through online learning, thus supporting the importance of accounting for infrastructural constraints in online learning design and provision (Abidi et al., 2017). COVID-19 was found to improve the potential for developing expertise through online learning (Hypothesis 3). Gender did not find support in the present analyses (Hypothesis 4). Finally, data-led model development yielded conceptually important insights that were not hypothesised through theory alone (Hypothesis 5).

Altogether, the significance and contribution of the present research are two-fold. First, insights from empirical research on the development of learner expertise have been extended from in-person settings to online learning. Second, the potential for machine learning to make theoretically significant, relevant, and applicable contributions to the field has been demonstrated: that is, by partnering theory with data closely during model development and by making thorough use of interpretable

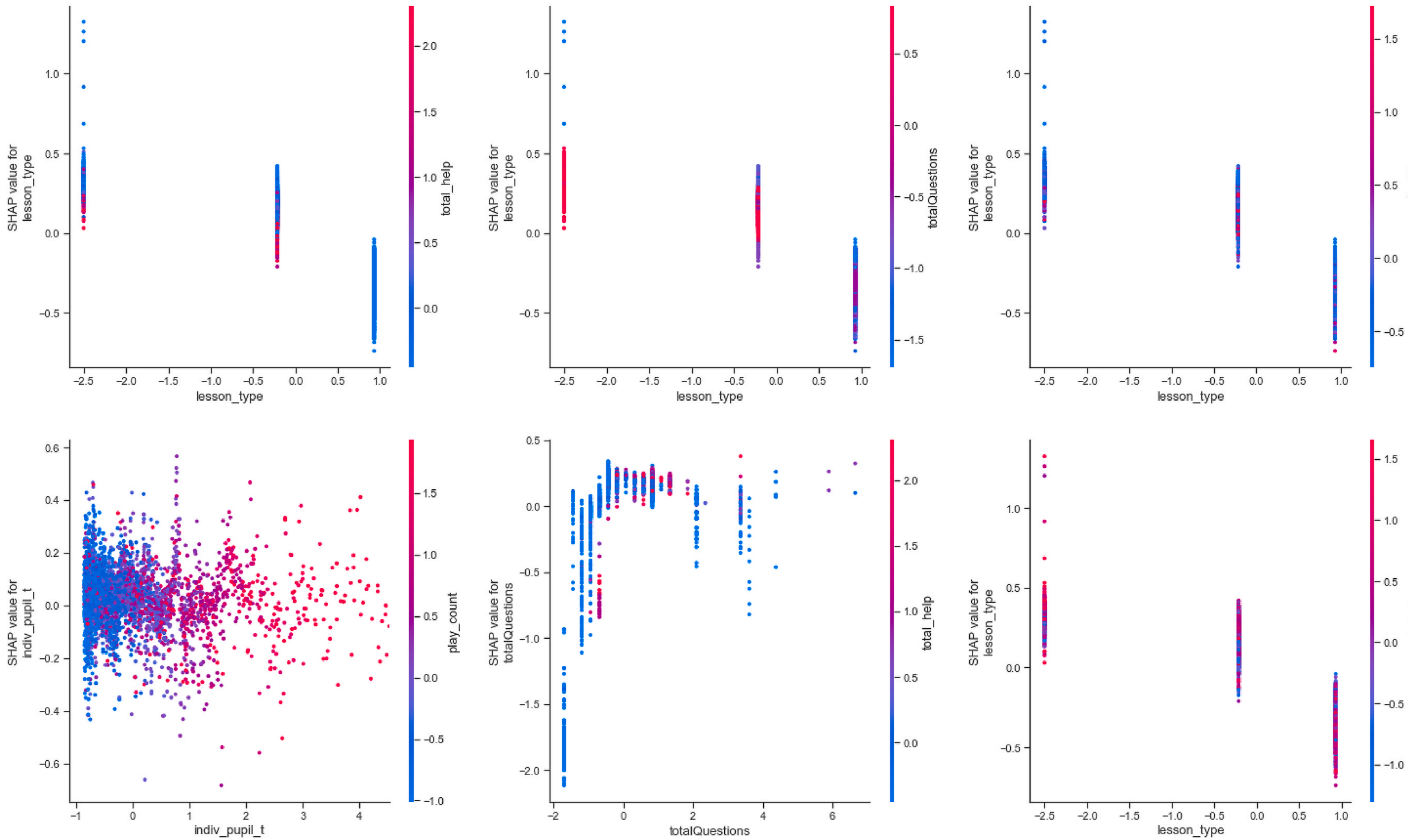

Fig. 15. Dependence plots for the six strongest interactants to emerge from the final model. Lesson_type: 1.5 = replay; 0.25 = tutor_pb; 0.75 = skip_lesson.

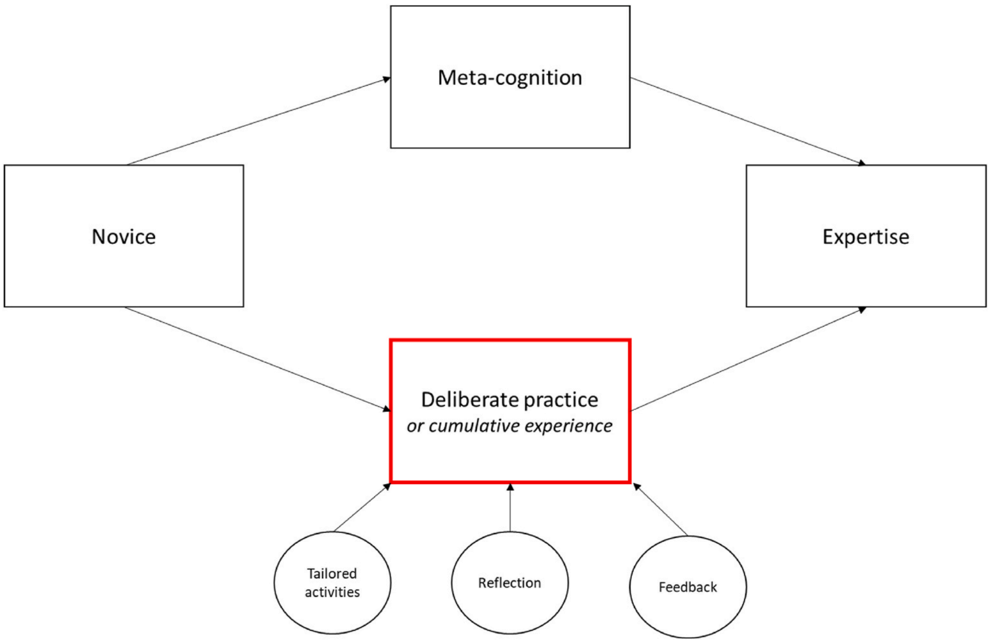

Fig. 16. Conceptual framework for how learner expertise is developed through online learning. This is the outcome of the final model, developed through theory and data. The present online learning platform thus provides opportunity for deliberate practice during cumulative experience on the online platform, which is one of two main avenues for developing learners' expertise in mathematics. According to the final model, deliberate practice during cumulative experience with the online learning platform must be accompanied by meta-cognition (independence and perseverance, rather than unhindered help-seeking) to yield subject expertise. This Figure relates to Hypothesis 1 in particular: namely, that cumulative experience will develop expertise in online learning within one particular platform and, as such, will predict increasing learning outcomes.

machine learning tools such as Shapley values and visualisations. This article has demonstrated the potential value of implementing data mining with machine learning in furthering insights that relate to effective online learning provision. Thus, the feasibility of substantive insights from data science research on online learning has been demonstrated.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Special thanks to Chun Pong Lau and Junaid Mubeen from Whizz Education for the continued support with obtaining the data necessary to conduct the present analyses. Part of the present research was conducted whilst the author was employed by the EdTech Hub with funding from the Foreign, Commonwealth and Development Office, UK.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.caeai.2022.100106.

References

Aas, K., Jullum, M., & Løland, A. (2021). Explaining individual predictions when features are dependent: More accurate approximations to Shapley values. *Artificial Intelligence, 298*, Article 103502. https://doi.org/10.1016/j.artint.2021.103502

Abidi, S. H., Pasha, A., Moran, G., & Ali, S. (2017). A roadmap for offering MOOC from an LMIC institution. *Learning, Media and Technology, 42*(4), 500–505.

Aboagye, E., Yawson, J. A., & Appiah, K. N. (2021). COVID-19 and E-learning: The challenges of students in tertiary institutions. *Social Education Research, 2*(1), 1–8.

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General, 117*(3), 288.

Adam, T. (2020). Between social justice and decolonisation: Exploring South African MOOC designers' conceptualisations and approaches to addressing injustices. *Journal of Interactive Media in Education, 7*(1), 1–11. https://doi.org/10.5334/jime.557

Ahmadon, F., Ghazalli, H. I. M., & Rusli, H. M. (2020). *Studying during pandemic: A review of issues from online learning in the middle of COVID-19. 2020 6th international Conference on interactive digital media (ICIDM)* (pp. 1–4).

Akuratiya, A., & Meddage, N. (2021). Readiness for online learning among students amidst COVID-19. *A Case of a Selected HEI in Sri Lanka, 5*(7), 2454–6186.

Al-Salman, S., & Haider, A. S. (2021). Jordanian university students' views on emergency online learning during COVID-19. *Online Learning, 25*(1), 286–302.

Alkhalil, S., Manasrah, A., & Masoud, M. (2021). Let's Learn with a Jigsaw! Implementing a Unique Collaborative Online Learning in an Engineering Course. In *International conference on information technology (ICIT)* (pp. 396–399), 2021.

Amarius, S., & Fredriksson, O. (2021). Digital games and collocations: A study of the relationship between L2 learners' gaming habits and knowledge of collocations. In *[Undergraduate project, Halmstad University]*. Google Scholar.

Anderson, T. (2004). Towards a theory of online learning. *Theory and Practice of Online Learning, 2*, 109–119.

Anderson, W. (2012). Independent learning: Autonomy, control, and meta-cognition. In *Handbook of distance education* (3rd ed.). Routledge.

Attewell, P., & Monaghan, D. (2015). Data mining for the social sciences: An introduction. In *Data mining for the social sciences*. University of California Press. https://doi.org/10.1525/9780520960596.

Babbs, P. J., & Moe, A. J. (1983). Metacognition: A key for independent learning from text. *The Reading Teacher, 36*(4), 422–426.

Bassok, M., & Holyoak, K. J. (1989). Interdomain transfer between isomorphic topics in algebra and physics. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(1), 153.

Basu, I., & Maji, S. (2022). Multicollinearity Correction and Combined Feature Effect in Shapley Values. In G. Long, X. Yu, & S. Wang (Eds.), *AI 2021: Advances in Artificial Intelligence* (pp. 79–90). Springer International Publishing. https://doi.org/10.1007/978-3-030-97546-3_7.

Baticulon, R. E., Sy, J. J., Alberto, N. R. I., Baron, M. B. C., Mabulay, R. E. C., Rizada, L. G. T., Tiu, C. J. S., Clarion, C. A., & Reyes, J. C. B. (2021). Barriers to online learning in the time of COVID-19: A national survey of medical students in the Philippines. *Medical Science Educator, 31*(2), 615–626.

Bing, W., & Ai-Ping, T. (2008). The influence of national culture toward learners'interaction in the online learning environment: A comparative analysis of shanghai tv university (China) and wawasan open university (Malaysia). *Quarterly Review of Distance Education, 9*(3), 327.

Biswas, B., Roy, S. K., & Roy, F. (2020). Students perception of Mobile learning during Covid-19 in Bangladesh: University student perspective. *Aquamedia, 4*(2), Article ep20023.

Boling, E. C., Hough, M., Krinsky, H., Saleem, H., & Stevens, M. (2012). Cutting the distance in distance education: Perspectives on what promotes positive, online learning experiences. *The Internet and Higher Education, 15*(2), 118–126.

Bouckaert, R. R., & Frank, E. (2004). Evaluating the replicability of significance tests for comparing learning algorithms. In H. Dai, R. Srikant, & C. Zhang (Eds.), *Advances in knowledge discovery and data mining* (pp. 3–12). Springer. https://doi.org/10.1007/978-3-540-24775-3_3.

Breddels, M. A., & Veljanoski, J. (2018). Vaex: Big data exploration in the era of Gaia. *Astronomy & Astrophysics, 618*, A13. https://doi.org/10.1051/0004-6361/201732493

de Bruin, A. B., Rikers, R. M., & Schmidt, H. G. (2007). Improving metacomprehension accuracy and self-regulation in cognitive skill acquisition: The effect of learner expertise. *European Journal of Cognitive Psychology, 19*(4–5), 671–688.

Bruns, J. A. (2010). *Democratic or gendered domain: Communication and learning styles in the online classroom*. Mankato: [Masters thesis, Minnesota State University] Cornerstone: A Collection of Scholarly and Creative Works for Minnesota State University.

Casad, B. J., Hale, P., & Wachs, F. L. (2015). Parent-child math anxiety and math-gender stereotypes predict adolescents' math education outcomes. *Frontiers in Psychology, 6*, 1597.

Catsambis, S. (1994). The path to math: Gender and racial-ethnic differences in mathematics participation from middle school to high school. *Sociology of Education, 67*(3), 199–215.

Chang, C.-L., & Fang, M. (2020). E-Learning and online instructions of higher education during the 2019 novel coronavirus diseases (COVID-19) epidemic. *Journal of Physics: Conference Series, 1574*(1), Article 012166.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Chen, X., Xie, H., Zou, D., & Hwang, G.-J. (2020). Application and theory gaps during the rise of artificial intelligence in education. *Computers and Education: Artificial Intelligence, 1*, Article 100002. https://doi.org/10.1016/j.caeai.2020.100002

Chi, M. T., Glaser, R., & Farr, M. J. (2014). The nature of expertise. In *Psychology press*.

Chiu, T. K. (2018). *Learner Expertise and Emotional Design in Multimedia Learning. International Society of the Learning Sciences, Inc.[ISLS].*

Cianciolo, A. T., Matthew, C., Sternberg, R. J., & Wagner, R. K. (2006). Tacit knowledge, practical intelligence, and expertise. In *The cambridge Handbook of expertise and expert performance* (pp. 613–632).

Cianciolo, A. T., & Sternberg, R. J. (2018). Practical intelligence and tacit knowledge: An ecological view of expertise. In K. A. Ericsson, R. R. Hoffman, A. Kozbelt, & A. M. Williams (Eds.), *The cambridge handbook of expertise and expert performance* (2nd ed., pp. 770–792). Cambridge University Press. https://doi.org/10.1017/9781316480748.039.

Coiro, J. (2011). Predicting reading comprehension on the Internet: Contributions of offline reading skills, online reading skills, and prior knowledge. *Journal of Literacy Research, 43*(4), 352–392.

De Mooij, M., & Hofstede, G. (2011). Cross-cultural consumer behavior: A review of research findings. *Journal of International Consumer Marketing, 23*(3–4), 181–192.

Dhawan, S. (2020). Online learning: A panacea in the time of COVID-19 crisis. *Journal of Educational Technology Systems, 49*(1), 5–22.

Ericsson, K. A. (2018). The differential influence of experience, practice, and deliberate practice on the development of superior individual performance of experts. In K. A. Ericsson, R. R. Hoffman, A. Kozbelt, & A. M. Williams (Eds.), *The cambridge handbook of expertise and expert performance* (2nd ed., pp. 745–769). https://doi.org/10.1017/9781316480748.038 Cambridge University Press.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*(3), 363–406, 10/d7gc4g.

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology, 47*(1), 273–305, 10/cc4h3c.

Gnanadass, E., & Sanders, A. Y. (2018). Gender still matters in distance education. In *Handbook of distance education* (4th ed.). Routledge.

Guiller, J., & Durndell, A. (2006). 'I totally agree with you': Gender interactions in educational online discussion groups: Gender interactions in educational online discussion groups. *Journal of Computer Assisted Learning, 22*(5), 368–381. https://doi.org/10.1111/j.1365-2729.2006.00184.x

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature, 585* (7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Herrington, J., Oliver, R., & Reeves, T. C. (2003). Patterns of engagement in authentic online learning environments. *Australasian Journal of Educational Technology, 19*(1). https://doi.org/10.14742/ajet.1701. Article 1.

Hofstede, G. (1986). Cultural differences in teaching and learning. *International Journal of Intercultural Relations, 10*(3), 301–320, 10/fwtktn.

Hwang, H., & Shin, J.-A. (2019). Cumulative effects of syntactic experience in a between- and a within-language context: Evidence for implicit learning. *Journal of Memory and Language, 109*, Article 104054.

Jafree, S. R. (2021). *The need for cultural interventions to improve girls' education during COVID-19 and beyond.* ISA eSymposium for Sociology.

Jewitt, S., & Ryley, H. (2014). It's a girl thing: Menstruation, school attendance, spatial mobility and wider gender inequalities in Kenya. *Geoforum, 56*, 137–147.

Jiang, H., Islam, A. A., Gu, X., & Spector, J. M. (2021). Online learning satisfaction in higher education during the COVID-19 pandemic: A regional comparison between eastern and Western Chinese universities. *Education and Information Technologies, 26*, 6747–6769.

Jones, N., Tapia, I. S., Baird, S., Guglielmi, S., Oakley, E., Yadete, W. A., Sultan, M., & Pincock, K. (2021). Intersecting barriers to adolescents' educational access during COVID-19: Exploring the role of gender, disability and poverty. *International Journal of Educational Development, 85*, Article 102428.

Kalyuga, S. (2006). *Instructing and testing advanced learners: A cognitive load approach.* Nova Science Publishers.

Kalyuga, S. (2008). Relative effectiveness of animated and static diagrams: An effect of learner prior knowledge. *Computers in Human Behavior, 24*(3), 852–861.

Kalyuga, S. (2009). The expertise reversal effect. In *Managing cognitive load in adaptive multimedia learning* (pp. 58–80). IGI Global.

Kaye, T., & Ehren, M. (2021). Computer-assisted instruction tools: A model to guide use in low-and middle-income countries. *International Journal of Education and Development using Information and Communication Technology, 17*(1), 82–99.

Khan, A., Ahmad, F. H., & Malik, M. M. (2017). Use of digital game based learning and gamification in secondary school science: The effect on student engagement, learning and gender difference. *Education and Information Technologies, 22*(6), 2767–2804.

Kilbourne, W., & Weeks, S. (1997). A socio-economic perspective on gender bias in technology. *The Journal of Socio-Economics, 26*(3), 243–260. https://doi.org/10.1016/S1053-5357(97)90018-4

Kim, Y., Sohn, D., & Choi, S. M. (2011). Cultural difference in motivations for using social network sites: A comparative study of American and Korean college students. *Computers in Human Behavior, 27*(1), 365–372.

King, M., Pegrum, M., & Forsey, M. (2018). MOOCs and OER in the global south: Problems and potential. *International Review of Research in Open and Distance Learning, 19*(5), 1–20. https://doi.org/10.19173/irrodl.v19i5.3742

King, M. W., & Resick, P. A. (2014). Data mining in psychological treatment research: A primer on classification and regression trees. *Journal of Consulting and Clinical Psychology, 82*(5), 895–905. https://doi.org/10.1037/a0035886

Kulasegaram, K. M., Grierson, L. E., & Norman, G. R. (2013). The roles of deliberate practice and innate ability in developing expertise: Evidence and implications. *Medical Education, 47*(10), 979–989.

Lang, J. C. (2021). From teaching excellence to expertise development: A pedagogical framework for developing expertise. *Creative Education, 12*, 907, 04.

Launois, P., Allotey, P., Reidpath, D., Maher, D., Certain, E., & Ross, B. (2019). Lessons learnt from a professional development MOOC: Engaging culturally and linguistically diverse learners from low- and middle-income countries. *European Journal of Open, Distance and E-Learning, 22*(2). Article 2 https://old.eurodl.org/?p=archives&sp=brief&year=2019&halfyear=2&article=809.

Lee, C. H., & Kalyuga, S. (2011). Effectiveness of on-screen pinyin in learning Chinese: An expertise reversal for multimedia redundancy effect. *Computers in Human Behavior, 27*(1), 11–15.

Leu, D. J., Kinzer, C. K., Coiro, J., Castek, J., & Henry, L. A. (2018). New literacies: A dual-level theory of the changing nature of literacy, instruction, and assessment. In *Theoretical models and processes of literacy* (pp. 319–346). Routledge.

Lim, D. H. (2004). Cross cultural differences in online learning motivation. *Educational Media International, 41*(2), 163–175. https://doi.org/10.1080/09523980410001685784

Lin, X., Featherman, M., & Sarker, S. (2017). Understanding factors affecting users' social networking site continuance: A gender difference perspective. *Information & Management, 54*(3), 383–395.

Lin, L., & Johnson, T. (2021). Shifting to digital: Informing the rapid development, deployment, and future of teaching and learning. *Educational Technology Research & Development, 69*(1), 1–5.

Liu, S.-H. (2017). Relationship between the factors influencing online help-seeking and self-regulated learning among Taiwanese preservice teachers. *Computers in Human Behavior, 72*, 38–45.

Liu, J. (2020). Absolute Shapley Value. *Proceedings of the VLDB Endowment, 12*. https://doi.org/10.48550/arXiv.2003.10076.

Liu, R., Alvarado-Urbina, A., & Hannum, E. (2020). Differences at the extremes? Gender, national contexts, and math performance in Latin America. *American Educational Research Journal, 57*(3), 1290–1322. https://doi.org/10.3102/0002831219876236

Lundberg, Scott (2018). bar plot — SHAP latest documentation. https://shap.readthedocs.io/en/latest/example_notebooks/api_examples/plots/bar.html.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems, 31*, 1–10. https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html.

Mathrani, A., Sarvesh, T., & Umer, R. (2021). Digital divide framework: Online learning in developing countries during the COVID-19 lockdown. *Globalisation, Societies and Education*, 1–16. https://doi.org/10.1080/14767724.2021.1981253, 0(0).

McKinney, W. (2010). Data structures for statistical computing in python. *Proceedings of the 9th Python in Science Conference, 445*, 51–56.

Meinz, E. J., & Hambrick, D. Z. (2010). Deliberate practice is necessary but not sufficient to explain individual differences in piano sight-reading skill: The role of working memory capacity. *Psychological Science, 21*(7), 914–919. https://doi.org/10.1177/0956797610373933

Molnar, C. (2021). *Interpretable machine learning: A guide for making black box models explainable*. Leanpub.com https://christophm.github.io/interpretable-ml-book.

Mseleku, Z. (2020). A literature review of E-learning and E-teaching in the era of Covid-19 pandemic. *SAGE, 57*(52), 588–597.

Nelson, L. J., & Fyfe, E. R. (2019). Metacognitive monitoring and help-seeking decisions on mathematical equivalence problems. *Metacognition and Learning, 14*(2), 167–187. https://doi.org/10.1007/s11409-019-09203-w

Nemanich, L., Banks, M., & Vera, D. (2009). Enhancing knowledge transfer in classroom versus online settings: The interplay among instructor, student, content, and context. *Decision Sciences Journal of Innovative Education, 7*(1), 123–148. https://doi.org/10.1111/j.1540-4609.2008.00208.x

Newman, R. S. (2002). How self-regulated learners cope with academic difficulty: The role of adaptive help seeking. *Theory Into Practice, 41*(2), 132–138.

Noah's Ark 2.0. (2020). December 3). 6 11 game theory shapley value SciShow. https://www.youtube.com/watch?v=90liQP8pIOM.

Nokes, T. J. (2009). Mechanisms of knowledge transfer. *Thinking & Reasoning, 15*(1), 1–36.

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.
OECD. (2019). *Measuring the digital transformation: A roadmap for the future*. OECD. https://doi.org/10.1787/9789264311992-en
OECD. (2020). *Covid-19: Global action for a global crisis—OECD*. https://read.oecd-ilibrary.org/view/?ref=120_120544-8ksud7oaj2&title=Education_responses_to_Covid-19_Embracing_digital_learning_and_online_collaboration&_ga=2.231710746.448289388.1632396451-1020876759.1632396451.
OECD. (2021). *The state of school education: One year into the COVID pandemic*. https://doi.org/10.1787/201dde84-en
Ogan, A., Walker, E., Baker, R., Rodrigo, M. M. T., Soriano, J. C., & Castro, M. J. (2015). Towards understanding how to assess help-seeking behavior across cultures. *International Journal of Artificial Intelligence in Education, 25*(2), 229–248.
Omura, M. (2021). *Effects of repeated reading instruction in online EFL classes for Japanese high school students [masters thesis*. Researchgate: Sophia University.
Palmer, D. J., Stough, L. M., Burdenski, J. T. K., & Gonzales, M. (2005). Identifying teacher expertise: An examination of researchers' decision making. *Educational Psychologist, 40*(1), 13–25, 10/fr939j.
Pasha, A., Abidi, S., & Ali, S. (2016). Challenges of offering a MOOC from an LMIC. *International Review of Research in Open and Distance Learning: IRRODL, 17*(6), 221–228.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.
Penney, B. (2019). *The English indices of deprivation 2019* (pp. 1–31).
Putnick, D. L., & Bornstein, M. H. (2016). Girls' and boys' labor and household chores in low- and middle-income countries. *Monographs of the Society for Research in Child Development, 81*(1), 104–122. https://doi.org/10.1111/mono.12228
Ratner, B. (2011). *Statistical and machine-learning data mining: Techniques for better predictive modeling and analysis of big data* (2nd ed.). CRC Press LLC http://ebookcentral.proquest.com/lib/cam/detail.action?docID=840391.
Reagans, R., Argote, L., & Brooks, D. (2005). Individual experience and experience working together: Predicting learning rates from knowing who knows what and knowing how to work together. *Management Science, 51*(6), 869–881.
Rebello, N. S., Cui, L., Bennett, A. G., Zollman, D. A., & Ozimek, D. J. (2017). Transfer of learning in problem solving in the context of mathematics and physics. In *Learning to solve complex scientific problems* (pp. 223–246). Routledge.
Reychav, I., & McHaney, R. (2017). The relationship between gender and mobile technology use in collaborative learning settings: An empirical investigation. *Computers & Education, 113*, 61–74.
Rodríguez-Pérez, R., & Bajorath, J. (2020). Interpretation of machine learning models using shapley values: Application to compound potency and multi-target activity predictions. *Journal of Computer-Aided Molecular Design, 34*(10), 1013–1026. https://doi.org/10.1007/s10822-020-00314-0
Rosé, C. P., McLaughlin, E. A., Liu, R., & Koedinger, K. R. (2019). Explanatory learner models: Why machine learning (alone) is not the answer. *British Journal of Educational Technology, 50*(6), 2943–2958. https://doi.org/10.1111/bjet.12858
Schlenz, M. A., Schmidt, A., Wöstmann, B., Krämer, N., & Schulz-Weidner, N. (2020). Students' and lecturers' perspective on the implementation of online learning in dental education due to SARS-CoV-2 (COVID-19): A cross-sectional study. *BMC Medical Education, 20*(1), 1–7.
Sculley, D., & Pasanek, B. M. (2008). Meaning and mining: The impact of implicit assumptions in data mining for the humanities. *Literary and Linguistic Computing, 23* (4), 409–424. https://doi.org/10.1093/llc/fqn019
Seeley, E. A., & Gardner, W. L. (2003). The "selfless" and self-regulation: The role of chronic other-orientation in averting self-regulatory depletion. *Self and Identity, 2*(2), 103–117. https://doi.org/10.1080/15298860309034
Shilling, C. (1991). Social space, gender inequalities and educational differentiation. *British Journal of Sociology of Education, 12*(1), 23–44.
Srite, M., & Karahanna, E. (2006). The role of espoused national cultural values in technology acceptance. *MIS Quarterly, 30*(3), 679–704.
Sternberg, R. J. (1999). Intelligence as developing expertise. *Contemporary Educational Psychology, 24*(4), 359–375, 10/dzvjsj.
Sternberg, R. J. (2001a). Giftedness as developing expertise: A theory of the interface between high abilities and achieved excellence. *High Ability Studies, 12*(2), 159–179, 10/bn8pss.
Sternberg, R. J. (2001b). Metacognition, abilities, and developing expertise: What makes an expert student?. In *Metacognition in learning and instruction* (pp. 247–260). Springer.
Stevens, M. (2020). Expertise, complexity, and self-regulated engagement: Lessons from teacher reflection in a blended learning environment. *Journal of Online Learning Research, 6*(3), 177–200.
Trommsdorff, G. (2010). Teaching and learning guide for: Culture and development of self-regulation. *Social and Personality Psychology Compass, 4*(4), 282–294. https://doi.org/10.1111/j.1751-9004.2009.00257.x
Tronstad, T. (2021). Occupational health in developing countries–a massive open online course. *Archives of Environmental & Occupational Health, 76*(5), 291–300.
Umar, K. I., Basheer, M. M., Isa, R., & Watsilla, H. (2017). Educational technologies: A way to enhance learning process in Nigerian tertiary institutions. *International Journal of Engineering Technology Research & Management, 1*(3), 10–16.
Undorf, M., Livneh, I., & Ackerman, R. (2021). Metacognitive control processes in question answering: Help seeking and withholding answers. *Metacognition and Learning, 16*(2), 431–458. https://doi.org/10.1007/s11409-021-09259-7
Venkatesh, V., & Zhang, X. (2010). Unified theory of acceptance and use of technology: US vs. China. *Journal of Global Information Technology Management, 13*(1), 5–27.
Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin, 140*(4), 1174.
Worcester, P. (2019). *June 6). A comparison of grid search and randomized search using scikit learn. Medium* https://medium.com/@peterworcester_29377/a-comparison-of-grid-search-and-randomized-search-using-scikit-learn-29823179bc85.
Yates, A., Starkey, L., Egerton, B., & Flueggen, F. (2021). High school students' experience of online learning during Covid-19: The influence of technology and pedagogy. *Technology, Pedagogy and Education, 30*(1), 59–73. https://doi.org/10.1080/1475939X.2020.1854337
Yeung, A. S. (1999). Cognitive load and learner expertise: Split-attention and redundancy effects in reading comprehension tasks with vocabulary definitions. *The Journal of Experimental Education, 67*(3), 197–217.
Yeung, K. L., Carpenter, S. K., & Corral, D. (2021). A comprehensive review of educational technology on objective learning outcomes in academic contexts. *Educational Psychology Review, 33*(4), 1583–1630.
Zhang, S., & Duke, N. K. (2008). Strategies for internet reading with different reading purposes: A descriptive study of twelve good internet readers. *Journal of Literacy Research, 40*(1), 128–162. https://doi.org/10.1080/10862960802070491