Mapping from proximity traces to socio-spatial behaviours and student progression at the school

Yan, Lixiang, Martinez-Maldonado, Roberto, Gallo Cordoba, Beatriz, Deppeler, Joanne, Corrigan, Deborah and Gasevic, Dragan (2022) Mapping from proximity traces to socio-spatial behaviours and student progression at the school. British Journal of Educational Technology, 53 (6). pp. 1645-1664. ISSN 0007-1013

Received: 26 April 2021 | Accepted: 3 February 2022

DOI: 10.1111/bjet.13203

ORIGINAL ARTICLE

British Journal of Educational Technology BERA

Mapping from proximity traces to socio-spatial behaviours and student progression at the school

Lixiang Yan[1] | **Roberto Martinez-Maldonado**[1] | **Beatriz Gallo Cordoba**[2] | **Joanne Deppeler**[2] | **Deborah Corrigan**[2] | **Dragan Gašević**[1]

[1]Centre for Learning Analytics at Monash, Faculty of Information Technology, Monash University, Clayton, Victoria, Australia

[2]Faculty of Education, Monash University, Clayton, Victoria, Australia

Correspondence
Lixiang Yan, Centre for Learning Analytics at Monash, Faculty of Information Technology, Monash University, 20 Exhibition Walk, Clayton, VIC 3800, Australia.
Email: lixiang.yan@monash.edu

Funding information
Roberto Martinez-Maldonado's research is partly funded by Jacobs Foundation.

Abstract

Identifying students facing difficulties and providing them with timely support is one of the educator's key responsibilities. Yet, this task is becoming increasingly challenging as the complexity of physical learning spaces grows, along with the emergence of novel educational technologies and classroom designs. There has been substantial research and development work focused on identifying student social behaviours in digital platforms (eg, the learning management system) as predictors of academic progression. However, little work has investigated such relationships in physical learning spaces. This study explores the potential of using wearable trackers for the early detection of low-progress students based on their social and spatial (socio-spatial) behaviours at the school. Positioning data from 98 primary school students and six teachers were automatically captured over a period of eight weeks. Fourteen socio-spatial behavioural features were extracted and processed using a set of machine learning classifiers to model students' learning progression. Results illustrate the potential of prospectively identifying low-progress students from these features and the importance of adapting classroom learning analytics to differences in pedagogical designs.

KEYWORDS
multimodal learning analytics, positioning tracking, predictive learning analytics, proximity, student progression

Practitioner notes

What is already known about this topic

- Learning analytics research on predicting students' academic progression is emerging in both digital and physical learning spaces.
- Students' social behaviours in learning activities is a key factor in predicting their academic progression.
- Emerging sensing technologies can provide opportunities to study students' real-time social behaviours in physical learning spaces.

What this paper adds

- Fourteen progression-related socio-spatial behavioural features are extracted from students' physical (*x*-*y*) positioning traces.
- Predictive learning analytics that achieved 81% accuracy in prospectively identifying low-progress students from their real-time socio-spatial behaviours.
- Empirical evidence to support the need for classroom learning analytics to have instructional sensitivity (ie, be calibrated according to the learning design).

Implications for practice and/or policy

- Sensing technologies and machine learning algorithms can be used to capture and generate valuable insights about higher-order learning constructs (eg, performance and collaboration) from students' physical positioning traces in classrooms.
- Researchers and practitioners should be cautious with generalised classification algorithms and predictive learning analytics that do not account for the pedagogical differences between different subjects or learning designs.
- Researchers and practitioners should consider the potentially unforeseen ethical issues that can emerge in using sensing technologies and predictive learning analytics in authentic, physical classroom settings.

INTRODUCTION

The early identification of students who may be facing challenges that can slow down their progress in their learning journeys is critical for teachers to provide them with timely support (Gray & Perkins, 2019). Timely and appropriate support can then translate into improved student engagement and achievement (Klem & Connell, 2004). Advances in learning analytics have made it possible to create a range of early-warning systems that can identify low-progress students based on their behaviour (Nam & Samson, 2019). The majority of these early-warning tools rely on prior information about the students and behavioural traces digitally captured from online learning platforms (Hellas et al., 2018). For example, predictions on student outcomes have been made based on student behaviours in MOOCs (Tomkins et al., 2016) and the LMS (Zacharis, 2015). However, little progress has been made in prospectively detecting low-progress students based on their activity in physical learning spaces.

Analysing physical trace data can serve to generate a deeper understanding of the growing complexity of learning spaces. These are being constantly enriched with novel digital technologies (Goodyear, 2020) and (re-)shaped according to emerging architectural approaches (eg, open learning spaces; Reh et al., 2011), which are challenging current pedagogical practices. Emerging sensing and analytic tools can potentially be created to help teachers cope with this complexity by automatically identifying students who may need closer attention. As sensing technologies are becoming ready for their widespread use in educational applications (Chua et al., 2019), analysing students' physical positioning data traces can be a promising direction to characterise students' social and spatial behaviours (socio-spatial) in the classroom. These traces can serve to extract indicators of student interaction with peers and teachers (Chng et al., 2020; Martinez-Maldonado, Echeverria, et al., 2020; Yan et al., 2021), physical activity (Saquib et al., 2018), sustained engagement (Chin et al., 2017; Chng et al., 2020), and help-seeking behaviours (Fernandez Nieto et al., 2021), all closely related to learning progression.

This paper presents a study that explores the potential of using wearable, indoor positioning trackers for automatically mapping socio-spatial behaviours of low-progress students at the school level from low-level proximity traces. Positioning data from 98 primary school students and six teachers were automatically captured over eight weeks using wearable trackers (called wearables for short). Based on the theoretical foundations of *proxemics* (Hall, 1966), measures of proximity among students and teachers were obtained. Based on these, fourteen socio-spatial behavioural features were extracted and processed using a set of machine learning classifiers to model student learning progression. To our knowledge, this is the first longitudinal study exploring the potential of modelling physical proximity traces in identifying low-progress students. This paper contributes to addressing this gap by exploring whether it is possible to model social interactions from physical positioning data with the purpose of identifying potentially low-progress students prospectively.

LITERATURE REVIEW

Students' social behaviours have been frequently investigated to predict student progression in online learning platforms (see review in Hellas et al. [2018]). Thus far, most of these prediction models have been applied to digital traces captured from online learning systems, such as digital forums, to estimate educational constructs like students' social interaction and social presence in a course (Brinton & Chiang, 2015; Joksimovic et al., 2015; Tomkins et al., 2016; Zacharis, 2015). By contrast, early detection of low-progress students in physical learning spaces has mainly relied on static data sources such as student demographics, prior progression, and psychometrics (Hellas et al., 2018). The potential progression-related insights that could be obtained from the analysis of real-time socio-spatial behaviours of students remain underexplored. In the remainder of this section, we briefly describe the area of study, focusing on understanding how people use the physical space to enable social interactions (proxemics), and related educational technology research.

Theoretical foundation: Proxemics

Recent developments in the learning analytics community have demonstrated the potential of automatically modelling socio-spatial behaviours from indoor positioning traces (Chng et al., 2020; Martinez-Maldonado, Echeverria, et al., 2020; Yan et al., 2021). These studies have built on the theoretical foundation of proxemics, which refers to the study of human space usage in social contexts (Hall, 1966). Mondada's (2013) work on interactional space has suggested the use of proximity as a proxy of potential interactions in physical spaces. In fact, findings in social psychology have demonstrated that proximity metrics can serve to estimate students' social interaction and the formation of social ties with others (Back et al., 2008). This means that, in schools, the physical distance or proximity among students and teachers during personal, interpersonal, and group activities can serve as a proxy of the social dynamics that emerge according to learning tasks and the characteristics of the learning space.

Prior research in social dynamics has illustrated the association between students' socio-spatial behaviours and their academic progression. For example, physical activity within learning spaces has been found to be positively associated with students' cognitive function and academic progression (Donnelly et al., 2016). Students' in-class interactions could reflect their social capital and peer relationships, which can be related to higher academic progression (Gasevic et al., 2013; Wentzel, 2017). Additionally, the probabilities of transitions between different interaction states can also hold potentially valuable information about students' learning performance (Chng et al., 2020). The probability of maintaining the same interaction state (eg, maintaining interaction with students) can reflect students' sustained engagement in individual or group learning activities, which has been positively related to academic attainment in observational classroom studies (Guthrie et al., 2012). Transitions between different interaction states (eg, from learning individually to interacting with students) can reflect students' help-seeking behaviours, which have been found to predict their grades (Ryan & Shin, 2011). Capturing and modelling these socio-spatial behaviours could potentially support teachers to identify low-progress students, whose academic progression is lower than the state government indication scores for their year-level. These students might experience problems during learning activities and might be at risk of falling behind academically.

Sensing technologies and learning analytics

Four types of sensing technologies have been used in studies focused on modelling social interaction based on positioning data. Both WiFi (Nguyen et al., 2020) and thermal sensors (Brennan et al., 2018) have been used to detect students' physical presence in classrooms. These technologies offer coarse spatio-temporal precision which works well to identify if two people are roughly in the same room. They are, however, unsuitable for capturing granular socio-spatial behaviours of individual students and teachers. Advances in computer vision can offer fine granularity of positioning data. For example, Chng et al. (2020) and Ahuja et al. (2019) used Kinect sensors and video cameras (respectively) to capture motion and posture data, which were then translated to *x*-*y* coordinates using a reference grid system, and further estimated proximity among students and teachers. However, this method is limited to either lecture-style or small classroom spaces since visual occlusion can damper the precision and continuity of the positioning tracking in larger and more dynamic learning spaces (ie, at a school; Martinez-Maldonado, Mangaroska, et al., 2020).

The fourth approach to capture students' positioning data involves using wearable trackers. Martinez-Maldonado, Mangaroska, et al. (2020) used sensors worn around the waist to automatically capture teachers' movement and positioning strategies in a classroom. Saquib et al. (2018) also used tiny wearable trackers attached to students' and teachers' shoes to visualise the physical dynamics in a Montessori school. Similarly, Yan et al. (2021) also used wearable trackers in the form of wristbands to capture individual social interactions and group social dynamics in large learning spaces, but their discussion limited to model behaviours for descriptive purposes using unsupervised techniques.

Research gap and research questions

The studies presented above jointly demonstrate the potential of using wearables to capture multiple spatial behaviour features from fine-grained positioning data. However, none of these studies has explored the potential of automatically modelling socio-spatial behaviours in order to identify low-progress students (using supervised machine learning techniques), despite the long records of significant relationships between socio-spatial behaviours and student progression (further elaborated in the next section). Thus, to explore this potential implication, the following research questions are investigated:

RQ1. Do students' socio-spatial behaviours captured using wearables exhibit significant relationships with their academic progression?
RQ2. How accurately can students' socio-spatial behaviours be modelled to identify low-progress students prospectively?
RQ3. What are the most important socio-spatial behavioural features that can be used to identify low-progress students prospectively?

METHODS

Study context

The current study took place in an open-plan primary school equipped with movable furniture that can be rearranged by teachers and students. Figure 1 shows a section of the open-plan learning space and Figure 2 shows the floor plan with live positioning data points from students and their teachers in the physical learning space. In this paper, we focus on Maths and Reading sessions taught within the building area. In these two subjects, students were allocated into four different groups based on their academic progression at the end of their previous school year. Each group was assigned to one teacher. However, students could interact with students from other groups at any time since they were all collocated in the same open-plan learning space. Our previous work explored how graph-based approaches could be used to describe interactions between students and their change over time, finding that teachers' ability grouping strategy has no impact on students' social interaction preferences Yan et al. (2021).

The pedagogical approach for Reading involved several instructed group-based activities, whereas students could choose to study in groups or individually in Maths. After signing informed consents, a total of 98 Year-6 students (47 females, 49 males, and two unspecified; 11 to 12 years of age) and six teachers (four full-time teachers, one aide, and one part-time teacher; three females and three males) participated in the study. All these students took both Math and Reading subjects, and the same teachers taught both subjects. Ethics approval was granted by [Anonymous] University and the Department of Education and Training of the State of [Anonymous] in [Anonymous]. Parental consent was also obtained prior to students' participation.

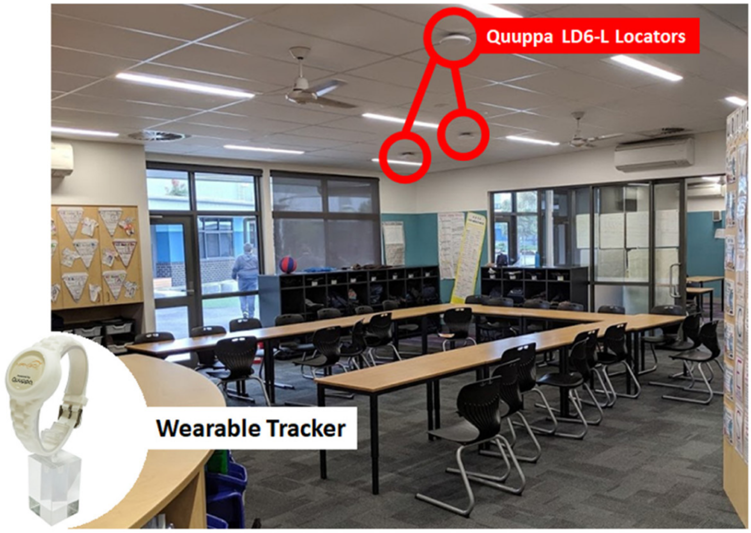

FIGURE 1 Open-plan learning space with positioning tracking system installed

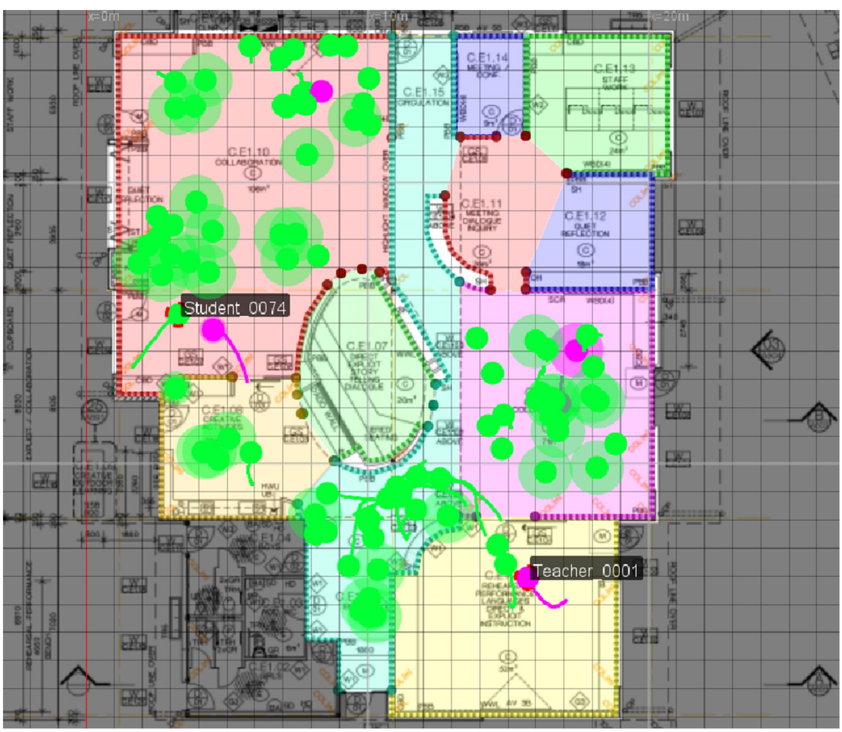

FIGURE 2 Floor plan of the open-plan learning space with green and purple dots representing students and teachers' recorded location, respectively

Apparatus and data collection

A total of 14 Quuppa LD6-L locators (quuppa.com) were placed at various locations on the ceilings of the learning spaces and connected to a server running Quuppa's proprietary positioning engine (as illustrated in Figure 1). Each participant was assigned a BLE (Bluetooth Low Energy) tag; Tatwah Mango BLE-WB200 wristbands for students and Jeewey JW-C1809C card tags for teachers. All tags were individually numbered and preset to transmit tracking information at 5Hz. The tracking information was received by the locators and translated by the tracking algorithms, utilising the Angle-of-Arrival methodology (Suryavanshi et al., 2019), into real-time *x*-*y* locations, accurate to approximately 200 mm and 20 ms.

Teachers in the study kept a register of the tag numbers and distributed wristbands to the students, recording the identifying number, at the beginning of every school day and collecting them back at the end of the day. If a student lost their wristband, a teacher gave them a replacement and amended the register. A single data point consists of a timestamp, a tracking identifier, and *x*-*y* coordinates of the learning space's floor-plan in meters (eg, 22/07/2019 11:38:24.000, Student0001, 7.3875, 20.675). A total of 62.15 million data points were recorded during Maths (35 sessions) and Reading (23 sessions) over eight school weeks from July 22 to September 13, 2019. The average duration of each Maths and Reading session was one hour.

Three months after the data collection, students' academic progression scores for Maths and Reading were measured using state-wide standardised testing for the 2019 school year. Students with missing progression scores or less than half attendance during the data collection period were excluded from the analysis. The final sample sizes were 96 and 87 students for Maths and Reading, respectively.

Data processing and feature extraction

In order to extract socio-spatial behaviour features from fine-grained physical positioning traces, the dataset was first normalised by averaging data to one data point per second. Linear interpolation was then applied to fill in any missing values caused by occlusion or temporary detachment of trackers (Gløersen & Federolf, 2016). This interpolation was limited to missing values between two valid data points, and for less than 60 consecutive missing values; otherwise, students were considered outside of the tracking area.

A total of 14 socio-spatial behaviour features were extracted for each student and session. These were grouped into three types in relation to (1) body movement, (2) social interaction, and (3) state-transitions (see Table 1 for details). The rationale and extraction procedures for each feature are elaborated on below.

Movement features

Physical activity can be extracted from students' physical positioning traces by calculating the time and distance of walking. *DistanceMoved* in a session was calculated as the sum of the Euclidean distance between a students' current and previous position on a one-second basis. *TimeMoved* was calculated as the sum of all the seconds for which *DistanceMoved* is greater than zero.

TABLE 1 Socio-spatial behavioural features from positioning traces

Behaviours	Features	Description
Movement	*DistanceMoved*	Distance walked (meters)
	TimeMoved	Time spent in walking motion (seconds)
Interactions	*TimeIndividual*	Time the student spent by herself (seconds)
	TimeStudent	Time spent in close proximity to peers (seconds)
	TimeTeacher	Time spent in close proximity to a teacher (seconds)
State	*I-I, S-S, T-T*	Probability of maintaining in the same state
Transitions	*I-S, I-T* *S-I, S-T* *T-I, T-S*	Probability of transit from one state to a different state, where: *I*—learning individually *S*—interacting with other students *T*—interacting with teachers

Interaction features

Previous works (Back et al., 2008; Chng et al., 2020) have estimated the potential occurrence of social interactions between two people by measuring the duration of collocation based on proximity. To model this, a four-step extraction process was performed:

1. Interpersonal distances among all students were extracted by calculating the Euclidean distances between each tag. This step involved calculating all possible pair combinations for each second.
2. A potential instance of social interaction was identified if two or more tags were within one-meter proximity of each other for more than ten consecutive seconds, as modelled in previous works (Chng et al., 2020; Martinez-Maldonado, Schulte, et al., 2020; Yan et al., 2021). This ten-second constraint minimises the false identification of unintended collocation, for example, when teachers are walking around during supervision or two students are passing by each other (Greenberg et al., 2014).
3. Each second for each student was labelled as *I* (individual), *S* (student), or *T* (teacher) representing different types of interaction if the student was alone, near peers or close to a teacher, respectively.
4. Three features about student interactions were then extracted by counting the number of seconds in each state, including *TimeIndividual*, *TimeStudent* and *TimeTeacher* (Table 1).

State-transition features

A total of nine different state transition features were extracted using Markov chains (Table 1). This extraction was performed by calculating a one-step transition matrix with three different states, including *I*, *S*, and *T*. For example, the feature *I-I* represents the probability of a student to stay in individual learning. The feature *I-S* represents the probability of a transition from individual learning to learning with other students.

Analysis

Three analyses were performed to investigate the relationships between students' socio-spatial behavioural features and their academic progression.

Correlation analysis—RQ1

A set of correlation analyses has been performed to investigate the linear relationships between students' socio-spatial behaviour features and their academic progression. Aggregated averages of each feature were calculated for progression in Maths and Reading and normalised to investigate the linear relationship between these features and students' academic progression using Pearson's correlation coefficient. The Bonferroni correction method was applied to adjust the significance threshold (initial alpha equal 0.05) for multiple comparisons. Significant correlations were summarised into tables to illustrate the set of socio-spatial behaviours that are linearly related to students' academic progression in Maths and Reading.

Predictive analytics—RQ2

State-of-the-art machine learning (ML) algorithms were used to early detect low-progress students using the socio-spatial behaviour features. The following steps were implemented to construct the predictive models:

1. Classifying *low-progress* students. Students with academic progression scores lower than the *state government indication scores* for their year-level were labelled as *low-progress* and were the positive class in the binary classification. This classification strategy was chosen as students who progressed below the *state government indication scores* for their year-level may have been at the risk of falling behind academically. This disadvantage could affect their future studies, especially since they are reaching the transition from primary to secondary education.
2. Balancing classes. The ratios of low-progress students to other students in both Maths and Reading were around 1:2. Thus, to deal with this class imbalance, the current study implemented a 3-fold cross-validation (repeated 20 times) and applied the Synthetic Minority Oversampling Technique (SMOTE; Chawla et al., 2002) exactly inside the cross-validation loop to oversampling the data while preventing potential data contamination (Farrow et al., 2019). After applying the SMOTE, the class would be balanced with an equal proportion of low-progress students and other students (class ratio 1:1; accuracy of dummy models would be 50%) in both Maths and Reading. The 3-fold cross-validation was chosen to accommodate the relatively small sample size (98 students) and reduce potential issues of the training set not representing the test set. This also ensures that there is sufficient data for the validation sets and the grid search process. Together, these two approaches provide more stable and accurate results.
3. Selecting ML classifiers. ML classifiers from the Python Scikit-Learn library (Pedregosa et al., 2011) were used to explore the optimal classifiers for identifying low-progress students. We used the following three types of ML algorithms that have been commonly used for binary classification problems in learning analytics (Hellas et al., 2018): (i) linear classifiers, including logistic regression (LR) and support vector machine (SVM); non-parametric classifiers, including random forests (RF) and k-nearest neighbours (KNN); and an artificial neural network (ANN) was developed using the multi-layer perceptron algorithm. These classifiers were trained separately for Maths and Reading as the pedagogical designs for these two subjects were different. All socio-spatial features listed in Table 1 were used in each classifier.
4. Evaluating model performance. Model evaluation metrics were used to compare the classifiers based on prediction performances (including accuracy, precision, recall, and Area Under the Curve; AUC) and inter-rater reliability (Cohen's kappa; k). Grid Search was

used for hyperparameter tuning to optimise recall and minimise false negatives (Joseph, n.d.). Details of the hyperparameter used for each model are available in the Appendix. The ultimate goal is to avoid incorrect low-progress classifications, which would misdirect teachers and leave these students unattended and without help. The mean and standard deviation (*SD*) of these metrics are reported as average values across each iteration of cross-validation.

Feature evaluation—RQ3

Each feature's contribution to the prediction was assessed using the Shapley values to enhance interpretability (Strumbelj & Kononenko, 2014). This metric was chosen because the Shapely Additive Explanation (SHAP) combined all the prior model interpretation techniques and was suitable for explaining all machine learning classifiers (Lundberg & Lee, 2017). SHAP feature analysis is also distinct from other permutation-based evaluations as it provides the most granular results regarding the feature-level influence on prediction. The best performing classifier was chosen as the kernel for running SHAP. Each spatial feature was assigned with a SHAP value based on its impact on the model. Based on the SHAP values, an additional feature selection process has been conducted to illustrate the model improvements by removing features with little contribution to the models.

RESULTS

Feature correlations—RQ1

In Maths, *progression* was positively correlated with *TimeStudent*, *I-I*, *MovedTime*, *MovedDistance*, and negatively correlated with *S-I* and *S-T* (Table 2, left). In Reading, *progression* was positively correlated with *I-S*, *TimeStudent*, *S-S*, and negatively correlated with *S-I*, and *TimeIsolate* (Table 2, right). These findings can be interpreted as, for example, the students who spent more time with other students, showed an increase in their academic progression. Although the correlations between *progression* and three features (*TimeStudent*, *S-S*, and *S-I*) were consistent across both subjects, other features were subject-dependent. These discrepancies can be explained by differences in pedagogical design. For example, *TimeIsolate* and *I-S* may be more relevant in the Reading classes since high *TimeIsolate* may indicate less participation in the instructed group activities which is against the pedagogical intention of the learning design for Reading, and vice-versa for *I-S*.

Classifier performance—RQ2

The classifiers' performance is shown in Tables 3 and 4 for Maths and Reading, respectively. Linear classifiers demonstrated better performance than others. In particular, LR with a LIBLINEAR solver and a regulation strength of one demonstrated the best performance in both Maths and Reading. Likewise, SVM with linear kernel and a regulation strength of one showed similar performance. The best classifier could accurately identify around 70% and 61% of the low-progress students in Reading and Maths, respectively, based on their socio-spatial behaviours. This is above the chance of random guessing (50% for dummy classifiers with balanced class after applying SMOTE). This result was expected from the moderate correlation results in RQ1, and the results from prior studies (Donnelly et al., 2016;

TABLE 2 Significant correlations between socio-spatial behavioural features and student progression in Maths (left) and Reading (right)

Maths		Reading	
Features	**Correlation (*r*)**	**Features**	**Correlation (*r*)**
TimeStudent	0.49***	*TimeStudent*	0.36***
MovedTime	0.30**	*TimeIsolate*	−0.26*
MovedDistance	0.29**	*I-S*	0.42***
I-I	0.39***	*S-S*	0.26*
S-S	0.43***	*S-I*	−0.28**
S-I	−0.39***		
S-T	−0.22*		

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

TABLE 3 Model performance before feature selection (Maths): Mean and *SD* (in parenthesis)

	Accuracy	Precision	Recall	Cohen's *k*	AUC
LR	0.77 (0.06)	0.66 (0.06)	0.70 (0.06)	0.50 (0.06)	0.75 (0.06)
SVM	0.75 (0.06)	0.62 (0.06)	0.66 (0.06)	0.44 (0.06)	0.73 (0.06)
RF	0.76 (0.06)	0.65 (0.06)	0.64 (0.06)	0.45 (0.06)	0.73 (0.06)
KNN	0.70 (0.07)	0.54 (0.07)	0.63 (0.07)	0.34 (0.07)	0.68 (0.07)
ANN	0.73 (0.07)	0.61 (0.07)	0.60 (0.07)	0.39 (0.07)	0.69 (0.07)

TABLE 4 Model performance before feature selection (Reading): Mean and *SD* (in parenthesis)

	Accuracy	Precision	Recall	Cohen's *k*	AUC
LR	0.69 (0.06)	0.45 (0.06)	0.61 (0.06)	0.29 (0.06)	0.66 (0.06)
SVM	0.68 (0.07)	0.44 (0.07)	0.59 (0.07)	0.27 (0.07)	0.65 (0.07)
RF	0.70 (0.07)	0.45 (0.07)	0.41 (0.07)	0.22 (0.07)	0.61 (0.07)
KNN	0.59 (0.09)	0.35 (0.09)	0.58 (0.09)	0.14 (0.09)	0.58 (0.09)
ANN	0.71 (0.08)	0.49 (0.08)	0.50 (0.08)	0.28 (0.08)	0.64 (0.08)

Guthrie et al., 2012; Wentzel, 2017), which all point to linear relationships between these features and student progression. The other classifiers (RF, KNN, and ANN) might have under-performed due to the relatively small sample size, which could have reduced the power of these classifiers to discover complex relationships between the features and student progression (Entezari-Maleki et al., 2009).

Despite training the classifiers separately for Maths and Reading, there is still a large discrepancy between the classifiers' performance for these two subjects. For Reading, the recall, precision, and Cohen's *k* values are concerning as these low values can potentially indicate high false-positive and false-negative cases, and poor inter-rater reliability. A potential reason behind the under-performing classifiers in Reading could be related to the pedagogical differences and irrelevant input features. For example, interaction features (ie, *TimeStudent*) could have a stronger contribution in predicting low-progress students as they might contain information about collaborative learning behaviours, which is essential in group-oriented reading sessions. Whereas, movement features might be less relevant since they contain little information about group learning. These interpretations are supported by the findings on feature importance, which are reported in the next section.

Feature importance—RQ3

The results of the feature evaluation through SHAP with the LR kernel (the best performing classifier in the prior section) are shown in Figure 3 (Maths) and Figure 4 (Reading).

Features were listed according to their importance and visualised using summary plots to illustrate the SHAP value distributions, where lighter blue and darker red represent lower and higher feature values, respectively. A positive SHAP value increases the likelihood of a student being predicted as low progress, whereas a negative SHAP value decreases this likelihood. For example, the T-S feature (see Figure 4) contributes more to increasing the likelihood of a student being classified as low-progress than to decreasing this likelihood.

Overall, the selected features had stronger contributions for Maths than for Reading, as shown by the narrower and T-shaped distributions in Reading. For both subjects, *TimeStudent* was the most significant feature, and it was negatively associated with the likelihood of a student being classified as low progress. That is, students who spent more time with other students were less likely to be classified as a low-progress student, and vice-versa. *TimeIsolate* was also a significant feature in both subjects but exhibited a positive association. Other features appear to have resonated with the subjects' pedagogical design. For example, both *MovedTime* and *I-I* had a significant positive association with Maths but became less relevant for Reading.

After feature selection, models with seven features, including *TimeIsolate*, *TimeStudent*, *S-I*, *MovedTime*, *I-I*, *T-I*, and *T-S*, showed the best performances in Maths (Table 5). LR remained the best performing classifier in accurately identifying low-achieving students (81%). For Reading, models with six features, including *TimeIsolate*, *TimeStudent*, *I-S*, *S-S*, *S-T*, and *T-S* demonstrated the best performances (74%; Table 6). These performance improvements may indicate the need to select appropriate features that are relevant to the pedagogical design of the learning activities.

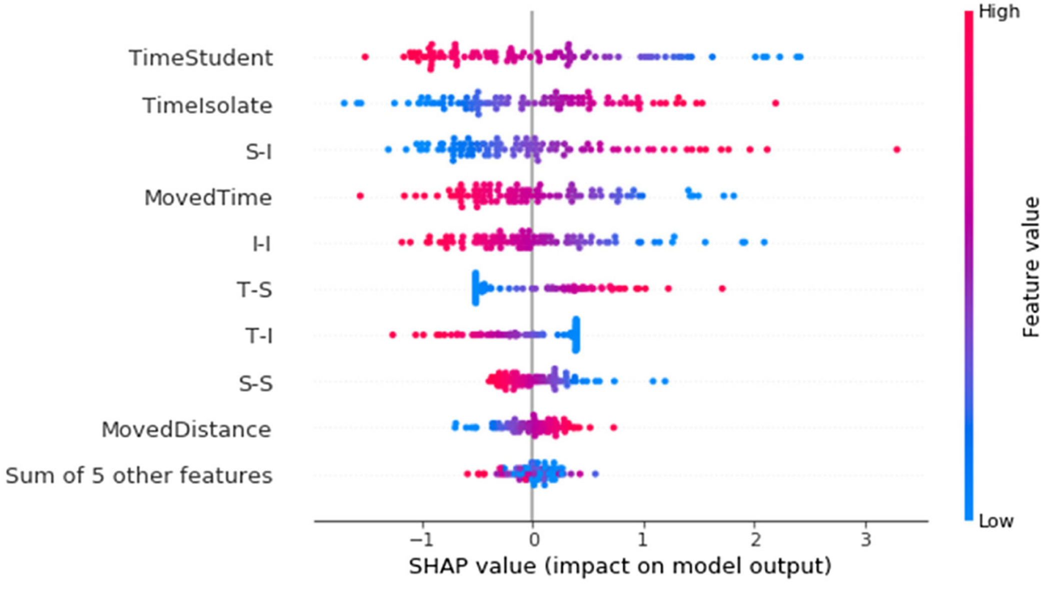

FIGURE 3 SHAP features importance for Maths

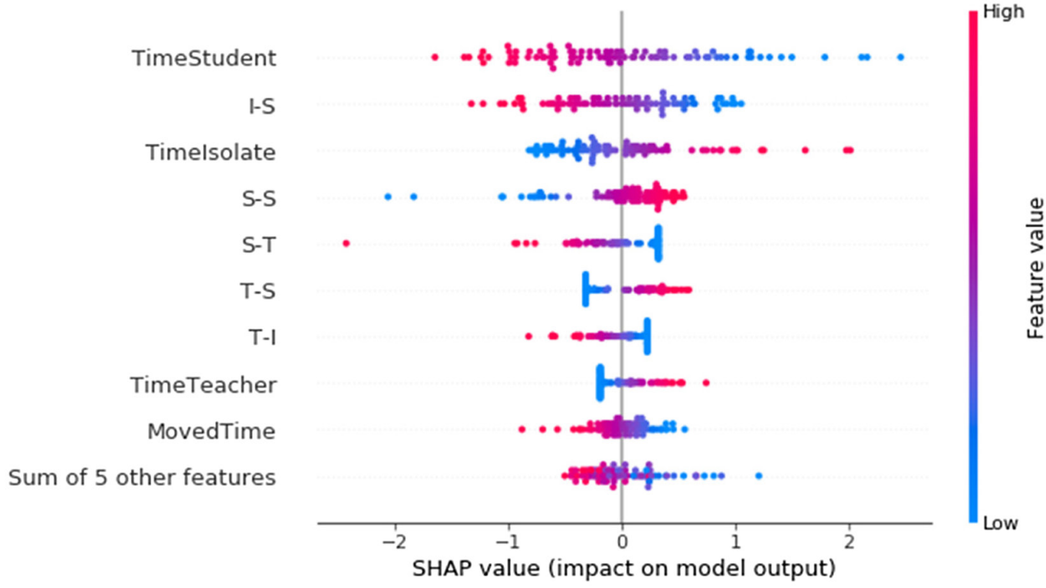

FIGURE 4 SHAP features importance for Reading

TABLE 5 Model performance (Maths) with selected features: *TimeIsolate*, *TimeStudent*, *S-I*, *MovedTime*, *I-I*, *T-I*, and *T-S*

	Accuracy	Precision	Recall	Cohen's *k*	AUC
LR	0.81 (0.06)	0.71 (0.06)	0.75 (0.06)	0.57 (0.06)	0.79 (0.06)
SVM	0.78 (0.06)	0.66 (0.06)	0.73 (0.06)	0.52 (0.06)	0.77 (0.06)
RF	0.78 (0.06)	0.68 (0.06)	0.67 (0.06)	0.50 (0.06)	0.75 (0.06)
KNN	0.76 (0.06)	0.64 (0.06)	0.68 (0.06)	0.47 (0.06)	0.74 (0.06)
ANN	0.74 (0.06)	0.61 (0.06)	0.66 (0.06)	0.43 (0.06)	0.72 (0.06)

TABLE 6 Model performance (Reading) with selected feature: *TimeIsolate*, *TimeStudent*, *I-S*, *S-S*, *S-T*, and *T-S*

	Accuracy	Precision	Recall	Cohen's *k*	AUC
LR	0.74 (0.08)	0.52 (0.08)	0.72 (0.08)	0.41 (0.08)	0.73 (0.08)
SVM	0.73 (0.07)	0.51 (0.07)	0.72 (0.07)	0.4 (0.07)	0.73 (0.07)
RF	0.73 (0.07)	0.53 (0.07)	0.49 (0.07)	0.31 (0.07)	0.66 (0.07)
KNN	0.73 (0.07)	0.52 (0.07)	0.70 (0.07)	0.40 (0.07)	0.72 (0.07)
ANN	0.75 (0.08)	0.55 (0.08)	0.58 (0.08)	0.38 (0.08)	0.69 (0.08)

DISCUSSION

As the complexity of physical learning increases, providing teachers with educational technologies that can assist them in teaching reflection and student supervision become essential. This paper illustrated the potential of using physical positioning and proximity traces and ML classifiers to identify potential low-progress students from their socio-spatial behaviours prospectively.

Main findings

In response to RQ1, the correlation results (Table 2) were consistent with prior studies and contributed new empirical evidence to support student progression's positive relationships with social interaction (Gasevic et al., 2013; Wentzel, 2017), physical activity (Donnelly et al., 2016), and sustained engagement (Guthrie et al., 2012). These findings can also provide initial evidence to support the use of wearables to capture progression-related socio-spatial behaviour features. These differences in correlations are likely to reflect differences in instructional designs. For example, the negative correlation between *TimeIsolate* and student progression in Reading but not Maths may imply that *TimeIsolate* is a feature that successfully captures instructional designs that promote collaborative interactions.

Regarding RQ2, ML classifiers proved more accurate than randomly guessing (accuracy of dummy model would be 50%) for both Maths and Reading. Yet, the overall performance is lower in Reading than Maths. The poor precision and low inter-rater reliability would impair the practical value of the current work in helping teachers to identify low-progress students. In particular, classifiers with a high false positive rate would still require teachers to make significant efforts to identify the actual low-progress students.

The potential reason behind this poor performance was revealed in the investigation on feature importance, through RQ3. Although both subjects shared some important features (eg, *TimeStudent* and *TimeIsolate*), the low contributing features in Reading could be responsible for the poor performance. These features (eg, *MovedTime* and *I-I*) were less relevant in Reading, a subject that uses collaborative learning more often, as part of its pedagogical design in this school. Whereas in Maths, students had more autonomy in choosing their desired learning format. Consequently, their learning progress could be reflected by more varieties of socio-spatial features, resulting in better model performance. Indeed, after the feature selection, the model performances for both Maths and Reading have increased.

Thus, the findings from RQ2&3 would oppose generalised classification algorithms in physical learning contexts. This finding also emphasises the need to account for the pedagogical differences between different subjects and select appropriate features when constructing prediction models, which resonates with Gasevic et al.'s (2016) call for course-specific models.

Implications

This work has several implications for future research and teaching practice. Firstly, the current findings on the significant associations between students' socio-spatial behaviours and their academic performance further stressed the valuable insights contained within students' in-class behaviours. Mining and utilising these insights could be a promising direction for future learning analytics research that aims to understand how socio-spatial behaviours contribute to academic progression and differentiating influential learning behaviours when multiple learning designs are deployed.

However, although wearables and positioning data, alone, are insufficient to provide a highly accurate prediction, they do contribute to the growing development of multimodal learning analytics (Dich et al., 2018; Worsley & Blikstein, 2018). Combining wearables and positioning data with other educational technologies, such as computer vision, natural language processing, and physiological sensors, would further strengthen the power of predictive learning analytics in triangulating students' actual learning progress. Additionally, developing and improving these predictive models can also contribute to complement the existing explanatory models on student learning (Chng et al., 2020; Martinez-Maldonado, Echeverria, et al., 2020; Yan et al., 2021), and further enhance the actionable insights that can be generated from students' behavioural traces. As the cost of these sensing technologies declines, large-scale implementation could become possible. By then, the automated, non-intrusive, and scalable nature of these sensing technologies could provide teachers with insights to support and scaffold student learning as well as to reflect on their teaching practices and the impact of their learning designs across various disciplines.

Ethical considerations

It is vital to consider the ethical issues surrounding the current work and, in general, predictive learning analytics. Labelling students based on data can lead to further marginalising students who may already be facing social or learning challenges (Osterholm et al., 2011). It can be argued that identifying students who may not socialise much at school or that could be flagged by an artificial intelligence algorithm as being "at-risk" may allow teachers to help these potentially low-progress students. Yet, this can also create grounds for discrimination and unconscious biases. In fact, psychology research has widely and convincingly demonstrated the potential dangers of discrimination resulting from labelling students, such as reduced self-esteem in students and expectations in teachers (Higgins et al., 2002). Thus, the purpose of predictive learning analytics and how it will be communicated to teachers require careful considerations and further work.

As shown in the current work, labelling students should not be universal, instead, it needs to be subject-specific with the aim of assisting learning supervision and helping teachers to provide timely support. For example, instead of planning an intervention, the teacher could reflect on better strategies to improve their learning designs to promote inclusion or social interactions. Meanwhile, actionable insights from explanatory models should also be made available while communicating results from machine learning algorithms to teachers, so they can understand why a student may be flagged as featuring "low-progress". With these additional insights, labelling could initiate the process for identifying and helping low-progress students instead of remaining as the end product of predictive algorithms. Most importantly, educating teachers about implicit biases and ensuring that they adopt inclusive pedagogy principles is essential to minimise labelling's adverse effects (Florian & Black-Hawkins, 2011).

Many concerns have also been expressed regarding privacy issues related to sensing technologies (Cukurova et al., 2020; Martinez-Maldonado et al., 2018). Given the dilemma between the escalated risk of unintended surveillance in capturing data from multiple channels, and the potential danger of making decisions based on insufficient data, further research is required to understand the unintended consequences of this trade-off more deeply and develop regulations for stakeholders who will be in direct contact with these data and analytics. Co-designing learning analytics solutions with schools and stakeholders would open the space to discuss potential unintended surveillance and design the mechanisms to use these technologies with integrity (Martinez-Maldonado, Mangaroska, et al., 2020).

Additionally, future research and potential practical adoption of sensors in physical classrooms should also comply with the regulatory requirements proposed by relevant government authorities (eg, the European Union General Data Protection Regulation). This compliance should involve but not be limited to ensuring data confidentiality, the autonomy of stakeholders, and adopting a transparent data handling procedure (Hoofnagle et al., 2019). Encouragingly, parents were very supportive of the current research as obtaining parental approval was highly efficient and successful. This phenomenon could result from the data security procedures that were employed to ensure data confidentiality (students' names were masked) and strictly limited data usage for research purposes. This supportive attitude resonates with findings in the UK, where adults showed favourable views about using artificial intelligence to tailor teaching and learning to individuals (Ipsos/MORI, 2017). A recent study also found that around 71% of Australian adults support the use of artificial intelligence to address educational challenges (Selwyn et al., 2020). These findings encourage academicians, practitioners, and researchers to further investigate the potentials of practically adopting educational technologies to support learning and teaching practices.

Limitations and future works

The current findings have two major limitations. First, wearable trackers' illustrated potentials in identifying low-progress students are limited to open-ended learning spaces and flexible classrooms. The lack of spatial movement in lecture-style spaces or classrooms with fixed seating would reduce the practicality of positioning data. In these learning contexts, exploring the predictive potentials of facial, gesture, and physiological data could be more promising (Ahuja et al., 2019; Raca & Dillenbourg, 2013; Sharma & Giannakos, 2020). Second, these granular data are at most an estimation of students' actual behaviours. Future studies should validate the accuracy of their granular data in capturing the targeted learning behaviour. Additionally, future studies could map from granular data to learning behaviours and then to individuals' cognitive process with multimodal data (Dindar et al., 2020). This additional step could provide further insights for teachers to understand the potential reasons behind students' low progression and encourage cross-disciplinary research. Future works may also adopt a multimodal approach, such as multiple sensor data (eg, video, audio, proximity, and biometrics), to triangulate the type of interaction that occurred as proximity data, alone, does not contain this information (Martinez-Maldonado, Echeverria, et al., 2020; Yan et al., 2021). Moreover, developing dashboards and making these socio-spatial analytics available in real-time to support teachers and students is also a promising direction. Empirical evidence to support the use of these technological tools is emerging from research into teachers' (Martinez-Maldonado, Mangaroska, et al., 2020) and students' perceptions (Mangaroska et al., 2021) of sensing technologies and learning analytics. Future works may also explore the predictivity of socio-spatial behaviours in other pedagogical factors, such as in-class engagement and role distribution during problem-based learning. Lastly, capturing additional demographic information, such as the students' socioeconomic status, may contribute to comprehending the predictive results with more contextual insights. However, a delicate balancing act is needed between ethics and increased data disclosure.

CONCLUDING REMARKS

As physical learning spaces become increasingly complex, students' motor and physiological traces could contain the key to generating meaningful learning analytics. These insights are necessary for assisting teachers in dealing with increased complexity and fulfilling their responsibility to support students equally. The current work illustrates the potential of using wearable trackers to identify low-progress students from their socio-spatial behaviours. Our work contributes to the growing body of classroom learning analytics research and the need for these analytics to have instructional sensitivity. In future research, utilising multimodal data could further strengthen the predictive and explanatory power of classroom learning analytics.

ACKNOWLEDGEMENTS

Open access publishing facilitated by Monash University, as part of the Wiley - Monash University agreement via the Council of Australian University Librarians.

CONFLICT OF INTEREST

There is no potential conflict of interest in this study.

ETHICS STATEMENT

Ethics and parental approval were obtained prior to the study.

DATA AVAILABILITY STATEMENT

Any request for data access should contact the corresponding author.

ORCID

Lixiang Yan https://orcid.org/0000-0003-3818-045X

REFERENCES

Ahuja, K., Kim, D., Xhakaj, F., Varga, V., Xie, A., Zhang, S., Townsend, J. E., Harrison, C., Ogan, A., & Agarwal, Y. (2019). Edusense: Practical classroom sensing at scale. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *3*(3), 1–26. https://doi.org/10.1145/3351229

Back, M. D., Schmukle, S. C., & Egloff, B. (2008). Becoming friends by chance. *Psychological Science*, *19*(5), 439. https://doi.org/10.1111/j.1467-9280.2008.02106.x

Brennan, A., Peace, C., & Munguia, P. (2018). Classroom size, activity and attendance: Scaling up drivers of learning space occupation. In *Proceedings of the 8th international conference on learning analytics and knowledge* (pp. 255–259).

Brinton, C. G., & Chiang, M. (2015). Mooc performance prediction via clickstream data and social learning networks. In 2015 *IEEE conference on computer communications (INFOCOM)* (pp. 2299–2307).

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Chin, H. B., Mei, C. C. Y., & Taib, F. (2017). Instructional proxemics and its impact on classroom teaching and learning. *International Journal of Modern Languages and Applied Linguistics*, *1*(1), 69–85.

Chng, E., Seyam, M. R., Yao, W., & Schneider, B. (2020). Using motion sensors to understand collaborative interactions in digital fabrication labs. In *International conference on artificial intelligence in education* (pp. 118–128).

Chua, Y. H. V., Dauwels, J., & Tan, S. C. (2019). Technologies for automated analysis of co-located, real-life, physical learning spaces: Where are we now? In *Proceedings of the 9th international conference on learning analytics & knowledge* (pp. 11–20).

Cukurova, M., Giannakos, M., & Martinez-Maldonado, R. (2020). The promise and challenges of multimodal learning analytics. *British Journal of Educational Technology*, *51*(5), 1441–1449. https://doi.org/10.1111/bjet.13015

Dich, Y., Reilly, J., & Schneider, B. (2018). Using physiological synchrony as an indicator of collaboration quality, task performance and learning. In *International conference on artificial intelligence in education* (pp. 98–110).

Dindar, M., Järvelä, S., & Haataja, E. (2020). What does physiological synchrony reveal about metacognitive experiences and group performance? *British Journal of Educational Technology*, *51*(5), 1577–1596. https://doi.org/10.1111/bjet.12981

Donnelly, J. E., Hillman, C. H., Castelli, D., Etnier, J. L., Lee, S., Tomporowski, P., & Szabo-Reed, A. N. (2016). Physical activity, fitness, cognitive function, and academic achievement in children: A systematic review. *Medicine and Science in Sports and Exercise*, *48*(6), 1197.

Entezari-Maleki, R., Rezaei, A., & Minaei-Bidgoli, B. (2009). Comparison of classification methods based on the type of attributes and sample size. *Journal of Convergence Information Technology*, *4*(3), 94–102.

Farrow, E., Moore, J., & Gasevic, D. (2019). Analysing discussion forum data: A replication study avoiding data contamination. In *Proceedings of the 9th international conference on learning analytics & knowledge* (pp. 170–179).

Fernandez Nieto, G., Mrtinez-Maldonado, R., Echeverria, V., Kitto, K., & Buckingham Shum, S. (2021). What can analytics for teamwork proxemics reveal about positioning dynamics in clinical simulations? *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1–24.

Florian, L., & Black-Hawkins, K. (2011). Exploring inclusive pedagogy. *British Educational Research Journal*, *37*(5), 813–828. https://doi.org/10.1080/01411926.2010.501096
Gasevic, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, *28*, 68–84. https://doi.org/10.1016/j.iheduc.2015.10.002
Gasevic, D., Zouaq, A., & Janzen, R. (2013). “Choose your classmates, your GPA is at stake!” The association of cross-class social ties and academic performance. *American Behavioral Scientist*, *57*(10), 1460–1479. https://doi.org/10.1177/0002764213479362
Gløersen, Ø., & Federolf, P. (2016). Predicting missing marker trajectories in human motion data using marker intercorrelations. *PLoS One*, *11*(3), e0152616. https://doi.org/10.1371/journal.pone.0152616
Goodyear, P. (2020). Design and co-configuration for hybrid learning: Theorising the practices of learning space design. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.12925
Gray, C. C., & Perkins, D. (2019). Utilizing early engagement and machine learning to predict student outcomes. *Computers & Education*, *131*, 22–32. http://www.sciencedirect.com/science/article/pii/S0360131518303191
Greenberg, S., Boring, S., Vermeulen, J., & Dostal, J. (2014). Dark patterns in proxemic interactions: A critical perspective. In *Proceedings of the 2014 conference on designing interactive systems* (pp. 523–532).
Guthrie, J. T., Wigfield, A., & You, W. (2012). Instructional contexts for engagement and achievement in reading. In *Handbook of research on student engagement* (pp. 601–634). Springer.
Hall, E. T. (1966). *The hidden dimension*, Vol. *609*. Doubleday.
Hellas, A., Ihantola, P., Petersen, A., Ajanovski, V. V., Gutica, M., Hynninen, T., & Liao, S. N. (2018). Predicting academic performance: A systematic literature review. In Proceedings companion of the 23rd annual ACM conference on innovation and technology in computer science education (pp. 175–199).
Higgins, E. L., Raskind, M. H., Goldberg, R. J., & Herman, K. L. (2002). Stages of acceptance of a learning disability: The impact of labeling. *Learning Disability Quarterly*, *25*(1), 3–18. https://doi.org/10.2307/1511187
Hoofnagle, C. J., van der Sloot, B., & Borgesius, F. Z. (2019). The European Union general data protection regulation: What it is and what it means. *Information & Communications Technology Law*, *28*(1), 65–98. https://doi.org/10.1080/13600834.2019.1573501
Ipsos/MORI. (2017). *Public views of machine learning*. Royal Society.
Joksimovic, S., Gasevic, D., Kovanovic, V., Riecke, B. E., & Hatala, M. (2015). Social presence in online discussions as a process predictor of academic performance. *Journal of Computer Assisted Learning*, *31*(6), 638–654. https://doi.org/10.1111/jcal.12107
Joseph, R. (n.d.). *Grid search for model tuning*. Retrieved January 8, 2021, from https://towardsdatascience.com/grid-search-for-model-tuning3319b259367e
Klem, A. M., & Connell, J. P. (2004). Relationships matter: Linking teacher support to student engagement and achievement. *Journal of School Health*, *74*, 262–273. https://doi.org/10.1111/j.1746-1561.2004.tb08283.x
Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In I. Guyon, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds), *Advances in neural information processing systems* Vol. *30*, (pp. 4765–4774). Curran Associates Inc. https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf
Mangaroska, K., Martinez-Maldonado, R., Vesin, B., & Gašević, D. (2021). Challenges and opportunities of multimodal data in human learning: The computer science students' perspective. *Journal of Computer Assisted Learning*, *37*, 1030–1047. https://doi.org/10.1111/jcal.12542
Martinez-Maldonado, R., Echeverria, V., Santos, O. C., Santos, A. D. P. D., & Yacef, K. (2018). Physical learning analytics: A multimodal perspective. In *Proceedings of the 8th international conference on learning analytics and knowledge* (pp. 375–379).
Martinez-Maldonado, R., Echeverria, V., Schulte, J., Shibani, A., Mangaroska, K., & Shum, S. B. (2020). Moodoo: indoor positioning analytics for characterising classroom teaching. In *AIED* (pp. 360–373).
Martinez-Maldonado, R., Mangaroska, K., Schulte, J., Elliott, D., Axisa, C., & Shum, S. B. (2020). Teacher tracking with integrity: What indoor positioning can reveal about instructional proxemics. *Proceedings of the ACM on IMWUT*, *4*(1), 1–27. https://doi.org/10.1145/3381017
Martinez-Maldonado, R., Schulte, J., Echeverria, V., Gopalan, Y., & Shum, S. B. (2020). Where is the teacher? Digital analytics for classroom proxemics. *Journal of Computer Assisted Learning*, *36*(5), 741–762. https://doi.org/10.1111/jcal.12444
Mondada, L. (2013). Interactional space and the study of embodied talk-in-interaction. In P. Auer, M. Hilpert, A. Stukenbrock, & B. Szmrecsanyi (Eds.), *Space in language and linguistics* (pp. 247–275). De Gruyter.
Nam, S., & Samson, P. (2019). Integrating students’ behavioral signals and academic profiles in early warning system. In *Artificial intelligence in education* (pp. 345–357). Springer.
Nguyen, Q., Poquet, O., Brooks, C., & Li, W. (2020). Exploring homophily in demographics and academic performance using spatial-temporal student networks. In *Proceedings of the 13th international conference on educational data mining (edm 2020)* (pp. 194–201).

Osterholm, K., Nash, W. R., & Kritsonis, W. A. (2011). Effects of labeling students “learning disabled”: Emergent themes in the research literature 1970 through 2000. *FOCUS on Colleges, Universities & Schools*, *6*(1), 3–9.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., & Vanderplas, J. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, *12*, 2825–2830.
Raca, M., & Dillenbourg, P. (2013). System for assessing classroom attention. In *Proceedings of the third international conference on learning analytics and knowledge* (pp. 265–269).
Reh, S., Rabenstein, K., & Fritzsche, B. (2011). Learning spaces without boundaries? Territories, power and how schools regulate learning. *Social & Cultural Geography*, *12*(01), 83–98. https://doi.org/10.1080/14649365.2011.542482
Ryan, A. M., & Shin, H. (2011). Help-seeking tendencies during early adolescence: An examination of motivational correlates and consequences for achievement. *Learning and Instruction*, *21*(2), 247–256. https://doi.org/10.1016/j.learninstruc.2010.07.003
Saquib, N., Bose, A., George, D., & Kamvar, S. (2018). Sensei: Sensing educational interaction. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *1*(4), 1–27. https://doi.org/10.1145/3161172
Selwyn, N., Gallo Cordoba, B., Andrejevic, M., & Campbell, L. (2020). *AI for social good—Australian attitudes toward AI and society report.pdf* (Version 1). Monash University. https://doi.org/10.26180/13159781.v1
Sharma, K., & Giannakos, M. (2020). Multimodal data capabilities for learning: What can multimodal data tell us about learning? *British Journal of Educational Technology*, *51*(5), 1450–1484. https://doi.org/10.1111/bjet.12993
Strumbelj, E., & Kononenko, I. (2014). Explaining prediction models and individual predictions with feature contributions. *Knowledge and Information Systems*, *41*(3), 647–665.
Suryavanshi, N. B., Reddy, K. V., & Chandrika, V. R. (2019, February). Direction finding capability in bluetooth 5.1 standard. In *International Conference on Ubiquitous Communications and Network Computing* (pp. 53–65). Springer.
Tomkins, S., Ramesh, A., & Getoor, L. (2016). Predicting post-test performance from online student behavior: A high school MOOC case study. In *International conference on educational data mining* (pp. 239–246). .
Wentzel, K. R. (2017). Peer relationships, motivation, and academic performance at school. In A. J. Elliot, C. S. Dweck, & D. S. Yeager (Eds.), *Handbook of competence and motivation: Theory and application* (pp. 586–603). The Guilford Press.
Worsley, M., & Blikstein, P. (2018). A multimodal analysis of making. *International Journal of Artificial Intelligence in Education*, *28*(3), 385–419. https://doi.org/10.1007/s40593-017-0160-1
Yan, L., Martinez-Maldonado, R., Cordoba, B. G., Deppeler, J., Corrigan, D., Nieto, G. F., & Gasevic, D. (2021). Footprints at school: Modelling in-class social dynamics from students' physical positioning traces. In *LAK21: 11th International learning analytics and knowledge conference* (pp. 43–54). Association for Computing Machinery. https://doi.org/10.1145/3448139.3448144
Zacharis, N. Z. (2015). A multivariate approach to predicting student outcomes in web-enabled blended learning courses. *The Internet and Higher Education*, *27*, 44–53. http://www.sciencedirect.com/science/article/pii/S1096751615000391

How to cite this article: Yan, L., Martinez-Maldonado, R., Gallo Cordoba, B., Deppeler, J., Corrigan, D., & Gašević, D. (2022). Mapping from proximity traces to socio-spatial behaviours and student progression at the school. *British Journal of Educational Technology*, 53, 1645–1664. https://doi.org/10.1111/bjet.13203

APPENDIX

Hyperparameters for each Classifier

	Hyperparameters	
Classifiers	Grid search parameters	Model parameters
LR	Penalty = [l1, l2, elasticnet, none] Regularization strength (inverse) = [0.001, 0.009, 0.01, 0.09, 1, 5, 10, 25] Solver = liblinear, lbfgs Random state = 0	Penalty = [l2] Regularization strength (inverse) = [1] Solver = liblinear Random state = 0

Classifiers	Hyperparameters	
	Grid search parameters	Model parameters
SVM	Kernel = linear, poly, rbf, sigmoid Regularization strength (inverse) = [0.001, 0.009, 0.01, 0.09, 1, 5, 10, 25]	Kernel = linear Regularization strength (inverse) = [1]
RF	Criterion = [gini, entropy] Max_features = [auto, sqrt, log2] Number of estimators = [100, 200, 300, 1000]	Criterion = [gini] Max_features = [auto] Number of estimators = [100]
KNN	Weights = [uniform, distance] Metrics = [euclidean, manhattan] Number of neighbors = [3, 5, 11, 19]	Weights = [uniform] Metrics = [euclidean] Number of neighbors = [5]
ANN	Solver = [lbfgs, sgd, adam] Activation = [identity, logistic, tanh, relu] Alpha = [1e-5, 1e-4, 1e-3, 0.01, 0.1] Learning_rate = [constant, invscaling, adaptive] Max_iter = 5000 Hidden layer sizes = [(100), (5,2)] Random state = 1	Solver = [lbfgs] Activation = [relu] Alpha = [1e-5] Learning_rate = [constant] Max_iter = 5000 Hidden layer sizes = [(5,2)] Random state = 1