Agent-Based Simulation, Machine Learning, and Gamification: An Integrated Framework for Addressing Disruptive Behaviour and Enhancing Student and Teacher Performance in Educational Settings

KHULOOD OBAID ALHARBI

Agent-Based Simulation, Machine Learning, and Gamification: An Integrated Framework for Addressing Disruptive Behaviour and Enhancing Student and Teacher Performance in Educational Settings

Khulood Obaid Alharbi

Supervisors:

Professor Alexandra I. Cristea

Professor Sue Black

A Thesis Presented for the Degree Of

Doctor of Philosophy in Computer Science

School of Engineering and Computing Sciences

Department of Computer Science

Durham University

DH1 3LE

United Kingdom

2025

Abstract

The classroom environment is a major contributor to the learning process in schools. Young students are affected by different factors in their academic progress, be it their own characteristics, their teacher's, or their peers'. Disruptive behaviour, in particular, is one of the main factors that create challenges in the classroom environment, by hindering learning and effective classroom management. To overcome these challenges, it is important to understand what causes disruptive behaviour, and how to predict and prevent it. While Machine Learning (ML) is already used in education to predict disruption-related outcomes, there is less focus on understanding the processes leading to the effect of disruptive behaviour on learning. Thus, in this thesis, I propose using Agent-Based Modelling (ABM) for the simulation of disruptive behaviour in the classroom, to provide teachers with a tool that helps them not only predict, but also understand how classroom interactions lead to disruptions. Reducing negative factors in the learning environment, like disruptive behaviour, is further supported by increasing positive factors, such as motivation and engagement. Therefore, the use of gamification is then introduced as a strategy to promote motivation and improve engagement by making not only the learning environment more rewarding, but also the ABM teacher simulation more appealing.

This thesis focuses on these issues by designing and implementing for the first time an integrated approach that combines ABM and ML with gamification to simulate classroom interactions and predict disruptive behaviour. The ABM models the complex interactions between students, teachers, and peers, providing a means to study the processes leading to behavioural issues. Meanwhile, ML algorithms help

predict learning outcomes with behaviours such as inattentiveness, hyperactivity, and impulsiveness.

The simulation has revealed insights, such as the impact of peer influence on student behaviour and the varying effects of different types of disruptive behaviour, such as inattentiveness, hyperactivity and impulsiveness, on academic performance. The improved performance of the hybrid ML-ABM is shown by measuring results of simulation with ML integration using metrics like MAE, RMSE and Pearson correlation. Moreover, the inclusion of gamification elements was shown to improve engagement by increased login frequency and course completion rates in a MOOC setting, as well as be effective and appealing for teachers using the ML-ABM.

In conclusion, this thesis presents the first comprehensive model that integrates ABM, ML, and gamification elements to explore educational outcomes in a disruptive classroom; it develops the first hybrid ML-ABM approach for predicting and managing classroom disruptive behaviour; it provides empirical evidence of the effectiveness of gamification in boosting student and teacher engagement; and it offers practical insights for educators and policymakers seeking to adopt innovative, technology-driven strategies for improving teaching and learning. The research lays a foundation for future studies, aiming to further explore and expand the capabilities of these technologies in an educational context.

Declaration

The work in this thesis is based on research carried out within the Artificial Intelligence and Human Systems (AIHS) Group at the Department of Computer Science, Durham University, the United Kingdom. No part of this thesis has been submitted elsewhere for any other degree or qualification, and it is all the author's work, unless referenced to the contrary.

Some of the work in this thesis has been published or currently under submission to be published as indicated below:

- Alharbi K., Cristea A.I., Shi L., Tymms P., Brown C. (2021) Agent-Based Classroom Environment Simulation: The Effect of Disruptive Schoolchildren's Behaviour Versus Teacher Control over Neighbours. In: Roll I., McNamara D., Sosnovsky S., Luckin R., Dimitrova V. (eds) Artificial Intelligence in Education. AIED 2021. Lecture Notes in Computer Science, vol 12749. Springer, Cham. *Chapter 4.*
- Alharbi K., Cristea A.I., Shi L., Tymms P., Brown C. (2021) Agent-Based Simulation of the Classroom Environment to Gauge the Effect of Inattentive or Disruptive Students. In: Cristea A.I., Troussas C. (eds) Intelligent Tutoring Systems. ITS 2021. Lecture Notes in Computer Science, vol 12677. Springer, Cham. *Chapter 4.*
- Alharbi, K., Alrajhi, L., Cristea, A. I., Bittencourt, I. I., Isotani, S., & James, A. (2020, June). Data-Driven Analysis of Engagement in Gamified Learning Environments: A Methodology for Real-Time Measurement of MOOCs. In International Conference on Intelligent Tutoring Systems (pp. 142-151). Springer, Cham. *Chapter 7.*
- Alharbi K., Cristea A.I. (2024) Hybrid Agent-Based Machine Learning Simulation of a Classroom Disruption Model. In International Conference on Modelling and Simulation. ECMS 2024. *Chapter 6.*
- Alharbi K., Cristea A.I. (2024) Disruptiveness-related Features as an Interpretable Predictor for Student Performance. Submitted to the Journal of Experimental Education. *Chapter 5.*

- Alharbi K., Cristea A.I. (2025) A Data-Driven Approach to Teacher Engagement: Gamification in a Hybrid ML-ABM Framework. Submitted to the International Conference on Artificial Intelligence in Education. *Chapter 8.*

Acknowledgements

Firstly, I express my gratitude to Allah for granting me the strength, wellness, and dedication to accomplish this thesis.

The completion of my doctoral thesis evokes profound emotions, as words fall short in capturing the journey's challenges and blessings.

I am indebted to my unwavering support system, my beloved parents, Modhi and Obaid, who have been by my side since my earliest steps in life and along this academic path. Their faith and prayers have provided me with strength in times of need.

I also extend my thanks to my siblings for their warm encouragement, particularly to my support team: Hanin Alharbi, Fadi Alahrbi, Ohuod Alharbi, and Tahani AlKhuliwi.

My deepest appreciation goes to my cherished family: my husband Mohammed Almansor, who has stood by me through every moment, and my son Abdulaziz Almansor, whose love and joy have filled my days and above all for making me a mother with all the determination in the world to face any difficulties along the way.

Additionally, I wish to acknowledge the invaluable advice and assistance of my research group members, especially my co-author Laila AL Rajhi.

Special thanks are also due to my supervisor, Alexandra Cristea, for providing the opportunity to embark on this PhD journey and for her cheerful support and precise guidance throughout its completion.

Lastly, I extend my gratitude to Qassim University and the Saudi Cultural Bureau in the UK for sponsoring this PhD program and the department of computer science in Durham University and its members for facilitating everything needed to complete this research.

Table of Contents

List of Figures

List of Tables

Dedication

This thesis is dedicated to my dearest brother Hani.
His name means 'blessed' and of him I was.
May he rest in peace.

CHAPTER 1

1 Introduction

1.1. Disruptive Behaviour in Educational Settings

Teachers are responsible for creating an environment conducive to learning by designing curriculum instructions that enhance the achievement of learning outcomes. Besides giving suitable and effective curriculum instructions, teachers are responsible for controlling the learning processes to ensure that learners of different capabilities can achieve their zone of proximal development without affecting their peers' learning process. Their efforts are mostly constrained by learners' disruptive behaviours [127]. Mupa & Chinooneka [143] argued that teachers who conduct learning monotonously or without appropriate or engaging course materials result in ineffective teaching, and eventually boredom among learners. Bored learners are more likely to talk, seek the attention of their neighbours, and generally engage in activities likely to discursive other learners and the teacher.

In a classroom, learners' disruptive behaviours are generally identified as inappropriate behaviour that challenges smooth teaching and learning. Disruptive behaviour affects teachers' class management efforts and instructional delivery, thus, many students hardly meet a lesson's expected learning outcome. Therefore, learners' disruptive behaviour could be described as behaviour exhibited by learners that school guidelines and teachers consider unethical, unacceptable, and capable of harming the academic activities of the disruptive students, their peers, and the school or constraining expected academic achievement [73]. Likewise, disruptive behaviour could be understood as a situation where students' patterns of actions oppose the learning institutions' rules, regulations, or conduct. In this case, disruptive behaviours range from breaking the school's stipulated and defined moral and ethical laws or resorting to behaviours that endanger self and/or others and affect effective curriculum instructional delivery.

Within a classroom setup, learners' disruptive behaviours manifest as learners' uncooperativeness and preventing themselves and others from completing given learning exercises or assignments. Besides, learners might ignore or disrespect classroom rules such as raising arms when seeking permission to inquire or contribute to class discussion, walking around the classroom during a lesson, fighting or whispering to peers, fidgeting, speaking out of tune, or moving furniture around during class [59]. A teacher is expected to exhibit certain characteristics and abilities in class, including maintaining learner discipline, besides teaching and guiding students to achieve expected learning outcomes.

1.2. Peers and Disruptive Behaviours

Peers play a pivotal role in the social and emotional development of children and adolescents, fostering essential social bonds through early childhood interactions such as play and make-believe activities [210]. However, while peer groups are important for developmental growth, they can also affect individual behaviour. The influence of peers begins in the formative years of childhood and intensifies during adolescence, a period with an estimated one to six students displaying disruptive behaviour [31]. Peer influence can manifest in various forms of disruptive behaviour that can damage the learning environment. Such behaviours may include encouraging each other to engage in truancy, substance abuse, or minor acts of theft. Subtler forms of influence can lead to talking out of turn in class, spreading rumours, or engaging in cyberbullying by spreading inappropriate content online about other students. Also, pupils believe that going against school norms increase popularity among peers; thus, the need for peer approval may lead them to disruptive behaviour [30]. The worry about their peers' perception of their actions or inaction drives them into inappropriate behaviour [174], which eventually disrupts learning.

1.3. Agent-Based Model (ABM) for Classroom Management

Classroom simulations are increasingly used for classroom management training by allowing educators to model and practice pedagogical strategies in a controlled, low-risk environment. A simulation environment can offer insights into student-teacher interactions and enable teachers to apply management techniques, especially with disruptive behaviour without the immediate pressures of a real classroom [98].

Simulated experiences are considered by teachers to be important training tools because they replicate the complexities of real-life teaching. Enabling teachers to recreate classroom scenarios allows them to transfer the skills and insights gained from their training into their future teaching experience. They effectively answer the question, "What will happen if I take this specific action?" by offering a space to experiment and observe the consequences in a realistic yet controlled setting [148]. In addition, research shows several practical strategies that can be modelled in classroom simulations. Such as setting clear expectations at the start of the term, which has been proven to significantly reduce classroom disruptions [153]. Moreover, simulations can include games like the Good Behaviour Game, which rewards positive behaviour and cooperation among students, promoting an overall positive classroom dynamic [153].

Promoting and developing classroom simulations provides teachers with a safe space to practice effective techniques for managing classrooms, ultimately leading to calmer and more productive learning environments.

1.4. Using Gamification to Mitigate Disruptive Behaviours

The increasing popularity and adoption of educational technology in teaching and learning offer teacher opportunities to enhance learning through gamification [183]. Besides improving learning by leveraging the practical representation of concepts, teachers use gamification mechanics and elements, including points, rankings, progressions, rewards, challenges, immediate feedback, rules, and time to control interactions and mitigate disruptive behaviours [171]. These elements promote players' engagement and motivation in a game setup, making them ideal for promoting student engagement and motivation. Practically, gamification offers a suitable environment for motivating players. Players see their progress after successfully completing game levels and are eligible for rewards, keeping them interested and motivated to complete the next levels. In a learning environment, gamification provides teachers and students with a more interactive, engaging, and effective learning experience.

Teachers can devise gamification strategies to focus on individual disruptive students or groups [183]. Based on game dynamics, the teacher creates a scenario where a disruptive student works alone and leverages his or her effort to accomplish assigned tasks according to given instructions and the expected outcome of the activity. This approach helps disruptive students be conscious of the learning process and expected learning outcomes. On the other hand, teachers can devise gamification strategies to support collaborative learning. By assigning collaborative work to disruptive learners, learners work in a team, each pulling their weight because the outcome affects each team member. The game mechanism helps improve interaction and engagement besides helping in promoting an individual and collaborative responsibility.

Students often express higher levels of task orientation when teachers use gamified learning than when they are subjected to a traditional learning approach [183]. As a result, they express less disruptive behaviour because they are oriented towards intrinsic motivation, for instance, the feeling of being part of the process, unlike in extrinsic motivation where learners would expect an external reward or they tend to disengage in the process to avoid failure [183].

Alternatively, teachers themselves need support when using digital systems, and gamification is a promising means to provide this, as will be elaborated in this thesis.

1.5. Research Goals

The growing challenge of managing disruptive behaviour in educational settings affects learning outcomes and poses difficulties for educators in maintaining an effective classroom environment. Traditional disciplinary methods often fail to address the nuanced nature of disruptive behaviour, prompting the need for more innovative and comprehensive solutions. Thus, the aim of this research is to explore how a combination of advanced technological tools—Agent-Based Modelling (ABM), Machine Learning (ML), and gamification—can effectively address and mitigate disruptive classroom behaviour while simultaneously enhancing student engagement and teacher performance.

Agent-Based Modelling (ABM) allows for the simulation of complex classroom interactions, offering insights into individual and group behaviours in educational settings [67]. By modelling these interactions, ABM enables educators to test various intervention strategies in a controlled virtual environment. However, ABM alone may lack the predictive capabilities needed for accurate results. Therefore, Machine Learning (ML) has been integrated into the model to provide predictive analytics, which can forecast disruptive behaviour based on key variables such as student attention levels, peer dynamics, and socio-economic factors [74].

In addition to the technical capabilities of ABM and ML, this thesis incorporates gamification elements to improve engagement for both students and teachers. Gamification, which includes features like badges, leaderboards, and rewards, has been shown to enhance motivation and participation in educational settings [50]. By introducing these elements into the simulated classroom environment, this research aims to create a more interactive and engaging space, ultimately improving behavioural management and learning outcomes.

In conclusion, this study addresses the need for more effective, technologically enhanced approaches to manage disruptive behaviour in classrooms. By developing a hybrid ABM-ML approach enhanced with gamification, this research aims to provide a robust, data-driven solution that offers meaningful insights for educators.

1.6. Research Questions

The formulation of research questions (RQs) is an essential step in guiding the direction of this thesis. Each question is designed to contribute to the development of an integrated framework combining ABM, ML, and gamification for enhancing educational outcomes. The rationale behind these questions is grounded in the need to explore and validate the efficacy of these technologies within educational settings, particularly in managing disruptive behaviours and improving engagement and performance. The following research questions were defined:

- *RQ1: How can Agent-Based Models be utilised to explore the influence of disruptive students on their peers and the roles of teaching quality, teacher control in a disruptive classroom?*

Disruptive behaviours, such as inattentiveness and hyperactivity and impulsiveness can significantly hinder the learning environment. The first RQ aims to simulate disruptive behaviour to explore the extent to which these behaviours affect other students' performance and classroom dynamics, considering the role of teacher control and classroom disengagement. By identifying the impacts of different types of disruptive behaviours using simulation, this research can inform targeted interventions to mitigate their negative effects.

- *RQ2: How can we predict and explore students' learning outcomes based on disruption-related features (Inattentiveness, Hyperactivity, Impulsiveness), using ML models and Explainable Artificial Intelligence (XAI)?*

The second RQ aims to develop predictive models for students' learning outcomes based on disruptiveness-related features. Inattentiveness, hyperactivity, and impulsiveness are common behavioural issues that can impede learning. By using these features in predictive models, this research seeks to identify students at risk of poor academic performance early for educators to provide the needed interventions. Additionally, this question addresses the need for transparency in predictive models through XAI. Understanding the relationship between learning outcomes and disruptive behaviours is crucial for researchers in education policymaking. XAI provides insights into how these features influence predictions, making the models more interpretable and trustworthy. This research aims to use XAI to elucidate the connections between disruptiveness and academic performance, to inform researchers in educational decision-making.

- *RQ3: How can Machine Learning (ML) be integrated into an agent-based model (ABM) to improve the simulation of classroom disruptive behaviour, and what parameters of ML prediction yield realistic results in this hybrid ML-ABM approach?*

  Having explored the impact of disruptive behaviour features with ML, the third RQ focuses on enhancing ABM simulations with ML techniques. ABM allows for the modelling of complex interactions within a classroom but integrating ML can improve the accuracy and predictive capabilities of these models. This RQ also identifies the key parameters that influence the accuracy of ML predictions within the hybrid ML-ABM model. Determining these parameters is essential for ensuring that the model produces realistic and reliable results.

- *RQ4: How can gamification strategies be implemented to increase engagement in an educational setting, and which gamification elements have the most significant impact on engagement, both in student-oriented systems and teacher-oriented systems?*

To address disengagement caused by disruptive behaviour in learning environments, the fourth RQ seeks to investigate the effectiveness of gamification. Gamification has been widely adopted in various educational contexts to enhance motivation and engagement. Therefore, this research provides further evidence on its impact within MOOCs for the reason that it is a learning environment suitable for application of predictive models. This question aims to identify which gamification elements are most effective in increasing engagement and how these effects can be monitored in real-time. Understanding the real-time impact of gamification elements will enable educators to adapt their strategies to maximise student engagement. Moreover, this RQ explores the potential of gamification to enhance engagement within ABM systems. By integrating gamification elements into ML-ABM model, this research aims to determine whether these elements can increase user engagement and, if so, identify which elements are most effective. Understanding the role of gamification in ML-ABM model can provide valuable insights for developing more engaging and interactive educational simulations.

1.7. Research Contribution

As highlighted above, teachers are ultimately responsible for students' disruptive behaviour. Irrespective of the learning institution's robust, clear, and concise instructions and regulations, teachers must devise suitable and effective strategies to ensure that learners avoid inappropriate behaviour. Understanding students' characteristics and motives, and adopting suitable class management might not only be cumbersome but also subjective because implemented strategies will be based on the teachers' perception of these factors.

This study proposes, for the first time, a simulation model to aid teachers in class management and mitigation of disruptive behaviour. The simulation is based on Agent-Based Modelling, thus, making students agents in a model that mimics a classroom environment where the agent interacts and simulates interactions in a classroom, including disruptive behaviours and potential suitable mitigation strategies. Using a simulation model gives teachers an opportunity to run different scenarios of disruptive behaviours and possible solutions before they can actually use them in their classroom. This approach minimises teachers' use of trial-and-error strategies to try and control and manage their classroom, which eventually might waste time or make them feel disempowered in case of frequent strategy failures.

As machine learning models is known for its predictive capabilities, it is then integrated into the ABM to enhance its performance through combining the strength of ABM in simulating disruptive behaviour and student interactions and predicting learning outcomes with ML. This represents a novel approach in using machine learning with ABM to create a simulated environment for exploring the impact of disruptive behaviour on learning outcomes.

To effectively capture teachers' requirements, the study used a semi-structured interview to understand the teachers' preference of gamification elements. Teachers' responses of gamification elements were collected and analysed to improve the simulation model. To the best of my knowledge, this is the first time such an ML-ABM approach has been developed and presented to collect teachers' feedback on gamification elements used and those that could be added to the model to improve teachers' engagement.

1.8. Thesis Outline

This thesis is structured into nine chapters. The remaining chapters are organised as follows:

In Chapter 2: I present an overview of related work to gamification and Agent-Based Modelling in education as well as Explainable AI and the integration between Agent-Based Modelling and Machine Learning.

In Chapter 3: I present the methodology followed in answering the research questions, including a description of the datasets used in this thesis, preparation steps and the evaluation metrics applied.

In Chapter 4: The first version of simulation model is presented, with the technical details of its development and the initial results found.

In Chapter 5: The answer to RQ2 is presented via the use of ML models in investigating the impact of disruptive behaviour in a physical classroom. Explainable AI was also used to provide an understanding of ML results.

In Chapter 6: An improved version of the ABM approach is proposed, as an answer to RQ3 through the integration of ML. The hybrid ABM-ML model is presented, and its results are described and discussed.

In Chapter 7: The answer to RQ4 is presented via the use of ML models in investigating the impact of gamification elements in an online platform.

In Chapter 8: I introduce an exploration of the efficacy of integrating gamification strategies within ABM system for teachers. Their feedback and evaluation are presented and discussed.

In Chapter 9: A detailed and comprehensive discussion of the findings of this thesis is presented. An explanation of the importance of the contributions of this thesis is provided as well as an identification of the limitations for all findings. Finally, suggestions of future research areas are also included.

In Chapter 10: In this chapter, the thesis is concluded by outlining key contributions and findings.

CHAPTER 2

2 Background and Literature Review

In classrooms and other learning environments, students are subject to multiple elements that contribute to or hinder their achievement. One of the main elements is disruptiveness, either their own or that of others, which has a different effect on groups of students, creating variations in their performance. The interactions that take place in the classroom and how they affect school children's achievement have received much attention by literature over the years [34, 39, 180]. Disruptive behaviour, such as inattentiveness, hyperactivity or impulsiveness, profoundly affects the learning environment [57].

Inattentiveness indicates moving between tasks, leaving one unfinished before losing interest; hyperactivity implies excessive movements in a situation where calmness is expected, while impulsiveness is a tendency towards quick reactions, without proper thinking, disregarding negative consequences of these reactions [205]. These types are symptoms of the attention-deficit hyperactivity disorder (ADHD) that has a clear negative impact on children's long-term academic performance [135]. DuPaul et al. [54] observed that inattentiveness involves a lack of focus, daydreaming, or simply not following instructions. The study acknowledged that inattentive students are susceptible to struggling with classroom activities, appearing distracted or disinterested [54].

On the other hand, Barkley [18] asserted that hyperactivity is characterised by excessive energy or movement. The study highlighted that hyperactivity in students often manifests as an inability to remain seated, excessive talking, or impulsive actions, which often disrupt lesson flow besides distracting other students. By modelling these disruptive behaviours, educators and researchers can devise strategies to mitigate their impact, thereby enhancing the educational experience for all students [38].

Agent-Based Modelling (ABM) is defined as a framework for modelling simulations between autonomous units, known as agents, within a particular environment with defined behaviours that influence their interactions [67, 123]. In other words, ABM creates agents to represent entities within a given environment and setting [203]. Research indicates that the ABM is used in education to enhance educational processes [186], support learning activity [158], and promote student engagement by simulating the emotional states of learners and teachers [185]. [67] noted that simulating an educational setting helps analyse emerging behaviours and patterns, which are critical in understanding classroom dynamics, instructional strategies, and student learning processes. Ingram and Brooks [89] carried out an attempt in simulating a classroom environment built in the NetLogo platform, to visually reproduce classroom activities through creating "playbacks" of specific lessons. The primary challenges they identified include developing rules to capture real behaviours accurately and validating the model's predictive ability. Thus, ABM presents itself as an excellent tool for modelling the type of behaviours targeted in this thesis.

On the other hand, Machine Learning (ML), a branch of machine learning, has enhanced data analysis and predictive modelling [94]. Machine learning algorithms are adept in analysing extensive student learning and pedagogy-related datasets to identify trends and patterns in students' performance, behaviour, and learning preferences [14, 70, 172]. These insights are crucial in tailoring learning experiences to individual needs. Machine Learning's predictive capability is essential in forecasting students' expected learning outcomes, identifying at-risk students, and recommending appropriate interventions. Also, Jordan and Mitchell [94] noted that machine learning algorithms are versatile and well-suited for various educational settings, including both physical and online platforms. Overall, ML adaptability and predictive power make it a valuable tool for enhancing educational outcomes and fostering a more personalized learning experience for students across diverse learning environments. ML has been identified in this thesis as being about to formulate targeted interventions by analysing disruption-related characteristics such as inattentiveness, hyperactivity, and impulsiveness.

Gamification in education involves integrating game-design elements into non-game contexts to bolster user engagement, motivation, and learning outcomes [51]. It introduces aspects such as point scoring and competition into educational activities. Research indicates that gamification can significantly enhance student engagement and motivation [8, 163, 179], especially in online learning environments like Massive Open Online Courses (MOOCs). [97] observed that infusing educational activities into gamification might make learning more interactive and enjoyable, potentially improving student retention and achievement [97].

Although ABM, ML, and gamification have each proven effective individually in enhancing educational processes, they have not yet been integrated into a model for addressing classroom disruptive behaviour. Baker and Inventado [12] and Kapp [97] noted that ML's ability to simulate educational environments, combined with its capability to analyse and predict educational outcomes, alongside gamification's ability to enhance engagement, promises substantial benefits. This thesis integrates these techniques to create a more efficient and engaging model by harnessing the unique strengths of each approach.

Based on the above, the current thesis explores how these elements can be brought together, to specifically target disruptive behaviour, which has not been done before.

Also, this chapter discusses prior research in these areas in more details, identifying gaps, and starting with the main target and motivation: disruptive behaviour in classrooms. It then explores modelling and simulation in education, machine learning in education and their explainability, existing research overlapping e.g. ABM and ML, and finally, gamification in education. The epilogue explains how this thesis brings all these strands together.

2.1 Disruptive behaviour in classrooms

2.1.1 Disruptive Behaviour and ADHD in educational settings

Research indicates that disruptive behaviours such as inattentiveness, hyperactivity and impulsiveness among learners have a complex and multifaceted impact on the learning process [55, 167]. A large body of research draws parallels between ADHD-related symptoms and disruptive behaviour, both leading to poor academic performance that needs addressing [14][62]. As defined in the American Psychiatric

Association's Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Association, 1994), ADHD is a behavioural condition that makes focusing on everyday requests and routines challenging. This disorder affects a person's life on several aspects, from a young age as well as later in life. For the education aspect, several works have been carried out to understand the effect of ADHD symptoms over student performance [81][86]. These showed that ADHD-related behaviour is an important factor in the student learning environment. In this paper, the focus is on disruptive behaviour in general, which may, but does not have to, encompass ADHD.

Research shows that the consequences of these disruptive behaviours extend beyond the individual students, resulting in distractions, loss of instructional time, and a general decline in classroom morale [57]. An early study [167] supports this claim, noting that peer characteristics influenced student behaviour and achievement. Baker [11] found that peers can shape students' attitudes and behaviours, either mitigating or exacerbating disruptive tendencies. Later, Chang and colleagues [38] delved deeper into this relationship, suggesting that learners' disruptive behaviours presented effects among their peers, including potential anxiety, decreased motivation, and lower academic performance.

Addressing students' disruptive behaviours and potential mitigation strategies, Emmer and Stough [56] explored the significance of educators in class management. They identified educators' quality of teaching and classroom management skills as essential in addressing learners' disruptive behaviours. They argued that clear behavioural expectations and consistent rule enforcement can create a structured learning environment that deters disruptive behaviour.

Also, Marzano [128] emphasised that effective instructional strategies and a strong teacher-student rapport can significantly reduce the frequency and impact of disruptive behaviours. By engaging students through interactive and tailored teaching methods, educators can minimise opportunities for off-task behaviour. Marzano also pointed out that a positive teacher-student relationship acts as a buffer, reducing the likelihood of escalation of disruptive behaviour and enhancing the overall classroom climate. Supporting this approach, Simonsen et al. [26] found that proactive management strategies, such as modifying tasks to suit diverse learner needs and maintaining an engaging pace, can pre-emptively address potential disruptions.

Concluding, whilst many works have accomplished predicting student learning outcome, based on features related to ADHD [149] [207] [95] [45] [195]. However, disruptive behaviour has been less explored, specifically with respect to the three disruptiveness-features I select here in this thesis: inattentiveness, hyperactivity and impulsiveness, although they have been found to be strongly related to ADHD [137][41]. In the data used for this thesis, 'Inattentiveness', 'Hyperactivity' and 'Impulsiveness' behaviours are scored with a scale following the diagnostic criteria for ADHD by the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Association, 1994) [135][136], due to the parallels that can be drawn between ADHD and disruptive behaviour.

2.1.2 Impact on Peer Learning and Classroom Environment

Examining the effect of disruptive behaviours from a teacher's perspective, Houghton et al. [86] noted that the frequency, rather than the intensity of disruptive behaviour, was more troublesome in a classroom setting. In classrooms, we usually find a number of pupils, up to a quarter of a class, who display some form of disruptive behaviour

[6]. These disruptive children can have a negative impact on the learning environment and peer dynamics [11][18][36][138][188]. Particularly, Sullivan [188] found that disruptive behaviours mostly divert learners' attention away from educational content, leading to reduced learning opportunities for all students in the classroom.

Similarly, Merrell and Gimpel [138] addressed the effects of disruptive behaviours in a classroom, not their capacity to strain peer relationships and social dynamics. They noted that some students might resent and avoid disruptive behaviours, which could impact the social harmony and the collaborative potential of the class. On the other hand, Carrell, Hoekstra, and Kuka [36] noted that disruptive behaviours waste teacher's time and resources. Consequently, teachers often need to devote extra time and resources to manage disruptive behaviours, potentially detracting them from their ability to provide equal attention to all students [146].

Thus, in this thesis, the effect of peer learning is modelled, as part of the ABM model proposed (see Chapter 4).

2.1.3 Consequences on Educational Outcomes and Classroom Dynamics

Research has shown that classrooms with high levels of disruptive behaviour are susceptible to declining overall academic performance. For instance,found that constant disruptions could lead to lower test scores and grades among disruptive students and the entire class. Moreover, Breslau et al. [29] argued that persistent disruptive behaviour can have long-term effects on students' educational trajectories, including higher dropout rates and lower college enrolment. Also, the effects of inattentiveness and hyperactivity, for instance, extend beyond disruptive students and their peers to include teachers and the overall learning environment [115, 147], thus necessitating adequate mitigation. The consequences of such behaviours have far-reaching implications for the educational environment and student's academic and social development. For instance, Carrell, Hoekstra, and Kuka [36] noted that disruptive behaviours, like excessive talking or inability to stay seated, can interrupt the flow of lessons, making it difficult for other students to maintain focus and absorb material.

On the other hand, Blank and Shavit [22] argued that students’ disruptive behaviours in an education setting affect classroom climate and morale. Particularly, persistent disruptive behaviour can lead to a negative classroom climate, where students feel stressed or anxious, potentially affecting their motivation and engagement [187].

A study by [29] reported that disruptive behaviours led to lower academic achievement among disruptive students and their peers. They stated that continuous disruptions can decrease academic performance for disruptive students and their peers, as consistent distractions hinder learning processes. Also, Sullivan et al. [188] noted that disruptive behaviours reduced student engagement and participation. They stated that students in a disruptive classroom may become less inclined to participate in discussions or activities, affecting their learning and academic confidence.

Also, [146] noted that disruptive behaviours create an inequitable learning environment because some students receive less attention and support due to the disproportionate focus on managing disruptive behaviours.

Thus, in this thesis, educational outcomes and classroom dynamics are simulated to allow teachers to experiment with different settings and evaluate if the outcomes change (see Chapters 4, 6, 8).

2.1.4 Influence of Teaching Quality and Teacher Control

Teacher-student interaction has a significant impact over student achievement [9]. Interactions can be positive, like social and pedagogical interactions [3], or negative, like disruption [11] – such as talking out of turn, aggression or leaving seat [14][7]. Therefore, teaching quality and teacher control significantly influence the management of disruptive classroom behaviours and enhance the learning environment. A study by Marzano [129] explored the contribution of instructional strategies towards mitigating disruptive behaviours. The study noted that high-quality teaching, including using interactive teaching methods, differentiated instructions, and active learning techniques, would enhance student engagement and minimise opportunities for disruptive behaviours. Moreover, Marzano [129] attested that active learning techniques, which shift the focus from passive reception to active participation, have been proven to foster a more dynamic and attentive classroom environment.

On the other hand, Hattie [82] examined the impact of teacher-student relationships in mitigating disruptive behaviours. Hattie [82] found that positive behaviours, including teachers showing respect, empathy, and genuine interest in their students' well-being, significantly reduced disruptive behaviours by allowing teachers more and better classroom control. Students feel valued and understood, which enhances their self-esteem and motivation. Hattie [82] noted that mutual respect and understanding fostered through these relationships encourage students to engage more positively with their learning and peers, creating a harmonious and productive classroom setting.

Also, Maag [120] noted that teachers mitigate disruptive classroom behaviours through positive reinforcement of students’ behaviour. Maag [120] observed that acknowledging and rewarding appropriate behaviour effectively promoted desired behaviours. Behaviour modification techniques can also be applied to reduce disruptive actions gradually.

Hence, in this thesis, Teacher Quality and Teacher Control become crucial variables, which can be set in the simulation by the teacher, to better understand how they influence the outcomes of the classroom learning process (see Chapter 4).

2.2 Modelling and Simulation in Education

Modelling in education is used to serve different purposes. An early study by Winne and Hadwin [204] introduces a model for self-regulated learning (SRL) that focuses on student management of their learning process. The model suggests that learning involves different phases. It starts with planning then executing strategies followed by monitoring progress and ends with evaluating outcomes. Vosniadou [202] also introduces a model that tracks how students' preconceptions can support or hinder their learning. Gobert and Buckley [71] emphasize that models are not just representations but also tools for thinking and reasoning. They suggested model-based learning where students can actively create and revise models to better understand scientific phenomena.

Gredler [75] argues that simulations are effective because they allow students to engage with complex, real-world problems in a controlled environment. They build experience that helps in developing their decision-making and critical-thinking skills. [184] investigate the effect of augmented reality games on students' scientific argumentation skills. In their experiment, students participated in a location-based simulation game where they investigated real-world scenarios. Their findings showed potential for simulation game in improving students' hypothesis formation, evidence gathering, and argumentation. Simulations can also be used as a substitute of physical experiments. Finkelstein [61] investigated whether virtual simulations could replace physical lab equipment in physics education. Their findings showed that, in many cases, students learned more effectively through computer-based simulations than with traditional lab setups, due to the immediate feedback and the ability to conduct multiple experiments quickly. In the field of medical and surgical training, the study by Kneebone [103] suggests that simulations allow trainees to practice surgical techniques in a risk-free environment, which enhances skill acquisition and confidence before working with real patients. Girvan and Savage [69] examined the use of virtual worlds as educational simulations to teach 21st century skills such as problem-solving, collaboration, and digital literacy. The study shows that interactive simulations that replicate real-world scenarios help students develop these skills more effectively. Modelling and simulation advances have impacted various areas of education from classrooms in primary learning to pilot and surgical training in professional learning. These studies emphasise the importance of utilising models to provide a safe environment for students to practice new skills and deepen their understanding of educational content as well as support various educational goals. The use of ABM and ML methodologies in simulating and predicting disruptive behaviour impact on learning outcomes discussed in Chapters 4, 5 and 6 underscores the value of modelling and simulation in education, as these tools enable a practical, interactive approach to addressing disruptive behaviours and tailoring interventions for improved classroom management.

2.2.1 *Agent Based Modelling*

Agent-based modelling (ABM) is a framework for modelling the simulation of interactions between agents in a defined environment with a set of behaviours that influences those interactions [15]. Agents interact with other agents and with the environment based on a set of behaviours driven from personal characteristics and preferences. An agent can represent an individual or a group whereas their relationship in a simulation represents social relations.

Agent-based modelling has been adopted in the field of education to serve different purposes. Some utilized it as a support of the learning activity by modelling games for younger students such as the case with Ponticorvo et al[159] where they introduced an ABM framework for developing digital games for children. In another study [46], ABM was used to simulate the evacuation process from a classroom during an emergency. They studied the simulation of a collaborated classroom and modelled the evacuation with ABM using 5 different possible configurations of a collaborative classroom and proposed the best match of classroom collaborative design and the location and type of exit doors. Other research considered using this tool to improve the educational process by observing the effect of students and lecturers' negative emotions over student engagement [72]. Their findings suggest that pupils' negative emotions are influenced by the teacher's characteristics, such as poor communication skills and poor teaching. Another model of student behaviour by [154] that focuses on cheating in assignments. Their model showed a strong connection between cheating and participating in extracurricular activities which called for their next step, using the model to find the balance between outside activities and student knowledge level. [10] proposed a proof-of-concept model of teacher's and pupils' interactions with educational content in a classroom. The model aimed to help educational researchers and stakeholders to improve prediction of pupils' learning outcomes and choice of interventions. However, this model did not account for important factors, such as students’ social interactions or disruptive behaviour, which can significantly impact learning outcomes. Disruptive behaviour is a critical element in classroom dynamics that, if unaddressed, can hinder both individual and group learning. Therefore, ABM in this thesis, as outlined in Chapters 4 and 6, utilises disruptive behaviour for simulating classroom settings in the aim of capturing nuanced interactions that inform effective teaching strategies and improve behavioural management.

2.3.2 Predictive Modelling in Education

Predictive models can provide educators with insights into the factors that most significantly impact student learning, guiding intervention strategies [84].

In their study of ML trends and their transformative potential in data analysis, Baker and Inventado [12] studied educational data mining and learning analytics, underscoring the significance of data-driven strategies in effectively addressing diverse learning needs and improving overall educational outcomes. They found that ML helps develop actionable insights, enabling educators to tailor their approaches to individual student needs. For instance, they illustrated the use of ML in predicting at-risk students based on behavioural and academic indicators, prompting timely and targeted interventions.

2.3 Role of Machine Learning in Educational Contexts

Machine Learning, a subset of artificial intelligence, involves algorithms that learn from data and make predictions or decisions. Baker and Inventado [12] noted that ML has transformative potential in reshaping educational strategies and emphasising its role in enhancing both teaching efficiency and student learning outcomes. It can analyse various types of data, including student engagement metrics, performance data, and interaction patterns [12]. The analysis facilitates a deeper understanding of student behaviours and learning processes. For instance, by scrutinising engagement metrics, ML algorithms can identify patterns that predict student success or flag potential areas of concern, such as disengagement or the likelihood of dropout [9]. Furthermore, Metz and colleagues noted that when these algorithms are applied to performance data, they can forecast academic outcomes, thereby assisting educators in tailoring their instruction to meet individual student needs more effectively. Additionally, ML's ability to analyse interaction patterns offers invaluable insights into the dynamics of learning environments, enabling the creation of more collaborative and supportive educational experiences.

2.3.1 Applications of Machine Learning in Education

The application of Machine Learning (ML) techniques in educational contexts, particularly for predicting student performance, underscores significant advancements. Jordan and Mitchell [94] highlighted the efficacy of various ML algorithms in dissecting complex educational data. They noted that regression analysis can quantitatively assess the impact of specific attributes on academic achievements. Also, they asserted that classification models are instrumental in categorising students based on the likelihood of experiencing academic difficulties. Moreover, they observed that neural networks can leverage their sophisticated pattern recognition capabilities to analyse non-linear and intricate relationships between a myriad of student behavioural attributes and their learning outcomes, thus offering critical insights to academic performance. [206] explores the integration of ML for analysing student performance data, creating personalized learning experiences, and predicting academic outcomes. Their findings emphasise the role of algorithms like neural networks, decision trees, and support vector machines in enhancing learning processes and optimizing student engagement. Another review by [142] covered six main themes of ML in online education, including adaptive learning, dropout prediction, intelligent tutoring, and performance analytics. They also review algorithms commonly used in these areas, like decision trees, random forests, and support vector machines. Some ML models are commonly used in research for their performance. Kotsiantis [104] proposed predicting student performance by a decision support system using students' e-system log data and student academic data. Five ML models were used in their study, namely: Support Vector Machine (SVM), Model Tree (MT), NN, Linear Regression (LR), and Locally Weighted Linear Regression. Also, Hu et al. [88] proposed a warning system for predicting student at risk in an online learning environment using time-dependant variables. They applied three classifiers; Regression Tree (CART), Logistic Regression (LGR), and Adaptive Boosting but CART outperformed the other two classifiers with 95% accuracy. Reddy and Rohith [24] used ML to identify the explainable characteristics that could show the potential for student poor performance using SVM, Random Forest (RF), Gradient Boosting, and Decision Trees (DT). Similarly, prediction of student dropout was investigated by Sara et al. [170] through a dataset consisting of 72,598 instances and 17 attributes. The classifiers used in this study were RF, CART, SVM, and Naïve Bayes (NB). In Chapter 5 of this thesis, ML was used to predict student performance using the classifiers: XGBoost, Gradient Boosting, Ada Boost, Random Forest, Extra Trees, Logistic Regression, KNN, and MLP for their common use in student performance prediction [14][152][89][39].

2.3.2 Predicting Outcomes Using Disruption-Related Features

There is growing interest among educational psychologists and data analysts in student learning outcomes, particularly in relation to disruption-related features', especially inattentiveness, hyperactivity, and impulsiveness [42, 201]. Understanding how these behavioural traits influence academic performance is crucial for educators and policymakers in devising effective interventions [125]. Kim et al. [100] examined the use of wearable technology to gather data on children's sleep and activity patterns for ADHD detection. Researchers found significant accuracy with ML models, suggesting potential early interventions through behavioural monitoring. [126] focused on classifying ADHD in children based on behavioural data. Researchers used a dataset of 45,779 children aged 3-17 from the 2018–2019 National Survey of Children's Health. Within this population, 11.4% were diagnosed with ADHD. They applied logistic regression to identify key risk factors, such as demographic and behavioural indicators. Another study using behavioural attributes, [116] employed sequential engagement patterns in learning management systems to monitor student behaviour and predict academic success. By analysing features like response time, attendance, and interaction frequency, the study provides insights into students' learning habits. Ter-Minassian et al. [190] examined machine learning models in predicting ADHD in primary school students by linking educational records with healthcare data. Their findings suggest that fair, accurate ML models could support early ADHD identification, especially in diverse student populations. These studies collectively show the value of integrating behavioural data into ML models for a more accurate prediction of targeted outcomes. Thus, ML capabilities are used to explore and predict student outcome using behavioural features, specifically for the first time, three disruptive behaviour features as shown in Chapter 5.

2.3.3 Enhancing Agent-Based Models with ML

Research indicates that integrating ML and ABM in an educational setting would present a cutting-edge approach to analysing educational dynamics and accurately predicting learning process outcomes [67, 94]. Gilbert [67] highlighted the versatility of ABM in representing diverse educational scenarios, from classroom interactions to institutional policy impacts.

Emerging research on adopting ML-enhanced ABM in class behaviour management has recorded positive outcomes in predicting the outcome of various behavioural interventions based on students' backgrounds and teacher responses [156]. Furthermore, integrating ML algorithms with ABM facilitated the identification of key parameters, such as student engagement levels and interaction frequencies that significantly influenced the model's outcomes. By accurately simulating these complex variables, the hybrid ML-AB models could generate more reliable predictions about the effectiveness of educational strategies, thereby informing better decision-making in educational settings. These findings underscore the potential of combining ABM and ML in creating robust, adaptable models that can significantly contribute to our understanding and improvement of educational systems.

This thesis aims to leverage ML techniques to predict and manage disruptive behaviours in real-time, as covered in Chapter 5. Traditional methods of behaviour management often rely on static interventions and retrospective analyses, which do not adequately address the dynamic nature of classroom interactions. By employing advanced ML algorithms, this research aims to develop predictive models that can identify patterns and trends in student behaviour, enabling educators to implement timely and effective interventions. These predictive models will be based on historical data and real-time inputs, offering a proactive approach to managing classroom disruptions. The thesis also aims to utilise ABM, as illustrated in Chapters 4 and 6, to simulate complex classroom environments and interactions.

2.3.4 Integration of Machine Learning in Gamification Analysis

The integration of Machine Learning (ML) in analysing the effects of gamification in learning and student engagement represents a significant advancement in educational technology. Jordan and Mitchell [94] noted that ML algorithms are versatile in managing large, intricate datasets; thereby, they are ideal for dissecting the multifaceted nature of gamified environments. Machine Learning algorithms can identify nuanced patterns and relationships within educational data, offering insights far beyond traditional analytical methods [23, 63]. Therefore, ML algorithms can be adopted to evaluate the complex interactions and outcomes associated with gamified learning environments.

Using ML to analyse the effects of gamification has transformed the understanding of educational dynamics. On the one hand, ML techniques help discern the specific gamification elements that most effectively boost student engagement and learning outcomes [106]. Machine learning techniques analyse the effectiveness of points and badges in motivating students and evaluate the role of leaderboards in fostering a competitive yet collaborative learning atmosphere [83]. By leveraging ML's powerful analytics, educators and technologists can fine-tune gamified elements, ensuring they align closely with educational objectives and student preferences.

In summary, the application of predictive models in education, particularly in managing disruptive behaviours, extends beyond identifying at-risk students. It empowers educators and policymakers to make informed, data-driven decisions in crafting interventions, allocating resources, and developing policies that directly address the needs and challenges identified through these models. Additionally, in this thesis, ML in educational contexts is used in Chapters 5, 7. In Chapter 7, ML was used to understand the effect of gamification elements in an online gamified learning environment on students' engagement [4]. In Chapter 5, ML was applied to investigate the impact of different disruptive behaviour on academic performance in a physical classroom environment [6][7].

2.4 Explainability of Predictive Models (XAI)

Modelling disruptive behaviour with ABM simulation and ML predictions, involves identifying patterns and predicting outcomes. However, if these models operate as "black boxes" without clear reasoning for their predictions, it can be difficult for educators and stakeholders to trust or understand the results. XAI techniques, such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations), would enable the thesis to present predictions in a transparent way. Explainable Artificial Intelligence (XAI) refers to methods and techniques in AI that help human experts understand the results of the solution [19]. It contrasts with the "black box" nature of many AI models, particularly in complex machine learning [172]. Within educational settings, XAI plays a crucial role in elucidating the relationship between learning outcomes and disruptive behaviours, like inattentiveness, hyperactivity, and impulsiveness [208]. In a recent study to examine the adoption of hybrid human-AI to predict student performance based on whether they will fail or pass, Xu, Moon, and Van Der Schaar [207] suggested a change in student disruptiveness to influence students' grades of assignments.

Gunning and Aha [78] stated that explainable AI is essential in enhancing the understanding of model decisions. It helps educators and stakeholders understand why AI models make certain predictions or decisions. This understanding is crucial for justifying and implementing interventions based on model predictions. Moreover, Gunning and Aha [78] noted that explainability helps build trust. They noted that decisions in an educational setting significantly affect students. Therefore, the ability to explain AI decisions builds trust among educators, students, and parents. Explainable AI techniques used in an educational setting include Local Interpretable Model-agnostic Explanations (LIME) [25], SHapley Additive exPlanations (SHAP) [90], and visualizations [99]. Local Interpretable Model-agnostic Explanations (LIME) explains AI model predictions by approximating them locally with an interpretable model. On the other hand, SHAP measures the impact of each feature on the prediction, providing insight into how different behavioural aspects contribute to learning outcomes. Lastly, visualisations encompass graphical representations of data and model decisions to simplify complex relationships and make them accessible and understandable to non-experts.

This thesis leverages machine learning models in utilising features of disruptiveness (Inattentiveness, Hyperactivity, and Impulsiveness) to predict student learning outcomes and followed these predictions with the use of XAI to elucidate these predictions. This research applies XAI to uncover the connection between features of disruptive behaviour and student performance to dig deep into this connection and provide a new understanding of this relationship. Chapter 5 details the performance of ML prediction with disruptive features (Inattentiveness, Hyperactivity, Impulsiveness), as well as the contribution of each feature to the prediction of student performance.

2.5 Enhancing Agent-Based Models with Machine Learning

The integration of Machine Learning (ML) into Agent-Based Models (ABMs) represents a significant advancement in simulating and understanding complex systems, especially in educational settings [176]. This integration allows for more sophisticated and nuanced models to better predict and adapt to dynamic educational environments.

2.5.1 Understanding the Synergy between ML and ABM

The collaboration between ML and ABM can be explained through four perspectives. First, they could be understood from their complementary strengths, including ABMs and emergent phenomena, the use of ML in pattern recognition and prediction, and Synergy in Predictive Modelling. Epstein [57] noted that Agent-Based Models (ABMs) are adept at simulating complex systems by modelling the interactions of individual agents. They are particularly effective in capturing emergent phenomena that arise from these interactions. For instance, ABMs can simulate classroom dynamics by modelling individual student behaviours and interactions. On the other hand, Jordan and Mitchell (2015) noted ML utilises its strength in recognising patterns and making predictions from large datasets to complement ABMs. They observed that ML algorithms can analyse historical data to identify trends and patterns that are not immediately apparent. For example, ML can predict student performance based on historical academic data and behavioural patterns. Sankaranarayanan & Portman [169] addressed the synergy between ML and ABM, noting that the integration of ABM and ML creates a powerful tool for predictive modelling. They observed that while ABMs provide a structural framework for simulating complex systems, ML contributes to its predictive analytics capability, enhancing the overall adaptability and accuracy of the simulations.

Second, collaboration between ML and ABM could be understood through data-driven behaviours in ABMs. This approach involves examining the incorporation of ML into ABMs, the realism and dynamics in simulations, and case study applications. Rosés, Kadar, & Malleson [166] observed that by integrating ML algorithms into ABMs, the agents' behaviours can be informed by data-driven predictions. This integration leads to simulations based on predefined rules and adaptations according to evolving data patterns. Also, research has established that the combination of ML and ABMs can produce more realistic and dynamic models, [94].

Third, the collaboration between ML and ABMs could be explained through their enhanced predictive power and adaptability. Epstein [57] explained the ML and ABMs synergy from the point of adaptability and simulations, asserting that the adaptability of ABM simulations is significantly increased by incorporating ML. The study highlighted that as real-world data is continuously fed into the model, the simulated agents can adapt and evolve, leading to more accurate and timely predictions.

Incorporating ML into ABMs offers a powerful approach to understanding and addressing the complexities of educational environments. Through data-driven simulations, educators and policymakers can gain invaluable insights into student behaviours, learning processes, and the impact of various educational strategies. As these methodologies continue to evolve, they hold the potential to enhance educational planning, policy-making, and classroom management significantly. In Chapter 6, this thesis provides - for the first time - a data-driven simulation utilising ABM's capability to simulate agent interactions and ML's proficiency in pattern recognition and prediction for classroom dynamics focusing on disruptive behaviour.

2.5.2 Methodologies for ML and ABMs Integration

One method for integrating ML with ABMs is the use of data-driven models to inform agent behaviours. Rosés, Kadar, & Malleson [166] observed that machine learning and ABMs can be integrated by training ML models with ABMs or using ABMs to generate data for ML. The former involves embedding ML models within individual agents or the environment to inform decision-making processes. This approach might be useful when real-world data is limited or difficult to obtain. Another method for ABM and ML integration is using ABM to generate synthetic data, which can then be analysed using ML algorithms. Platas-López et al.[161] explored the uses of ML to overcome challenges of ABM. Another approach for ML and ABM is feedback loops, where ML models provide predictions that influence agent behaviours in ABM, and the outcomes of these behaviours feed back into the ML model to provide a refined outcome. Sankaranarayanan & Portman [169] discuss how feedback loops enhance ABM by adjusting agent behaviours based on ML predictions. As agents interact and the system evolves, ML algorithms continually update predictions based on the emergent behaviours within the ABM, leading to increasingly accurate simulations.

In conclusion, the integration of ML into ABMs has been successful across various domains, demonstrating the potential of this approach in educational settings which is yet to be explored. It has been used in adaptive learning environments, policy simulation and decision-making, class management and behavioural interventions, and predicting and addressing dropout rates. Traditional ABM has been effective in simulating the complex interactions within classrooms, allowing researchers to observe the emergent behaviours resulting from individual actions and environmental factors. However, the predictive accuracy of ABMs can be limited by their inability to incorporate historical data and adapt dynamically to new information. By integrating ML algorithms, which can analyse vast datasets and generate predictive insights, the model's ability to simulate real-world scenarios with higher accuracy is significantly improved. In Chapter 6, this thesis presents a hybrid ML and ABM approach that leverages the strengths of both approaches, aiming to create a more robust tool for educational researchers and policymakers.

2.5.3 Parameters for Realistic Hybrid ML-AB Models

The integration of Machine Learning (ML) and Agent-Based (AB) models in educational research necessitates a comprehensive understanding of the parameters contributing to these hybrid models' realism and effectiveness. Identifying and optimising key parameters is crucial in ensuring that the models accurately reflect real-world educational dynamics and can provide actionable insights.

2.5.4 Identification of Key Parameters in Hybrid ML-AB Models

Key parameters in hybrid ML-AB models include agent characteristics, learning environment dynamics, behavioural factors, and data-driven parameters [209]. In ABM models, agents represent entities such as students, teachers, or educational resources. Key parameters include agent attributes (e.g., learning styles, behaviour patterns), decision-making rules, and interaction protocols. These parameters must accurately reflect the diversity and complexity of real-world educational settings.

Parameters related to the learning environment include classroom size, teaching methods, and curriculum content [3]. These elements impact how agents (students and teachers) interact within the model. On the other hand, behavioural factors incorporate parameters representing disruption-related behaviours, such as inattentiveness, hyperactivity, and impulsiveness. Accurately modelling these behaviours is essential for studying their impact on learning outcomes.

ML components in hybrid models utilise parameters derived from real-world educational data. These might include historical academic performance, demographic information, and behavioural assessments [1]. The selection of these parameters is guided by their predictive power and relevance to the research questions. This thesis aims to develop predictive models for students' learning outcomes with parameters including disruptiveness-related features (see Chapter 5). Inattentiveness, Hyperactivity, and Impulsiveness are common behavioural issues that can impede learning. By using these features in my predictive models, this research seeks to utilise ML's power to make accurate predictions for the integration with ABM as shown in the development of hybrid ML-ABM approach in this thesis (see Chapter 6).

2.5.5 Validation and Testing of Hybrid Models

The process involves validation against real-world data, cross-validation, scenario testing, and iterative testing [92]. Validation against real-world data encompasses comparing model outputs with actual educational data to assess accuracy. This involves validating the model's predictions on key metrics like student performance, engagement levels, and behavioural incidents. Also, the validation and testing process uses cross-validation techniques to test the model's reliability and generalizability across different datasets and educational contexts. Another step, scenario testing, involves conducting scenario-based tests where hypothetical educational interventions are applied within the model [200]. Analysing the outcomes of these scenarios helps understand the potential impact of different educational strategies. Finally, iterative refinement entails continuously refining the model based on validation outcomes [176]. This iterative process involves adjusting parameters, updating data inputs, and re-evaluating model performance.

This research in this thesis aims to identify the key parameters that influence the accuracy of ML predictions within the hybrid ML-ABM approach. Determining these parameters is essential for ensuring that the model produces realistic and reliable results. In Chapters 4 and 6, this research explores various parameters and their impacts, optimising the model for better performance prediction. It also presents empirical validation of the results of ML-ABM approach detailed in section 8.2.

2.6 Gamification in Education

Gamification is a method that is usually applied for the aim of increasing motivation and ultimately decreasing disengagement. In education, gamification impacts motivation by providing incentives in the form of clear goals, immediate feedback, or a sense of progression. This is essential in educational environments, whether in online learning like MOOCs, where students often feel disconnected, or in physical classrooms, where they may lose interest without immediate acknowledgment or get distracted by disruptions in the classroom.

2.6.1 Impact of Gamification on Student Engagement

Gamification in education refers to incorporating game elements into the learning process to enhance student engagement and motivation. This innovative approach aims to transform the learning experience from a traditional, often passive absorption of information into an interactive, rewarding journey [51, 97].

An early research study by Deterding et al. [51] into gamification established gamification's capacity to enhance motivation and user experience. The study noted that game mechanics, including points, badges, and leaderboards, can be incorporated into learning platforms to create a more engaging and interactive educational environment. They observed that the interactive environment fostered a heightened sense of achievement and progress among learners, thus significantly boosting their motivation and engagement.

Later, Kapp [97] empirically supported the effectiveness of gamification in learning. He found that well-implemented gamification strategies and game mechanics in educational content captivated students' attention and encouraged their continued participation and interaction with the course material. His findings suggest that when students find the learning process enjoyable and challenging, their intrinsic motivation is significantly enhanced, which potentially promotes learning outcomes.

Similarly, [38] addressed the real-time aspects of gamification in an education setting, including the dynamic nature of learner engagement. The authors found that a gamified learning environment supports real-time feedback mechanisms and adaptive learning, significantly enhancing students' responsiveness to curriculum content and instructions, thus, catering to individual students' needs and preferences. This study asserted that integrating gamification into MOOCs enhances student engagement and catalyses creation of more personalised and effective online learning experiences.

Likewise, Kapp [97] emphasises the role of gamification in creating an engaging learning environment. Kapp's research demonstrates that incorporating these game mechanics in educational content captivates students' attention and encourages continued participation and interaction with the course material. His findings suggest that their intrinsic motivation is significantly enhanced when students find the learning process enjoyable and challenging, akin to a game. This motivation is crucial in MOOCs, where self-driven learning is pivotal. Both sets of authors contribute to the understanding that gamification, when effectively integrated into MOOCs, can transform a conventional educational model into an engaging and interactive learning journey, leading to improved student engagement and motivation.

Gamification has been widely adopted in various educational contexts to enhance motivation and engagement as stated. However, empirical evidence on its specific impact within MOOCs was limited at the time this part of the research was conducted. Therefore, in my work on gamification in education for MOOCs, I tackle engagement in a gamified online learning, in Chapter 5. As the design of gamified learning systems is usually theory-driven, there is a lack of runtime feedback, non-gamified scaffolding, and under-exploitation of interaction data. Whilst the theoretical basis is very important in designing purpose-fit gamified systems, in the context of large-scale online learning like MOOCs, it is not feasible to propose a one-size-fits-all design of gamification. For this reason, it is very important to take into account the data generated from the system in order to better understand the users' interactions, and refine the offering. Moreover, gamification in MOOCs was explored in a first instance (Chapter 7, [4]), to better understand gamification as a vehicle for motivation in the ABM approach, as implemented later on (Chapter 8).

2.6.2 Teachers and Gamification

The available related research explored the integration of gamification across various educational disciplines. Wells and Fotaris [64] examine the perceptions of trainee teachers towards gamified methods in London schools, revealing enthusiasm for new pedagogical strategies. Heras et al. [33] advance gamification in chemical engineering with simulators designed to support digital transformation. Baldeón et al. [15, 16] and Dermeval et al. [49] demonstrate how gamification can be embedded in complex systems theory and intelligent tutoring systems respectively, enhancing both student and teacher engagement and fostering a dynamic learning environment. Each study presents gamification as a potent tool for enhancing educational outcomes through active participation and innovative teaching methodologies.

The research by Wells and Fotaris [64] investigates the integration of gamified learning methods in educational settings through the lens of trainee teacher perceptions. This research delves into the potential of game-based learning to enhance student engagement and teaching effectiveness in East London schools. Trainee teachers expressed enthusiasm for adopting new pedagogical strategies and becoming agents of change, despite encountering obstacles like traditional teaching mindsets and structural limitations within schools. The study underscores the need for substantial support and training for teachers to effectively implement gamified learning, which could transform educational environments by making them more aligned with contemporary student experiences and expectations. The findings suggest a promising avenue for fostering greater engagement and motivation among students, thereby enhancing learning outcomes and teacher success in utilising innovative educational tools.

The work by Heras et al. [33] focuses on enhancing the educational landscape in chemical engineering. The paper presents a detailed framework for developing pedagogical simulators that integrate gamification elements to make learning more engaging and effective. The framework, termed P2Si, is designed to support the digital transformation in education necessitated by the fourth industrial revolution, emphasising the importance of process models in chemical and biochemical engineering education. The study highlights how these simulators can be used as educational tools to improve student engagement, facilitate a deeper understanding of complex processes, and enhance teaching effectiveness. By incorporating elements such as explanatory models, tailored learning designs, and participatory design involving students as co-designers, the framework aims to make educational simulators more interactive and responsive to the needs of learners.

In conclusion, while the available related scholarly works have explored the integration of gamification across various educational disciplines, there has been a noticeable gap in research specifically focusing on gamification for teachers. Compared to the research in this thesis, which examines how gamification can be utilised to enhance teacher engagement and effectiveness, most studies primarily focus on student outcomes. In Chapter 8, this research addresses this gap by providing an in-depth analysis of the impact of gamification on teachers within the context of the ABM approach proposed, demonstrating how elements like real-time statistics and leaderboards can significantly improve teaching practices and teacher motivation. This focus on teachers is essential, as their engagement is crucial for the successful implementation of gamified learning environments and it aligns with the broader aim of enhancing educational outcomes through innovative strategies.

2.7 Epilogue

Based on the evidence explored from the existing literature presented in this chapter, the integration of ABM, ML, and gamification offers a comprehensive approach to enhancing teaching and learning processes within the realm of educational technologies. Each methodology contributes unique strengths to educational environments, addressing a variety of challenges, from enhancing engagement, to managing behaviour and advancing predictive analytics. This synthesis leverages the distinct capabilities of each method to tackle issues such as engagement, behavioural management, and the development of predictive analytics.

ABM is particularly noted for its dynamic simulation of complex systems, which is highly relevant in educational settings, where interactions between students and teachers can be complex and varied. By utilising ABM, educators can develop virtual models of classroom environments to simulate interactions between agents. This method allows for an exploratory approach to understanding the impact of various pedagogical strategies on student engagement and learning outcomes. The insights gained from these simulations are invaluable, as they reveal the impact from the interactions of individual agents, providing a deeper understanding of classroom dynamics and the potential effects of different instructional strategies [87, 121, 203]. While ABM has shown its potential in the education field, whether in the learning content or the learning process, little attention has been given to its application to classroom interactions—specifically, the impact of disruptive behaviour on the learning environment. Most existing models either focus on individual behaviours or specific outcomes, without accounting for the disruptions in the classroom. Similarly, while predictive models using ML can identify patterns in student performance and behaviour, they lack the ability to simulate the interactions that occur in a classroom. ML complements ABM by adding a layer of data analysis and predictive modelling capabilities. The predictive power of ML is crucial in forecasting learning outcomes and identifying at-risk students, thereby enabling the implementation of timely and effective educational interventions [14, 94]. Thus, this thesis aims to develop an ABM-ML approach that utilises both techniques and incorporate disruptive behaviours and their effects on classroom dynamics. This model could provide educators and researchers with insights to help them in mitigating these behaviours and optimise learning outcomes.

During this analysis, a notable gap in research in the application of hybrid ABM-ML techniques to predict and manage disruptive behaviours in classrooms has been found. Traditional methods often rely on analyses and static intervention strategies, which may not account for the dynamic nature of classroom interactions. This thesis leverages ML algorithms (see Chapter 5) to predict learning outcomes based on disruptive behaviours, providing educators with helpful insights to mitigate disruptions. Furthermore, it utilises ABM to simulate complex classroom environments and interactions as ABM offers a powerful tool to model these interactions, allowing for the exploration of various scenarios and the identification of effective strategies for managing classroom behaviour and enhancing student performance. By identifying the specific impacts of different types of disruptive behaviours using ML techniques combined with ABM simulation (see Chapter 6), this research can inform educators to mitigate their negative effects.

Lastly, gamification enhances educational settings by incorporating game-design elements such as points, badges, and leaderboards to boost engagement and motivation. This strategy creates an interactive experience that not only increases involvement but also promotes persistence, significantly enhancing retention and performance, particularly in online platforms like MOOCs [8, 51, 163]. Gamification can help combat effects of disruptiveness, but also is considered in this thesis as a general motivational element, not just for students, but also for teachers.

The synergy from integrating ABM, ML, and gamification elements has the promise to create a more cohesive and adaptive learning environment. ML can enhance ABM simulations by providing data-driven insights of agent behaviour, thereby increasing the accuracy of the simulation. Simultaneously, gamification benefits from ML algorithms in evaluating the effectiveness of game elements and increasing engagement with the model.

The next step of this thesis is to explain the methodology employed to address the RQs in Chapter 3, to address the gaps as described in this chapter. Thus, this thesis presents and investigates the impact of disruptive behaviour, peers and teacher characteristics on academic performance by simulating classroom interactions in Chapter 4. In Chapter 5, further investigation of disruptive behaviour impact in physical classroom is performed using ML and XAI. Chapter 6 presents an improved version of classroom ABM, with the incorporation of ML. Lastly, Chapters 7 and 8 employ quantitative and qualitative methods to explore the impact of gamification in educational settings.

CHAPTER 3

3 Methodology

3.1 Prologue

After highlighting the thesis research questions in Chapter 1, and discussing relevant literature in Chapter 2, this chapter provides an explanation of the research methodologies used to address the research questions. This includes a high-level description of the data sources and approaches used in the process of answering the research questions provided in section 1.6. More detailed descriptions follow in the individual research chapters, which address specific research questions, and are referenced in this chapter where necessary.

3.2 Research Design and Strategy

A research design represents a researcher's plan to collect and analyse data to answer a particular research question [151]. Therefore, a research design is a roadmap followed during research to achieve the research purpose. A research design specifies the types of methods and procedures that a researcher would use to collect and analyse information. On the other hand, a research strategy outlines how research is conducted, including elements of data collection and synthesis [124]. Pandey and Pandey [151] noted that researchers can choose from exploratory, case studies, grounded theory, and experiments, among other strategies, as general ways of conducting a study. This thesis utilises a data-driven simulation approach to address disruptive behaviour in educational settings through the integration of Agent Based Modelling (ABM) and Machine Learning (ML). The research is conducted in multiple stages, with research questions tailored for every stage's objective. The high-level structure of the research design for answering these research questions is presented in the following:

RQ1: How can Agent-Based Models be utilised to explore the influence of disruptive students on their peers and the roles of teaching quality, teacher control in a disruptive classroom? (Chapter 4)

To understand the effect of teachers and peers using a simulation model.

- PIPS data (see section 4.2) was used to design a simulation model and create and visualise a simulated classroom with students' disruptive behaviour represented with features from PIPS.
- The designed model is run with different scenarios in relation to teacher and peers' characteristics to illustrate their effect, by comparing Pearson correlation coefficient between simulated and PIPS data, in Sections 0 and 4.5.

RQ2: How can we predict and explore students' learning outcomes based on disruption-related features (Inattentiveness, Hyperactivity, Impulsiveness), using ML models and Explainable Artificial Intelligence (XAI)? (Chapter 5)

To explore the effect of disruptive behaviour and other factors in classroom environment.

- Prediction output is defined as two classes representing student performance that is followed by grouping of the data to improve classifier performance (Section 5.4).
- Interpretation of ML predictions using SHAP values to explain the relationship between student performance and disruptive behaviour features (section 5.5).

RQ3: How can Machine Learning (ML) be integrated into an agent-based model (ABM) to improve the simulation of classroom disruptive behaviour, and what *parameters of ML prediction yield realistic results in this hybrid ML-ABM approach? (Chapter 6)*

Machine learning is integrated into Agent-Based Model to better illustrate the effect of disruptive behaviour in classroom using ML.

- A hybrid ML-ABM approach is developed by combining ABM simulation of classroom disruptive behaviour and Machine Learning predictions as demonstrated in Section 6.2.
- The performance of ML-ABM approach is measured using Pearson correlation, MAE and RMSE as shown in Section 6.3.10.

RQ4: How can gamification strategies be implemented to increase engagement in an educational setting, and which gamification elements have the most significant impact on engagement, both in student-oriented systems and teacher-oriented systems? (Chapter 7, Chapter 8)

To explore the effect of gamification on student engagement in MOOCs, different classifiers were applied.

- Gamification elements were extracted and analysed from data (CameleOn 3.4.1) as well as the definition of student engagement (Section 7.2.7).
- ML Classifiers predict that a student is engaged (1) or not engaged (0), based on the identified gamification elements, as shown in Section 7.3.

To understand the effect of adding gamification elements to the ABM simulation for use by teachers, these steps were followed:

- ABM was presented to teachers to collect their feedback on the use of the system using Toda’s taxonomy (Section 8.2).

- Teachers' views on the existing gamification elements as well as their preference of gamification elements were collected via semi-structured interviews.

Following, the main techniques used in the research are described. A specific research methodology is further detailed in the following chapters when it relates to a specific research question.

3.3 Research Methodology Types and Research Stages

Research methodology represents a theory of how a study is conducted [151]. It considers assumptions, principles, and procedures of a specific approach used in the research study. The research methodology restates the research problem and helps researchers to identify, choose, or develop suitable techniques for collecting data. A research methodology can either be quantitative or qualitative. A quantitative methodology is used in studies designed to test theories and facts or examine the relationship between variables and expected study outcomes [124, 140]. In a quantitative research study, a researcher uses sampling methods to select study participants and standardised tools like questions to collect data. The study relies on statistical tools to analyse collected data and test predetermined hypotheses about the relationship between study variables [140]. Quantitative research is independent of the researcher because it leverages standardised methods and techniques to collect, analyse, and interpret relationships between variables.

On the other hand, a qualitative approach is adopted in studies focusing on explaining a phenomenon or understanding a specific research question or developing theories to improve understanding of a specific topic or concept [32]. A qualitative study leverages interpretive and naturalistic approaches in explaining the topic or phenomenon in question in their natural context to make sense according to people's meanings [44]. Therefore, a researcher is key to the outcome of a qualitative study, especially in collecting and interpreting data. Qualitative studies utilise unstructured interviews, semi-structured interviews, open-ended interviews, document analysis, observation, and participant observation data collection methods to enhance understanding.

This thesis employs a qualitative strategy to explore the adoption of gamification elements in the agent-based model (ABM) through interviews (in Chapter 8) and a quantitative approach using machine learning predictions and simulation of classroom interactions (in Chapters 4, 5, 7 [4, 6]).

The qualitative methodology allowed for the use of observation and semi-structured interviews to collect teachers' views on the potential adoption of gamification elements in the simulation model that is designed to help them mitigate disruptive behaviours in the classroom. On the other hand, quantitative methods enabled the measurement of simulation's efficacy by analysing results from the simulation model. Statistical analyses quantified the impact of disruptive behaviour to assess the model's outcome. Thus, this research follows both quantitative and qualitative strategies in meeting the objectives of every stage designed to conduct the research. Figure 1 illustrates the overall stages undertaken in this research, with each stage detailed as follows:

1. Development of the Agent Based Model (ABM):

Simulation of the interactions between students, teachers, and peers to explore the effects of disruptive behaviour and teacher interventions. ABM is utilized for the simulation as it allows for the exploration of complex social dynamics and testing of interventions in a controlled, risk-free environment.

2. Machine Learning Predictions:

In this stage, ML was used to predict academic performance based on disruptive behaviour features (e.g., inattentiveness, hyperactivity, impulsiveness). ML complements ABM by enhancing the model's predictive capabilities; therefore, it was introduced to enhance the performance of the ABM.

3. Explainability (XAI):

XAI can offer insights to the ML performance. Thus, SHAP (SHapley Additive exPlanations) values is used to make ML models interpretable. XAI may not be interpretable by teachers; however, the added explanations can provide transparency to ML models to enrich the findings contributing to the body of research.

4. Hybrid ML-ABM Approach

ML is integrated with an ABM model to improve the performance of the model. This approach utilises the predictive power of ML in predicting learning outcomes with disruptive behaviour and employs ABM for simulating classroom interactions.

5. Integration of Gamification

Gamification has been shown to increase motivation and behaviour engagement. Therefore, gamification elements were explored in an online educational system named CameleOn. Gamification was then introduced to the simulation model for increased behaviour engagement. Teachers' feedback was collected about gamification to improve the model, aiming to make it more engaging and useful for classroom management.

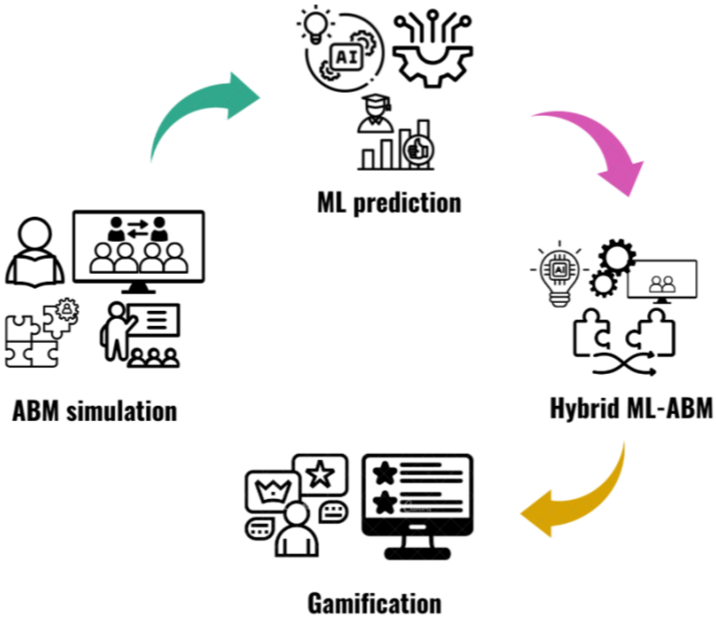

Figure 1 Overall research stages.

3.4 Data

The research in this thesis incorporates two datasets, PIPS and CameleOn, each serving unique purposes for this study:

1. PIPS Dataset: The PIPS (Performance Indicators in Primary Schools) dataset provides valuable data for simulating classroom interactions and disruptive student behaviours within the ABM. This dataset includes both academic and behavioural metrics, such as inattentiveness, hyperactivity, and impulsiveness. By using this data, the ABM can evaluate the effects of different levels of disruptive behaviour on academic performance.
2. CameleOn Dataset: The CameleOn dataset provides data on student engagement in a digital environment through gamification elements like badges, levels, and achievement tracking. Although ABM is not intended for student use, gamification features are incorporated into the ABM to increase

motivation and engagement for teachers who will interact with and benefit from the model. This dataset helps examine how gamification could enhance teacher engagement with the ABM, potentially making it a more engaging tool for exploring and managing disruptive behaviour.

Together, these datasets contribute to this research: the PIPS data enables an understanding of the impact of disruptive behaviours on performance, while the CameleOn data facilitates an exploration of gamification as a method to increase engagement for the integration with the ML-ABM approach. The following sections offer a detailed description of each dataset.

3.4.1 CamaleOn Dataset

The CamaleOn dataset was used for answering research question RQ 4. The work with this dataset is further depicted in Chapter 7 and published in [4]. CamaleOn[1] is a Brazilian Gamified Intelligent Tutoring System. Officially launched in 2012, its aim is to increase the accessibility of educational resources to Brazilian students to increase their chances of passing the Vestibular Exam. Students must compete with others from different schools in Brazil by scoring higher in this exam. It is held in three consecutive days where students with the highest scores are the ones who go through to the next day. In the last day of exam, the students would get accepted into the university if they had a higher score than the rest.

The Vestibular Exam consists of multiple-choice questions that are based on high school curricula in different topics such as mathematics, physics and Portuguese Literature. CameleOn allows students to choose a subject where they would like to improve and start learning by viewing the material and answering questions.

[1] https://plataformacamaleon.com.br/

Every subject has a multi-level map where students can move from one level to the next with higher difficulty. At each level, there are optional helping materials that student can view to provide extra help and understanding of the level's topic.

Data that I used for the research within this thesis (Chapter 7 [4]), collected from CamaleOn represents 8270 students, a sample size much above the required statistically applicable one for the student population of Brazil (for confidence level 95%, confidence interval ±5%, sample size calculator from Surveysystem.com, for the population of Brazil at 211 million people a minimum of 384 people is needed). Students solved 307814 problems, watched 1131 videos, received 236345 badges, and logged in 67752 times. Data was collected on their behaviour (Logs) to build a Student Model [168]. Behaviour reflects interaction data between the students and the various elements of their online learning environment, such as problems, resources, etc.

3.4.2 PIPS Dataset

The PIPS Dataset was used for the ABM simulation and training in Chapters 4, 5, 6 . This dataset came from schools that participated in the Performance Indicators in Primary Schools (PIPS) monitoring system [197] [198][2], in which young students were assessed at the start of their first year in elementary school and again at the end of that year. PIPS was developed by researchers at the Centre for Evaluation and Monitoring (CEM) in Durham University, the UK, during the time this data was collected. CEM is one of the largest providers for children assessment and a part of the Cambridge family. Consent for data collected by CEM is detailed in the official website (www.cem.dur.ac.uk). Local authorities and schools participating in PIPS pay to use it. It provides assessments that are administered on pupils then returned to CEM

[2] RR344_-_Performance_Indicators_in_Primary_Schools.pdf (publishing.service.gov.uk)

to be analysed for the purpose of monitoring the progress of pupils. Feedback that contains detailed information about pupils' academic achievement is then returned to schools. As a large number of schools participate in PIPS, the data it processes is of different variations that allow for research to be conducted on some of it. The research in this thesis is using data collected by PIPS in the academic year 2007/2008 as well as the following year 2008/2009 that was approved to be used for the purpose of the research for this thesis by the data provider (usage as described in Chapters 4, 5, 6 and [6, 7]. PIPS also recorded students' non-academic attributes that represent demographic variables, such as *gender*, eligibility for Free School Meals (*FSM*) as well as the *IDACI* score (Income Deprivation Affecting Children Index)[3]. FSM is a binary variable with a value of 1, which means that the student is eligible for a school meal and 0 otherwise. *IDACI* rank takes a range of 1 to 32462, and is defined by the Ministry of Housing as "a subset of the Income Deprivation Domain which measures the proportion of the population in an area experiencing deprivation relating to low income"[4]. This is a vast dataset, comprehensive is collected from 3,315 classes from 2,040 schools from all over the UK, with an average of 26 students per class. Another PIPS dataset from year 2008 collected in Australia and Scotland has over 11,000 records that will be later referred to as PIPS2008. In this chapter, the UK PIPS2007 will be used as it is larger. The PIPS2007 contains a total of 73,372 students' records 33,269 of which are male and 31,4434 are female while 8,669 of the remaining records are not specified. A total of 7,831 receives a school meal and 44,088 does not. Figure 1 shows the general growth in math and reading for all students.

[3] https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019
[4] https://opendatacommunities.org/def/concept/general-concepts/imd/idaci

Trimming outlier with standards deviation threshold = 3 [112], the resulted records are over 39,000 of students that include, beside the above variables, scores on the Initial and End-of-year assessments of Math and Reading, as well as the processed data:

- *Start Math score* is the initial math score of the student from the baseline assessment (0-63 range);
- *Start read(ing) score* is the initial reading score of the student from the baseline assessment (0-169 range);
- *End Math score* is the final score in math from end year assessment (0-69 range);
- *End Read (ing) score* is the final score in reading from end-year assessment (0-178 range).
- *Gender*: Boolean value 0 for male and 1 for female.
- *Student ID*: a unique number that distinguishes each student in the dataset.
- *Growth*: a value calculated from Start and End Math scores. A demonstration is shown in Figure 2.

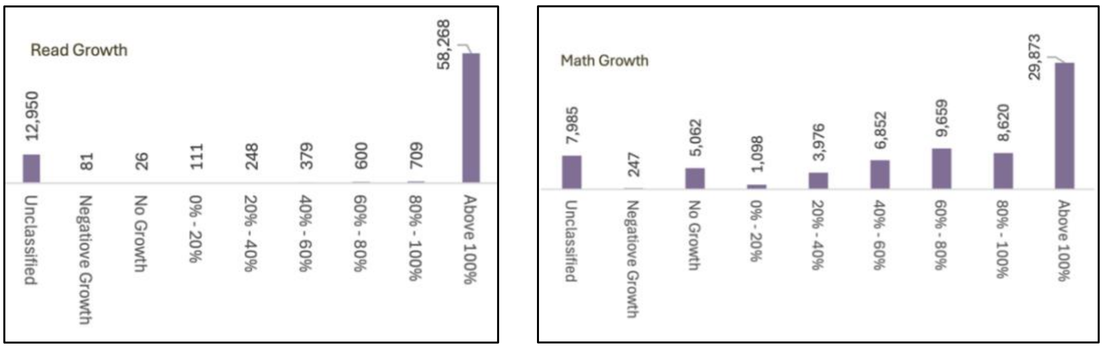

Figure 2 Growth of Read (left) and Math (Right)

The growth in math and reading is calculated as follows:

(3. 1)

For disruptiveness scores (calculated as shown in section 3.4.3), the median is used to indicate the level of disruptiveness if the score is larger than the mean, the student is considered highly disruptive and if the score equals the median, the student is moderately disruptive; otherwise is considered less disruptive.

Figure 3 shows the disruptive level of the students in the dataset for all disruptive features: Inattentiveness, Hyperactivity, Impulsiveness.

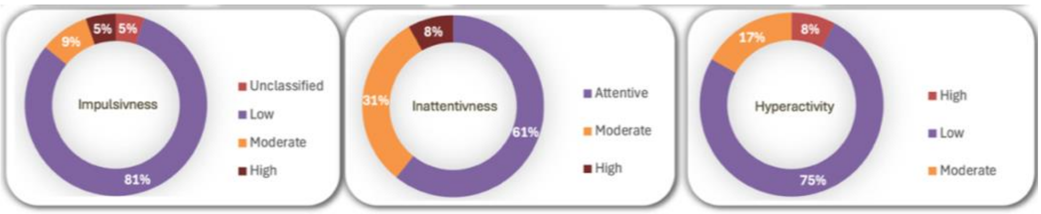

Figure 3 Disruptive level of all features: Inattentiveness, Hyperactivity, Impulsiveness

Figure 4 and Figure 5 show the level of disruptive features based on the percentage for male and female students respectively.

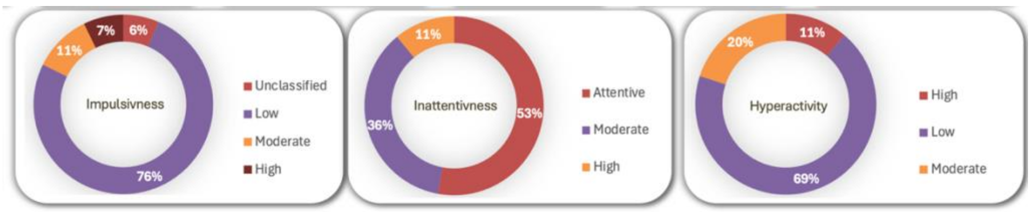

Figure 4 Level of disruptive features of male student percentage

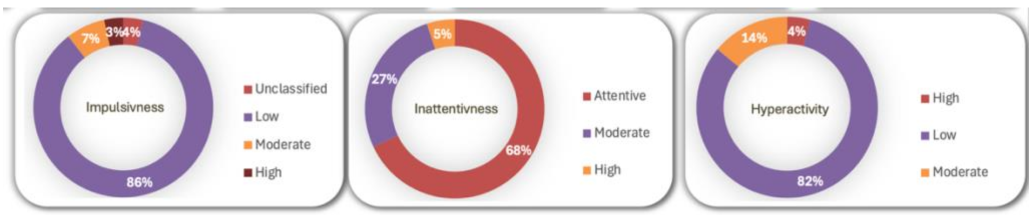

Figure 5 Level of disruptive features of female student percentage

We see from the last two figures that female students show lower disruptive behaviour in all features with 68% attentive females compared to 53% attentive males and 82% low hyperactivity score of female students compared to 69% for male students. Lastly, female students have 86% of low impulsiveness score while male students have 76%. Male students have a higher percentage of a highly disruptive score in hyperactivity, impulsiveness and inattentiveness with 11%, 11%, and 7% respectively.

Among disruptive behaviour features, students with low disruptive feature also have a lower score in the other two as shown for Inattentiveness in the figure where records of Hyperactivity and Impulsiveness show a lower score of Inattentiveness.

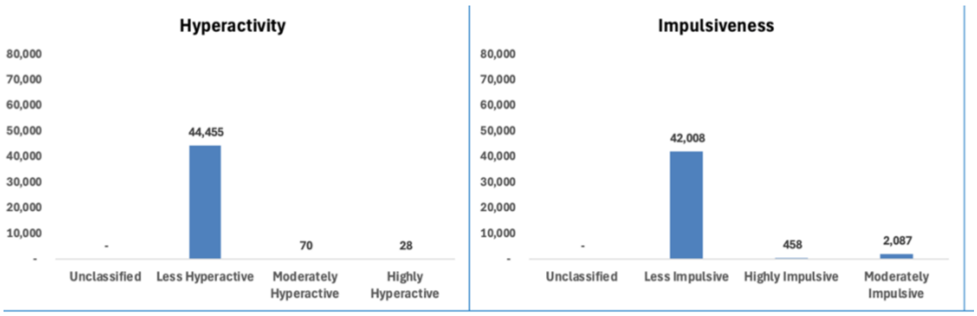

Figure 6 Hyperactivity and Impulsiveness frequency for students with Low hyperactivity

According to Figure 6, the majority of records fall in the low Hyperactivity and low Impulsiveness class with 44,455 and 42,008 records, respectively.

3.4.3 Evaluation Metrics

After explaining the datasets used for answering the research questions of this thesis, the next step is to present the methods and evaluation metrics applied on the findings of this thesis.

Disruptive Behaviour and Academic Performance Metrics: Disruptive behaviours (inattentiveness, hyperactivity, impulsiveness) were measured using teacher-reported scales. This is part of the PIPS data used in this thesis, and has been collected and processed prior to the work in this thesis. The assessment provided by PIPS focuses on Mathematics and Reading. In this assessment, students are shown pictures of people performing some actions and then the students are asked to identify these actions then move to identifying letters then reading commonly used words. Maths is assessed by identifying numbers first, followed by counting. Also, sums are assessed and questions become harder as the students answer each question. Once the student starts having hard time answering, the assessment is stopped. Each student is assessed individually with their teacher, with each assessment taking about 20 minutes. This assessment is repeated at the end of the year with an extended number of questions. The assessment's reliability and validity were tested by administrating the assessment by a researcher at CEM following the teacher's, with a random sample of students. The correlation between re-assessment and original assessment reached 0.9 which shows that it is reliable. At the end of the school year, another random sample is taken for reassessment but using Word Recognition and Phonic Skills (WRAPS), a commercially available reading test, and Basic Number Diagnostic Test for math. The correlation between original assessment and commercially available tests was also found high

The assessment process also provided a score, given by the teacher, for symptoms of disruptive behaviour, using an 18-steps behaviour rating scale with every item representing a specific behaviour (i.e., *inattentiveness* items from 0 to 9, *hyperactivity* from 10 to 15 and *impulsiveness* from 16 to 18) [136] for each student at the end of the school year. The higher the number a student scores on the scale of Inattentiveness, for example, the higher the potential for students to be Inattentive. The items on these scales follow the diagnostic criteria for ADHD by the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Association, 1994) [135][136]. The use of teacher-assessed scores for disruptive behaviour is important for this research because teachers interact with students on a daily basis and can provide a more accurate and consistent observation of behaviour over time. This method allows for an assessment that may be more sensitive to the conditions of classroom behaviour and provide valuable insights that standardized tests may overlook [165]. Therefore, teacher-assessed measurement may provide a more relevant measure of disruptive behaviour in the educational environment.

ML Evaluation Metrics: ML algorithms both shallow or deep were applied in different chapters of this thesis, especially in Chapter 5 and Chapter 07. The data explained in the previous sections are labelled data used for training ML models. A proportion of the data is used for training and another different portion is used for testing the model performance. The process of dividing the available data into training and testing set used in this thesis is by K-fold cross validation. This method provides different pairs of indices for train and test sets. Using this process, the model is trained using data that is split into K number of folds for K number of times. On every training iteration, one-fold or part is left out and the model is tested using that part. All models used in this thesis are trained and validated with k-fold cross validation with K= 10. The full description of ML models is presented in Chapter 5 and Chapter 7.

The models are then evaluated using performance evaluation metrics. Typically, classification models are evaluated using F1-score, overall accuracy while regression models are evaluated using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). Other performance evaluation metrics are also used, like Precision and Recall. The mathematical equations of the evaluation metrics are explained as follows:

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{3. 1}$$

$$Precision = \frac{TP}{TP+FP} \tag{3. 2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3. 3}$$

$$F1 - Score = \frac{2TP}{2TP+FP+FN} \tag{3. 4}$$

Where represents True Negatives and is True Positives is False Negatives and is True Positives.

3.5 Educational Contexts

This research draws on three distinct educational contexts in the design of a data-driven ML-ABM simulation. Each context has unique characteristics that could make the findings more comprehensive and applicable across different educational landscapes. Thus, one reason for using different contexts is that of generalisation. The other reason for using a different context is that the area of the exploration in this thesis, that of disruptive learning, affects not just one set of stakeholders, but several (such as learners, teachers, etc.). Finally, the reason for using these specific contexts is also related to convenience sampling: teachers are notoriously busy and have difficulty to work with researchers beside their regular work; working with students and student data is also adding various challenges, some of them of ethical nature, which will be explored in the next section. The contexts thus targeted in this thesis are briefly presented below.

1. Primary Schools in the UK (Chapters 4,56)

This refers to traditional in-person teaching in UK primary schools, specifically focusing on math and reading classes. This context involves structured classroom environments with young learners and direct teacher-student interactions. The UK primary school setting provides insights into classroom interactions, especially how disruptive behaviours such as inattentiveness or hyperactivity or impulsiveness affect learning outcomes in a formal, traditional setting. Therefore, it offers a baseline for understanding disruptions in a typical face-to-face educational environment which was used in answering RQ1, RQ2 and RQ3.

2. Secondary Schools/ MOOCs in Brazil (Chapter 7)

A Brazilian online learning platform offering Massive Open Online Courses (MOOCs) to secondary school students. This context was used to answer RQ4.It focuses on behaviour engagement in a large-scale, learning environment. MOOCs provide valuable data on how gamification elements can influence behaviour engagement in a traceable digital behaviour of learning environment, which can be utilised to explore the effect of gamification elements for incorporation into the ML-ABM simulation system.

3. Teachers in Saudi Arabia (Chapter 8)

Saudi Arabian primary school teachers provide insights into the use of an educational system, the hybrid ML-ABM, tailored for teachers. To answer RQ4, teachers provided insights regarding the integration of gamification elements into ML-ABM simulation system, providing detailed feedback on which features best enhance motivation and engagement.

3.6 Ethical Considerations

The data used in this research project is intended for research purposes only and the necessary usage permits required for this thesis were all obtained. This included permission to use CameleOn and PIPS data[5]. The interviews carried out in this research with the aim of testing and validating the ABM model follow the ethics policy and procedures of computer science research[6] that states the following:

"Experiments that do not affect or use any personal data from participants and that do not raise any other ethical issues (for example testing and validating a piece of software) do not require approval."

3.7 Epilogue

The preliminary objective of this thesis is to help teachers understand the effect of disruptive behaviours on individual and classroom academic performance through the development of a data-driven hybrid ML-ABM approach of classroom interactions. This chapter explains the datasets PIPS and CameleOn that were used in answering the research questions. It covered the simulation ABM prototype designed

[5] CameleOn dataset was received and approved to be used for research from Dr. Armando M Toda. PIPS dataset was received and approved for research from Prof. Peter Tymms.
[6] Ethics Policy and Procedures

for understanding peers and teachers' effect in simulated classroom. The study highlights the incorporation of ML learning outcomes prediction of disruptive behaviour for the integration with ABM in a hybrid approach, along with SHAP values to provide explainable insights into model outputs.

The upcoming chapters address each research question in detail through the design, implementation, and results phases. Chapter 4 focuses on understanding the effects of teacher and peer characteristics on student performance, utilising the ABM simulation model to investigate these dynamics. Chapter 5 explores the impact of disruptive behaviour on academic outcomes, applying ML predictions to explain these relationships in the classroom context. Chapter 6 describes the development of the hybrid ML-ABM model, which is designed to improve the simulation's ability to predict and simulate classroom interactions. Chapter 7 investigates the role of gamification in promoting engagement in an online system, using ML to analyse its influence on user's performance: the student in this case-. Chapter 8 then explores the perspectives of teachers on gamification elements, with insights gathered from interviews to refine the ML-ABM system's usability. The next chapter starts with answers to the research question RQ1.

CHAPTER 4

4 Implementation of Simulation Prototype

4.1 Prologue

Classroom interactions and their impact on student achievement have been a focus in educational research for many years [33, 40, 167]. But the exploring role of disruptive behaviour in these interactions through simulation remains under explored (see section 2.2 Chapter 2). This chapter applies simulation to deal with disruptive behaviours, particularly inattentiveness and hyperactivity and impulsiveness, which are known to weaken students' academic achievements.

The comprehensive research problem of this thesis is exploring how student disruptive behaviours affect both individual and peers' academic outcomes in a classroom setting via simulation. This chapter is grounded on using an Agent- Based Modelling (ABM) perspective. ABM formulated the preliminary methodology by simulating classroom environments, and focusing on the role of disruptive behaviour. Different sources of influence, like peer disruptiveness, teacher quality, and the level of teacher control, are explored through this simulation. In this chapter, I seek to answer the following research question:

RQ1: How can Agent-Based Models be utilised to explore the influence of disruptive students on their peers and the roles of teaching quality, teacher control in a disruptive classroom?

4.2 Approach

I used Agent-Based Modelling (ABM) to create a simulation of the learning process interactions. This is because the target stakeholders for our research question are human stakeholders in education, such as educational researchers, teaching administrators, teachers and, ultimately, students. We need to move beyond prediction in education with data mining and machine learning algorithms where I use available variables to predict outcomes [193] and be able to simulate how changing variables (e.g. the way of teaching a class) would change the outcomes, as well as note the different effects at different points in time (e.g. during a class, at the end of a class, at the end of a given number of classes, etc.).

From a technical point of view, the model was built using Mesa, which is an ABM framework in Python licensed by Apache2 [130]. Mesa provides a browser-based interface to visualise the model, which allows the use of interactive tools while running the model. This is especially useful during this COVID-affected time, when most interaction has moved online. Moreover, as it is coded in Python, it also has access to Python's large analysis tool library, such as SciPy for scientific computing, Pandas for data analysis and Matplotlib for visualisation.

From a visualisation point of view, this particular implementation is simulating a classroom. Here, a classroom is presented as a 5 x 6 grid to satisfy the limit of class size being 30 students per class in the UK [47]. Shown as coloured circles, students start the class session in a random state of either learning, passive, or disruptive. The state becomes a learning state (in green) when the student has a low disruptive behaviour score. It turns into a disruptive state (in red) if the student has a high disruptive behaviour score or the student's Disruptive Tendency score exceeds the threshold (Disruptive Tendency and Disruptive threshold are defined in Section 4.2.1), where 1 tick in the model represents 1 minute. When a student is being disruptive, he or she may affect the state of their neighbours, depending on the neighbours' disruptive score and the level of Teacher Control and Teaching Quality. As previously stated, every student has two disruptive behaviour scores: Inattentiveness and Hyperactivity, ranging from 0 to 9 and 0 to 6, respectively (using the ranges defined by PIPS). These values could in the future be set at the start of a class; for now, our model initialises each randomly. Students also have other attributes that will be explained in section 4.2.1.

A math lesson lasts for 45 minutes (as recommended by the Department for Education and Skills, 2002), where a student will be moving between the three states: passive, learning and disruptive (as defined based on the PIPS data). Figure 7 shows a flow chart of the model I have created to illustrate the change of the student state.

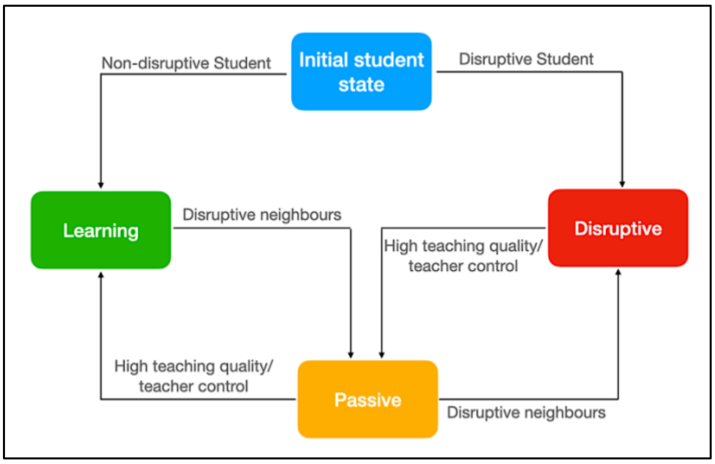

Figure 7 Agent Based Model flow chart

4.2.1 Definition of Variables

The model offers first switch variables that can be manually altered for each run, as described below. These are partially informed by variables recommended by PIPS researchers, and partially self-derived. I discuss implications of choices in section 4.2.2

Inattentiveness and Hyperactivity switch: This variable switch can be tuned to indicate a high or low level of Inattentiveness/Hyperactivity behaviour in a class.

Teaching Quality/Teacher Control switch: This switch varies the quality/control of teaching, ranging from 1 (weak) to 5 (excellent); this scale is arbitrarily defined for this model and has not been taken from PIPS for the purpose of understanding the effect of this variable as a part of the learning environment factors.

Attention Span switch: This variable represents the length of simulation time (ticks) where the student maintains their learning state.

The model also computes a number of derived variables during the simulation runs, defined as follows below.

Initial Disruptive Tendency: Students will be allocated this value based on their Inattentiveness. I propose to compute it using the following formula:

$$DT_{initial}(s,c) = \frac{I(s)-\mu(s,c)}{\sigma(s,c)} \tag{4. 1}$$

Where $I(s)$ is the Inattentiveness score of student s; μ and σ are the mean and standard deviation values of Inattentiveness' scores for class c of student s.

Disruptive Tendency: This variable will change over time - students who are disrupted frequently will be affected and their disruptive tendency will increase. The length of time a student will be in a disruptive or a learning state will be affected by a student's own characteristics, as well as that of the teacher's and peers':

$$DT(s,c,T_{current}) = \left(\frac{D(s,T_{current})}{T_{current}} - \frac{L(s,T_{current})}{T_{current}}\right) + DT_{initial}(s,c) \tag{4. 2}$$

Where $D(s,T_{current})$ represents the number of ticks (minutes) when the student s was in a disruptive state since $T_{current}$, while $L(s,T_{current})$ represents their learning state's ticks. The higher the disruptive tendency becomes, the higher the chance that the student will change to a disruptive state; $T_{current}$ represents the number of ticks passed since the beginning of the school year.

Math attainment level: This variable accounts for individual differences between students; it is derived from their initial score in Math as follows [189]:

$$A(s,c) = \frac{Smath(s,c) - \mu_{smath}}{\sigma_{smath}} \tag{4. 3}$$

Similar as for the disruptive tendency, I use the z-score of student s's initial assessment in the Math subject, Start Math, $Smath(s)$, defined below, because I wish to obtain information on varying from an average value, as opposed to absolute values. μ and σ are the mean and standard deviation values of Start Math scores for class c of student s that is computed before the simulation is initialised either from PIPS data or model generated random data for Start Math variable.

Start Math: This variable can be taken from PIPS or produced randomly by the model for each student. Its range (0-69) corresponds to the PIPS data range. Here, I took the values from PIPS to simulate a realistic environment.

Start Math scaled: As number of ticks the students learn indicate here their final score in Math, I have rescaled the Start Math score to represent minutes of learning:

$$Smath_{scaled}(s) = \left(e^{Smath(s)}\right)^{\frac{1}{n}} \tag{4. 4}$$

$Smath(s)$ is the Start Math score of student s. I use n in the exponent to fit the logarithmic function to map the 'learning Minutes' into 'Score' in a similar manner as the work of [131], who used the logarithmic function to map 'Teacher feedback' into 'Score'. To fit the logarithmic function, I use the total number of minutes the students would possibly have in a school year, which equals to end-time =8550. Since $log\, 8550^{n}$= 69, I calculate n to be ≈ 7.621204857.

End Math: The simulated End Math score is shown in the Equation 4.5 where $L(s, T_{end-time})$ represents the total learning time student s had throughout the simulated year:

$$Emath(s) = log(L(s, T_{end-time}) + Smath_{scaled}(S))^n \tag{4. 5}$$

Disruptive threshold: represents one standard deviation above the mean disruptive tendency of the class [20, 62].

4.2.2 Functionality

ABMs often simulate a multi-agent environment where interactions are complex. In my attempt to simulate a realistic classroom environment, the designed functions of the model were chosen based on resources and arbitrarily defined thresholds. As Railsback & Grimm [161] note, arbitrary thresholds are introduced to manage the complexity of ABMs and simplify agent behaviours in ways that are still meaningful, but more manageable, computationally. Moreover, such thresholds support the exploratory and flexible nature of ABMs, where different scenarios can be explored, and rules are further refined based on observed results.

In my simulation model, some scales were arbitral (e.g., a scale of 1 to 5 for Teaching Quality and Teacher Control). The upper bound of 5 was chosen to represent an excellent teacher, while 1 represented a teacher with poor control or engagement. These values were defined for the purpose of understanding their influence [122].

As per Figure 7, students are to be in a learning state if one of the following occurs:

- Disruptive Tendency is lower than the Disruptive Threshold of class. This threshold was introduced to ensure that minor disruptions do not immediately pull a student out of learning, which mirrors real-life classrooms where small disruptions can be tolerated without severely impacting a student's ability to focus [79].

- Disruptive Behaviour is low, and Teaching Quality or Teacher Control is high [199].
- Current state is passive, and more than half of the neighbours are in a learning state. This creates a positive peer influence that can pull the passive student into the learning state [101, 102].

Students are to be in a passive or disruptive state if one of the following situations occurs:

- Disruptive Tendency is higher than the Disruptive Threshold, passive if Teacher Control or Teaching Quality is high and disruptive if low. This reflects the critical role that effective teaching plays in managing classroom disruptions[56].
- Current state is disruptive, but Teacher Control is high; passive state and disruptive if Teacher Control is low. This rule reflects the teacher's ability to mitigate disruptions by re-engaging students through intervention, classroom management, or instructional methods [110] [199].
- Disruptive Behaviour is low, and Teaching Quality is low.
- Two neighbours are disruptive; passive state, more than two neighbours are disruptive; disruptive state. This illustrate the effect of disruptive peers, where their existence creates a negative environment that makes learning difficult for the surrounding students[108].

- Ticks of learning state exceed the attention span value. This reflects cognitive limits, where students can only focus for a certain amount of time before their attention decreases[160].
- Disruptive Behaviour is high, and previous state is passive; disruptive state.

The agent (defined in Section 2.2.1) in my model represents a student who will remember his/her previous state and choose the next state based on earlier states. For example, if a student falls in a disruptive state for long, they can change to either passive or learning, based on characteristics or statuses of the teacher and neighbours.

The model's simulation visualisation (Figure 8) will display the changes in student states during a minute (tick) in a lesson, with a line graph (below) that updates as the model runs. The graph follows the total number of disruptive students and learning students in every tick of the model. The black line represents the average End Math score of the class, while the red line represents the number of disruptive students. Figure 8 shows a screenshot of running the model with the variable Inattentiveness switched to zero. The grid contains cycles that represent the different states of the students during the simulation.

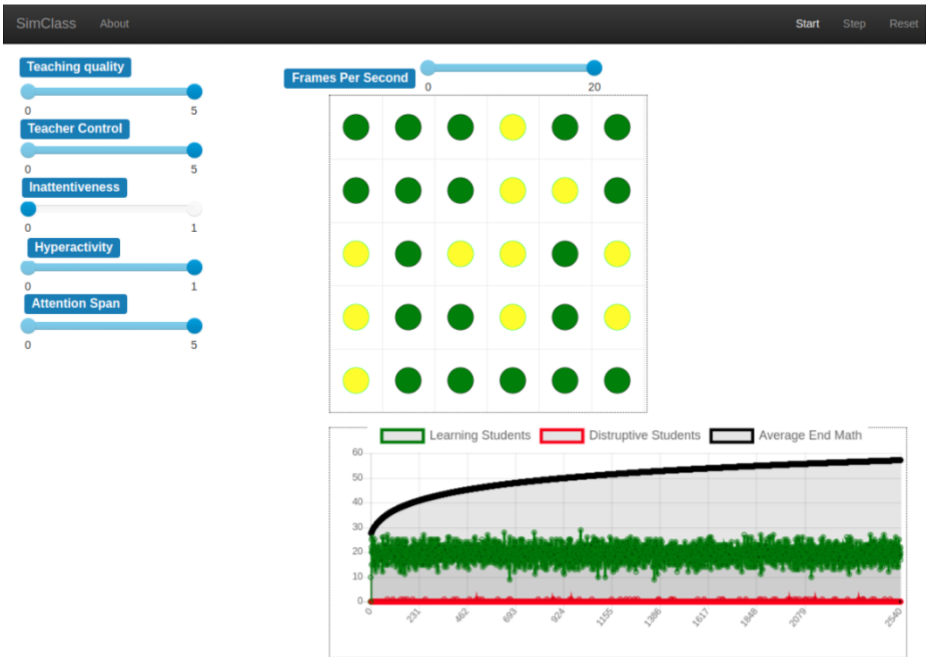

Figure 8 Running the SimClass model

4.2.3 Data Analysis

To answer the first research question, RQ1, and understand the effect of disruptive students on other students (here, the whole class), I explore the relationship between disruptive behaviour and End Math scores from PIPS dataset (here, representing general attainment – see Section 3.4.2). Specifically, I compute this End Math average score in classes with a high number of disruptive students and then compare this with classes with a lower number of disruptive students. I define the (set of) disruptive students as DS⊆S:

$$DS = \{s \in S, where\ ds(s) \geq M\} \quad (4.6)$$

$$M = \{median(ds(s)) |\ s \in S\} \quad (4.7)$$

Where S is the set of all students, s is an individual student, *ds(x)* is the disruptive score function, and M is the median. The median, rather than mean, was chosen to define the threshold, because the data, according to Shapiro's test, is not normally distributed [2]. According to the data from PIPS, Inattentiveness has a median of 5, while Hyperactivity has a median of 3.

Out of 3,315 classes in the dataset, there were 2,337 classes with students categorised as disruptive. To have a deeper look into the data, I calculated the percentage of disruptive students per class and the average of the End Math score for that class and compared the two.

Table ***1*** shows the correlation test results, where it can be seen that the percentage of disruptive students has a higher negative correlation (of -0.16) with the average of End Math, as opposed to -0.04 and -0.06 of their start and end scores, respectively. This suggests an effect of the number of disruptive students in a class over the general attainment - represented by End Math scores - in that class.

Table 1 Correlation test between disruptive behaviour and Math scores

	Start Math	End Math	Average End Math
Inattentiveness	-0.27	-0.33	-0.07
Hyperactivity	-0.14	-0.18	-0.06
Percentage of disruptive students	-0.04	-0.06	-0.16

4.3 Results

Running the simulation model for 8,553 ticks represents a 45-minute math lesson a day for 190 days in a year [117]. I here present 3 runs with different parameter inputs to observe their effect on student End Math scores. Each run was repeated 50 times to ensure stability and consistency of the model [162]. To demonstrate the realism of the ABM, I evaluate how accurately the model represents real-classroom outcomes by adjusting parameters across three distinct scenarios and comparing the simulation results to actual classroom performance metrics. This ensures that the model outputs align with realistic data. Additionally, the fundamental functions and design of the model rules are grounded in findings from the literature, providing a robust theoretical foundation for the simulation as presented in section 4.2.2. Results are shown in Table 2.

Run 1: In the first simulation run, I set all parameters with the maximum value for each (Teaching Quality and Teacher Control = 5, Inattentiveness/Hyperactivity = 1 and Attention span = 5). I chose this setting to be the baseline to allow me to explore the different impact of each parameter in other runs.

Run 2: In this run of the model, I switched off Inattentiveness and kept the rest of the parameters at maximum value in order to understand the effect of Inattentiveness variable over the results when compared with the baseline.

Run 3: Here, I aimed to observe the impact of Teaching Quality; therefore, all parameters had the maximum possible values of their ranges, except Teaching Quality, which was given the lowest possible value from its range, i.e., 1 out of 5.

Table 2 Results of End Math and Disruptive Tendency variables of three runs

	Math		Disruptive Tendency	
	First tick (Start Math)	Last tick (End Math)	First tick	Last tick
Run 1	27.43	43.08	1.16	0.12

Run 2	27.43	66.16	0.73	-0.53
Run 3	27.43	36.45	1.05	-0.07

4.4 Discussion

Three different parameter inputs into the simulation model provided different results. Therefore, I computed Cohen's d to present the effect size between the three runs (see Table 3). An effect size of.2 is considered small,.5 medium and.8 large [40]. It can be seen that the effect size is large between the runs. I used t-test and found the difference between End Math scores of the three runs to be statistically significant.

Table 3 Cohen's d and t test between End Math scores of all runs

	End math (Run 1)	End math (Run 2)	End math (Run 3)
End math (Run 1)	-	1.43 (p = 4.13e-42)	7.81 (p= 6.41e-07)
End math (Run 2)	-	-	9.12 (p = 3.09e-37)
End math (Run 3)	-	-	-

In the case of the third simulation, when Teaching Quality was reduced, the End Math results produced by the model were the lowest, with an average of 36.45, indicating that students made the least progress in maths of all runs. This means that Teaching Quality as a characteristic of the teacher influenced the attainment of the class by the end of the year. Additionally, it can be seen that students had also the highest disruptive tendency in this run. Please note that I have used the term "Teaching Quality", to represent the level of teacher's ability to maintain student focus and minimise disruptive behaviour during a lesson [111]. Teachers who employ interactive and inclusive strategies enhance students' motivation and involvement in the learning process [52], which strongly connect it to class engagement [145]. In contrast, the highest average of End Math scores was seen in the second run, when the Inattentiveness switch was off, resulting in 66.16 for the average End Math score, which presents an answer to Run 2 showing a negative effect of disruptive students in a class over their attainment. An average of 43.08 falls in between the previous two in the baseline run, when all variables used in the model had the maximum value allocated for each range.

To compare with the real-world PIPS data[7], I ran a Pearson correlation test for the three different simulation runs and the results are presented in Table 4.

Table 4 Correlation test between simulation runs results and model variables (8,553 ticks)

	End Math (Run 1)	End Math (Run 2)	End Math (Run 3)	End Math (PIPS)
Start Math	0.71	0.74	0.66	0.70
Inattentiveness	-0.31	-0.09	-0.38	-0.34
Hyperactivity	-0.13	-0.11	-0.12	-0.18

It can be seen from Table 4 that the correlation results of the three runs are close to End and Start Math of PIPS data, which was (computed separately to be) 0.70. The nearest correlation score to PIPS data can be seen in the first run, with 0.71, where all parameters had the maximum values possible. These results can be used for finding

[7] Please note however that PIPS data is only available for Start Math and End Math, thus only the start and end of the simulation process. (incomplete meaning)

the best adjustment of the model such as adding random elements of learning, changing ticks' representation and adjusting neighbours' affect.

Next, I consider my various parameters used in more details. I have used here inattentiveness as a disruptive feature, for instance to reflect a state where the mind is wandering, but that is not necessarily always the case. Inattentiveness can be passive if it refers to daydreaming. Mind wandering is an unintentional shift of attention from a task to unrelated thoughts, while daydreaming is a deliberate shift in attention from the external environment, to internal thoughts [178]. Unlike daydreaming, mind wandering is less goal-directed and more likely to involve random, task-irrelevant thoughts [134, 178]. Mind wandering can be disruptive if it occurs during tasks requiring concentration, such as learning or problem-solving [177]. Research indicates that mind wandering is common during learning, and can negatively affect comprehension and retention [139].

However, I do not have a direct measure of disruption, thus anything in the model is a proxy. Follow-up work might also look into the relation between disruptive tendency as a starting point on the road to impacting on personality.

I have here simulated, analysed and compared results at classroom level and compared averages. I showed the link between pupil disruption and math attainment for pupils and for classes, i.e. at two levels. This naturally leads to multi-level models for future simulations.

Beside the 3 runs presented here, I have run the simulation with various other parameters. More structured experiments are planned to run many models with slight variations from one to another, gradually moving toward each of the extremes represented here as Run 1, Run 2, Run 3, and graph the results.

A related issue, which would be addressed by multiple runs, is the stability of the models: how much variation is there when the parameters hardly change? Start Math scaled, introduced here, is currently rather deterministic – if we know how much time has been devoted to maths, we will know the score in maths. Yet, children's maths scores rise and flatten and rise again and stagnate in unexpected ways. Future work could contain an element of randomness to note if results change significantly. Next, I explore the effect of teacher and peers on student performance in different runs.

4.5 Observing Teacher and Peers' Effect

In this section, I take into consideration the level of teacher control as an added influence on pupil state transitions. Specifically, I aim to answer Teacher Control and peers-related part of the first research question:

RQ1: How can Agent-Based Models be utilised to explore the influence of disruptive students on their peers and the roles of teaching quality, teacher control in a disruptive classroom?

In the previous section, I have presented how I have created a simulation of the learning process interactions using Agent Based Modelling (ABM). In this simulation, I present a classroom with 30 pupils where a pupil will change between three different states: learning, passive or disruptive. Functionality of this model and technical details follow the ones in the previous section 4.2.2. The model offers first switch variables, disruptive behaviour and teacher characteristics - switches that indicate a high or low level of pupils' disruptiveness and teacher characteristics [7]. Another switch was added for this work to explore the effect of disruptive pupils in close proximity [27], Neighbours' Effect Threshold switch as it reflects to which degree a pupil affects his neighbours, with a range of 1 (high) to 4 (low). The effect is high if one pupil is enough to change a neighbour's state and low if it takes 4 pupils to trigger an effect. Other variables are math attainment level , which accounts for student learning differences, Start Math , which can be taken from PIPS data or assigned randomly by the model. I use a logarithmic function to map the 'learning Minutes' into 'Score' [131] to compute the End Math variable, *Emath(s,c),* computed as follows:

$$Emath(s,c) = log(L(s,c,T_{end-time}) + Smath_{scaled}(s,c))^n + A(s,c) \quad (4.8)$$

Where $L(s,c,T_{end-time})$ represents the total learning time until the last tick $T_{end-time}$ that student s from class c had during the simulated year. I present here 3 runs with different parameter inputs, to observe their different effect on the pupil End Math scores. In the previous section 4.2, I presented the results of three parameters: all maximum values, low Inattentiveness and low Teacher Quality. In this section, instead, I have examined the following parameters:

Neighbours' Effect Run: In the first simulation run, I am exploring the effect of another pupil's characteristic: Neighbours' Effect. I set this variable to one (out of its range 1 to 4) to understand the impact of very high neighbour's effect[107], when compared with other runs.

Hyperactivity Run: Here, I switched off Hyperactivity and kept the rest of the parameters at maximum value to understand the no-Hyperactivity Effect.

Teacher Control Run: Here, all parameters had the maximum possible values of their ranges, except Teacher Control, which was given the lowest possible value of its range, i.e., 1 out of 5: to explore no-Teacher Control Effect.

Results and Discussion

As an initial step to answer the Teacher Control and peer's- related part of RQ1, I explored the relationship between disruptive behaviour and End Math scores to understand the effect of disruptive pupils on other pupils and found a negative correlation between the percentage of disruptive students and average End Math score of the class [7]. This suggested an effect of the number of disruptive pupils in a class over the general attainment. I computed Cohen's d for the three runs and found the effect size to be is large or medium [40]. Table 5 shows the results of the average End Math score for the runs.

Table 5 Results of average End Math of three runs

Run	First tick (Start Math)	Last tick (End Math)
Neighbour's Effect	27.43	28.71
Hyperactivity	27.43	64.32
Teacher Control	27.43	30.60

Thus, when the Neighbours' Effect increased, the End Math results produced by the model were the lowest, with an average of 28.71, indicating that pupils made the least progress in maths of all runs which shows peers' disruptiveness over pupils' attainment. In contrast, the highest result was seen when the Hyperactivity switch was off, resulting in 64.32 for the average End Math score, and an average of 30.60 in the low Teacher Control run which provides an answer to RQ1 by showing a positive effect of low disruptive pupils in a class and a negative effect of low Teacher Control over their attainment. To compare with the real-world PIPS data[8], I ran a Pearson correlation test for the three simulation runs (see Table 6).

[8] Please note however that PIPS data is only available for Start Math and End Math, thus only the start and end of the simulation process.

Table 6 Correlation test between simulation runs results and model variables

	End Math (Neighbour's Effect Run)	End Math (Hyperactivity Run)	End Math (Teacher Control Run)	End Math (PIPS)
Start Math	0.98	0.40	0.69	0.70
Inattentiveness	-0.14	-0.17	-0.06	-0.34
Hyperactivity	-0.16	-0.19	-0.17	-0.18

The nearest correlation score to PIPS data can be seen in the third run, with 0.69. A high correlation is seen in the first run with the highest degree of Neighbour Effect, due to low progress resulting in little difference of pupils between End Math and Math score. These results can serve for further improving the use of the model by providing the simulation of several factors in the learning environment.

4.6 Epilogue

This chapter has presented an ABM model design to understand the effect of disruptive young students in a classroom environment using the PIPS data. The model simulates the interactions for one school year. The results show an increase in average End Math scores when the Inattentiveness variable is reduced, which confirms the effect of disruptiveness in a class over attainment, conforming to the PIPS data. In contrast, a decrease in the average End Math scores was seen when the Teaching Quality and Teacher Control was reduced, showing the effect of teacher characteristics over students' attainment in an answer to RQ1 of this thesis. The model was created using a user-friendly front, which allows users to adjust the model easily to find the best way of applying pedagogical strategies. I also presented how I improved the design of the ABM model to reflect the effect of disruptive young pupils in a classroom environment over their neighbours, supported via an experimentation with these parameters. The findings presented provide a positive causality relationship between decreased disruptiveness and increased performance, while high levels of disruption and low teacher control contribute to the low achievement. However, a limitation of this model is that socio-emotional variables that also define students in class and source of disruption are not factored in the model. The model can be helpful for analyse the interactions occurring within classroom. Future developments of the current model may include addressing the different facets of the students' characteristics, as well as improving model accuracy to increase the usability of the model. In the next chapter, I provide a response to RQ2, via the application of ML to extend the analysis of disruptive behaviour on classroom performance. Also, the use of XAI explains the ML predictions' outcomes.

CHAPTER 5

5 Exploring the Impact of Disruptive Behaviour on Student Performance

5.1 Prologue

Disruptive behaviour, such as Inattention, Hyperactivity, and Impulsivity, can significantly impact student achievement in education. Not only does it affect the disruptive student, but it can also undermine the performance of the other classmates, and lead to varied educational outcomes among the different groups. While significant in importance, education literature on the predictive aspects of different disruptive behaviours (inattentiveness, hyperactivity, impulsiveness) was scarce at the time of this research (see Chapter 2; section 2.3), particularly, interpretable models that provide educational stakeholders with opportunities to identify and mitigate their effects.

This chapter further develops the analysis built on top of the ABM framework from the previous chapter (Chapter 4). The previous chapter simulated classroom interactions in a school year. This chapter takes a crucial first step in improving the current ABM by examining the predictive capacity of ML in predicting the impact of disruptive behaviours on student performance. The primary goals of this chapter are twofold: (1) I aim to create a predictive ML model that predicts how different behaviours (inattentiveness, hyperactivity, impulsiveness) impact learning outcomes, and (2) build a transparent interpretative framework, using SHAP. This chapter uses robust algorithms, such as XGBoost, Gradient Boosting, AdaBoost, Random Forest, MLP etc., to achieve high predictive classification accuracy. Also, the interpretability of SHAP gives us insights into the contribution of each disruptive feature to the academic performance, so that the educators can take some action. Consequently, this ML exploration is a foundational step, exploring ML capabilities in order to establish an improved ABM that incorporates the predictive power of ML into its framework. Through this predictive model, this chapter addresses RQ2, regarding how disruptive traits relate to learning outcomes and how XAI reveals those relations. This study provides a basis for using interpretable model predictions, with both the interpretability of XAI providing the explanatory power of disruptiveness in classroom, and enabling educators and stakeholders to profile and support students in the management of their disruptiveness.

This chapter is a critical step in the methodology of the thesis, moving from simulation to predictive modelling. This chapter is based on my paper [4]. Its main aim is to find the features of disruptive behaviour that can predict student performance, exploring, for the first time, the effect of three features, Inattentiveness, Hyperactivity and Impulsiveness, over student performance. The research question that will be fully answered in this chapter is:

RQ2: How can we predict and explore students' learning outcomes based on disruption-related features (Inattentiveness, Hyperactivity, Impulsiveness), using ML models and Explainable Artificial Intelligence (XAI)?

5.2 Approach

To answer the second research question, RQ2, in predicting student performance from disruptiveness-related features, I first simplify the prediction problem from continuous values into a classification problem, and then find the best prediction algorithm. To accomplish this, I use the *Jenks' natural breaks* method [93] a well-known method used in similar research [141, 152]. This method is based on 'natural breaks' that classify data, based on breaks that naturally exist in the data. I use this method because it maximises variance between groups and minimise it within the group. It requires a prior setting of the number of breaks *n* (setting explained in section 5.4). The breaks are used to label the data for classification followed by data grouping by the same method to improve classification consequently, as shown in Figure 9. To address the issue of class imbalance between the two classification classes, I opt to use an under-sampling method called Tomek Links [194], since I have a higher number of records in one class compared to the other. To answer the research question RQ2, I focus on students' learning outcome, by examining their improvement, not simply their final achievement score. I propose to calculate the improvement variable as follows:

$$S_{imp} = S_{e\_math} - S_{s\_math} \tag{5. 1}$$

Where S_{s_math}, S_{e_math} represents students' Start- and End Math scores, respectively. As student improvement is affected by the environment, including the school they attend and their teacher, I introduce this variable to use it for student grouping in the data with the Jenkspy method to split the feature space along the available data features (disruptive behaviour, performance) to optimise the information gain of the classifier models and thus their resulting performance.

I then experiment with state-of-the-art classifiers: XGBoost classifier, Gradient Boosting, Ada Boost, Random Forest, Extra Trees, Logistic Regression, KNN, and MLP for their excellence in student performance prediction, as mentioned in section 2.3.1, Chapter 2. Another reason is that these classifiers still outperform NNs for tabular data [77], as used in this thesis. The Kruskal test is then computed between the prediction of the best performing classifier and the others, for establishing the statistical significance of the difference between their F1-scores.

Then, to answer the second part of RQ2, I explain the classification outcomes, using SHAP. SHAP stands for SHapley Additive exPlanations and is a method proposed by Lundberg and Lee [118], to explain the prediction results of supervised machine learning models using feature importance values called SHAP values. According to Lundberg and Lee, a SHAP explanation has a stronger agreement with human intuition than other explanation methods, such as LIME (Local Interpretable Model-agnostic Explanations) [164] and DeepLIFT (Deep Learning Important FeaTures) [175]. I select SHAP as it provides the local explanation that LIME provides, as well as a global explanation, i.e., the explanation for the entire model [119]. I am, *for the first time, applying SHAP to observe improvement and understand the relationship between each disruptiveness feature and student performance* as well as other features in the prediction model for different categories of students for deeper explanation. To do so, I use the SHAP instance-level interpretation as building blocks for group interpretation. Figure 9 shows an overview of the process used in answering RQ2.

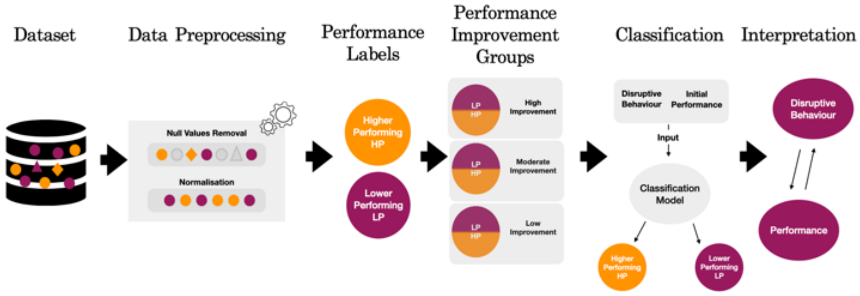

Figure 9 Overview of prediction and explanation methodology steps

5.3 Data

The source of data used for the work in this section of the chapter is found in Chapter 3, section 3.4.2. Utilising feature engineering- a technique for improving machine learning (ML) model performance [53]- additional proposed processed features derived from the PIPS dataset are described as follows:

- *countFSM*: calculated variable that represents the number of students who receive a Free School Meal in the classroom.
- *percentageFSM*: calculated variable that represents the percentage of students who receive a Free School Meal in the classroom that I compute from existing features by finding the percentage of students who receives a free school meal in each class c in our defined formula as follows:

$$percentageFSM_c = \frac{\sum_{s\in C} FSM_s}{\sum_{s\in C} 1} \tag{5. 2}$$

Where FSM_s is a Boolean determining if student *s* receives a free school meal and $\sum_{s\in c} 1$ is the total number of students in class *C*. This variable was introduced to observe the effect of class level variables as well as student level variables.

In total, 10 variables are used as input features for classification task: Start Math, Start Read, Inattentiveness, Hyperactivity, Impulsiveness, countFMS, percentageFMS, as well as individual characteristics: gender, FSM, IDACI, which are defined in Chapter 3.

5.4 Results

Using the Elbow method [191], the optimal number for breaking the data is found to fall between *n*=2 and *n*=3. Applying the Jenks natural breaks method, I tested first with *n=3* for the number of breaks, but as this resulted in a low performance of the classifiers, was reduced, in a first instance, the number of breaks to *n=2*. When analysing these groups, I found them to have distinct features and differences in performance, as well as being of a similar size. This led to a natural division into the *high performance* (HP) and *low performance* (LP) groups. HP has a higher mean for both *Start Math* and *End Math*, while LP has the opposite. LP, however, has a higher mean of *inattentiveness*, *hyperactivity*, and *impulsiveness*. Statistical difference between the resulting two classification groups was tested using the Kruskal test [133]. as the groups were not normally distributed – according to the Shapiro test, I found them to be statistically different ($p<<0.05$) between all features. To solve the issue of performing multiple comparisons, I used the Bonferroni correction for the 8 comparisons [173]; nevertheless, the p-value remained <0.05 after the correction.

Table 7 displays the full statistical description of the two groups normalised with the *z-score* method (by scaling the features to have a mean of 0 and standard deviation of 1).

Table 7. PIPS variables and values for high (HP) performers versus low performers (LP) (z-scores).

	Low performing (LP)		High performing (HP)	
	Min	Max	Min	Max
Count	16681		22109	
Start Math	-2.12	3.03	-3.04	3.07
End Math	-3.14	1.24	-1.48	4.45
Start read	-2.56	3.33	-2.59	3.02
End Read	-2.60	3.13	-2.97	2.29
Inattentiveness	-0.84	2.73	-0.50	3.84
Hyperactivity	-055	3.90	-0.40	3.56
Impulsiveness	-049	2.95	-0.35	4.45
FSM	-0.49	2.03	-0.30	3.28
IDACI	-1.46	2.06	-1.85	1.66

Looking at the statistics of both groups, it can be said that group HP displays lower disruptive behaviour and better learning outcomes than group LP. The distributions of disruptive behaviour features between Group HP and LP are further analysed in Figure 10.

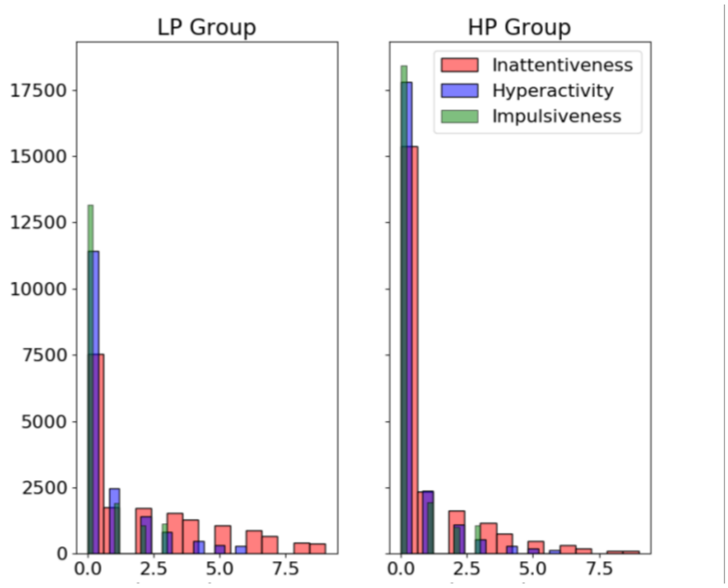

Figure 10. Inattentiveness, Hyperactivity, Impulsiveness score (x-axis) and frequency (y-axis) of HP and LP groups.

Figure 10 shows that students in group HP had a higher frequency of scores larger than zero in both Hyperactivity and Inattentiveness, which indicates that students in that group had higher scores in those two features. Impulsiveness has the lowest values in both groups, close to zero; and Hyperactivity comes second; while Inattentiveness has the highest values for both HP and LP groups.

The next step is to use a machine learning algorithm to classify the students into these groups, based on their learning outcome variable, End Math. First, I attempted to improve the prediction accuracy by adding more features as input variables [96]. I propose here to use the defined poverty variable which correlates negatively with the classification variable (-0.17; which is higher than the correlation coefficient of FSM and IDACI separately, for which the correlation is ±0.15). Table 8 shows the results of classification algorithms.

Table 8 Performance of disruptiveness-related classification models (10 features)

Classifier	Acc	Precision	Recall	F1-score	P-value
XGBoost	0.79	0.79	0.75	0.77	-
Gradient Boosting	0.79	0.79	0.75	0.77	0.69
Ada Boost	0.79	0.79	0.75	0.77	0.85
Random Forest	0.77	0.74	0.77	0.76	0.93
Extra Trees	0.77	0.74	0.77	0.76	0.79
Logistic Regression	0.77	0.78	0.71	0.74	0.14
KNNeighbors	0.76	0.76	0.72	0.74	0.41
MLP	0.71	0.66	0.74	0.69	0.42

It can be observed from Table 8 a similar performance of XGBoost, Gradient Boosting and Ada Boost. The p-value shown in Table 8 shows no statistically significant difference between any of the F1-scores and the best in class.

Generally, ML models performed significantly better than NNs models. The higher F1-score model MLP was included and excluded CNN, which achieved an F1-score below 0.60.

To further to improve the performance, for its positive impact on later model interpretation (where an accurate model allows for a better interpretation) [114], I used hyperparameter tuning with grid search (GridSearchCV from Scikit learn [155]). The algorithm which performed best after tuning the parameters of all classifiers was XGBoost, as its result showed an increase in F1-score that reached 86% and Gradient Boosting as second best with an F1-score of 0.85%.

Thus, analysing this data, I chose to continue with XGBoost for the following SHAP explanation, as it had the advantage of performing well. I then used natural breaks once more, this time, based on the calculated variable , resulting in at n=3. The resulted break characteristics in terms of specific improvement, based on our proposed variable, s_imp, are shown as a box plot in Figure 11 and the PIPS variables and values (Z-scores) of the resulting groups are listed in Table 9,10, 11

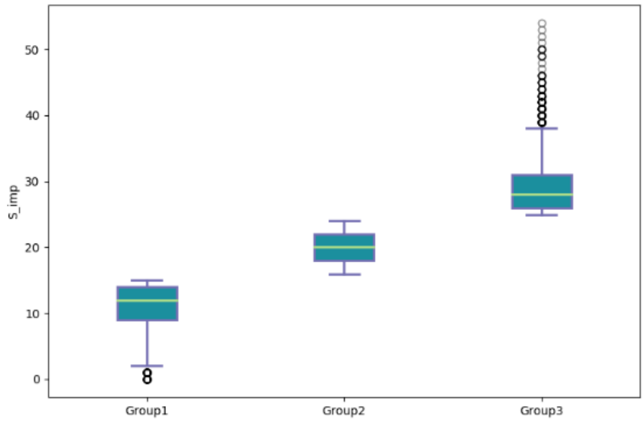

Figure 11. Improvement box plots for the three groups

Table 9 PIPS variables and Values of Group 1

	Min	Max
Count	10902	
Start Math	-2.34	3.00
End Math	-3.03	3.03
Start read	-2.02	3.17
End Read	-1.99	3.01
Impulsiveness	-0.76	2.81
Hyperactivity	-0.52	4.02
Impulsiveness	-0.47	3.07
FSM	-0.44	2.24
IDACI	-1.50	1.92

Table 10 PIPS variables and Values of Group 2

	Min	Max
Count	18608	
Start Math	-2.60	2.96
End Math	-3.00	3.25
Start read	-2.19	3.11
End Read	-2.62	2.51
Inattentiveness	-0.46	3.66
Hyperactivity	-0.46	4.83
Impulsiveness	-0.44	3.25
FSM	-0.38	2.59
IDACI	-1.68	1.79

Table 11 PIPS variables and Values of Group 3

	Min	Max
Count	9927	
Start Math	-2.48	2.96
End Math	-3.00	2.84
Start read	-2.24	3.06
End Read	-2.69	2.19
Impulsiveness	-0.57	4.21
Hyperactivity	-0.44	5.14
Impulsiveness	-0.43	3.31
FSM	-0.37	2.68
IDACI	-1.71	1.78

It can be seen from Table 9, 10, 11 that each group has different characteristics, with Group 1 exceeding (slightly, but statistically differently $p<<0.05$) in Start math compared to Group 2 and 3; but having the lowest mean End Math and End Read, thus being the least improved group. Group 3 has the lowest start scores in math and read but highest mean End Math score, which makes it the most improved group. As for Group 2, its values fall in between the first and third group, therefore it becomes the medium improvement group. I followed the analysis with classification of HP and LP for each improvement group – Group 1, Group 2 and Group 3 – separately to improve the classifiers performance, by creating groups with similar features and increasing the information that each feature provides about the predicted classes. The resulting predictions, according to the Shapiro test, are not normally distributed. Therefore, the Kruskal test was conducted, to test the statistical difference between the models' predictions and the XGBoost prediction for the three groups, with results shown in Table 12.

Table 12 Statistical difference in model prediction for three groups

	Group 1		Group 2		Group 3	
Classifier	F1-Score	P-value	F1-Score	P-value	F1-Score	P-value
XGBoost	0.91	1.00	0.92	1.00	0.91	1.00
Gradient Boosting	0.91	0.79	0.92	0.73	0.91	0.33
Ada Boost	0.91	0.81	0.92	0.39	0.91	0.90
Random Forest	0.90	0.001	0.91	0.00	0.90	0.04
Extra Trees	0.90	0.00	0.90	0.00	0.90	0.84
Logistic Regression	0.86	0.00	0.88	0.00	0.89	0.00
KNNeighbors	0.72	0.00	0.61	0.00	0.82	0.00
MLP	0.76	0.00	0.73	0.001	0.87	0.00

In general, it can be seen that the classifiers that have highest performance are the boosting algorithms: Gradient Boosting Extreme Gradient Boosting (XGBoost) and Adaptive Boosting (Ada Boost) that are known to reduce variance [65], followed by the tree-based algorithms. KNNeighbours uses similarity between neighbours, which might not be performing well due to the nonlinearity of the data. As previously mentioned, for deep learning models, like MLP, the nature of the data might be affecting their performance. [77]'s findings uncover the reason for lower performance of NNs compared to shallow ML to be "irregular patterns in the target function, uninformative features, and non-rotationally-invariant data where linear combinations of features misrepresent the information". These results provide an answer to RQ2, as the XGBoost classifier performed well in predicting the learning outcome, using disruptiveness-related features. 10 features were used that contain the *academic and financial factors in addition to disruptive behaviour factors, which have not been used for prediction of student academic performance before*. Hyperparameter tuning with GridSearchCV was used to improve the performance of the classification models using grid search algorithm, as well as tenfold cross-validation, to train the models. I used the grid search algorithm provided by Scikit learn library to find the optimal parameters of XGBoost for our case (base_score=0.5, booster='gbtree', colsample_bylevel=1, colsample_bynode=1, colsample_bytree=1, gamma=0, learning_rate=0.03, max_delta_step=0, max_depth=6, min_child_weight=1, missing=None, n_estimators=1000, n_jobs=1, nthread=None, objective='binary:logistic', random_state=0, reg_alpha=0, reg_lambda=1, scale_pos_weight=1, seed=None, silent=None, subsample=0.9, verbosity=1). My findings fall in line with [13], in predicting performance with SHAP interpretations using features including ADHD-related one (see Chapter 2, Section 2.1.1, for a discussion on disruptive behaviour and its relation to ADHD); however, their F1-score achieved 87% for predicting Arithmetic skill using five algorithms. In the current work, I outperformed their work, by exploring eight algorithms, with F-Measure reaching 92% predicting End Math, using disruptive, financial, and academic related features. Hence, this study is broader, in terms of algorithms used, data size and providing better performance for the predicted classes. I provide a relation between three different features of disruptive behaviour and student improvement, and an explanation of that prediction for a deeper understanding of this relation.

5.5 Interpretation

As stated, the results of the classification model are further interpreted using the SHAP framework. Figure 12 shows a plot of the SHAP values for all features, where we can see the highest to be the Start Maths feature and the lowest, the hyperactivity feature. The rest of the features have a varying degree of influence, according to their SHAP values, but we can see that Inattentiveness has the highest value among features that represent disruptive behaviour and Start Math has the highest importance, followed by Start Read. We can also see that the number of students receiving a school meal (countFSM) has a higher value than gender. This interpretation method allows us to compare the effect of student level features as well as class level features, such as percentage of disruptive students, which is unique to this work.

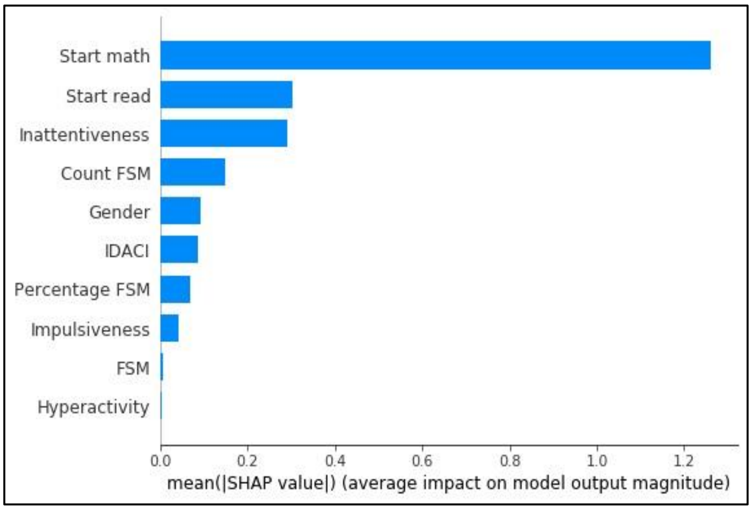

Figure 12. Feature importance as the mean absolute SHAP value

Figure 13 illustrates the summary plot that shows feature effects and feature value as dots. The position of the dots on the right and left of the x-axis represents the positive and negative relationship between the feature and the predicted value: if the dot lies on the right side, it has a positive SHAP value, and if it falls on the left side, it has a negative SHAP value. It shows the direction of the relationship between the feature and predicted value. Positive SHAP values indicate higher performance while negative SHAP values indicate lower performance. The colours represent the instance value from high (red) to low (blue). For example, an instance of a high Start Math score would be represented as a dot in red and on the right side of the plot, because Start Math has a positive SHAP value in our predictive model. We notice that the high scores of Start Math feature have positive SHAP values and the low scores of Start Math have negative SHAP values, while high inattentiveness scores having negative SHAP values. This works towards answering the second part of RQ2. We can also see the same effect, albeit lower, from the number of free school meals students receive in the class (countFSM). This implies that the effect of financial status on classroom level can have a higher effect on students' learning outcome; this is called the compositional effect where students achievement is affected more by the socioeconomic status of the classroom than their own [80].

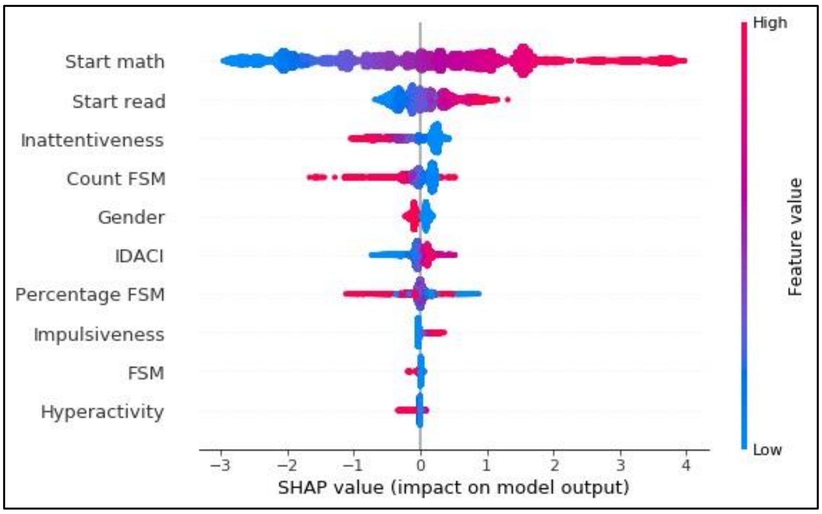

Figure 13. SHAP summary plot of features instances

Next, I examine the feature dependence, by plotting the features that are most frequently associated with each other, and the classified values are shown in Figure 13, 14, 15. Figure 14 shows a frequent association between gender and inattentiveness behaviour, with zero representing male and one representing female. We see a clear link of higher inattentiveness among male students than female students. I notice a negative relationship between the male gender and our prediction value, and we can see that the negative association with prediction also increases with inattentive scores of female students, which indicates that the higher the female students' inattentiveness scores, the lower their predicted learning outcome. The case is surprisingly different with male students, as their relationship with learning outcome value falls on the positive side, and inattentiveness scores do not show a particular association with the increase or decrease of the learning outcome, suggesting that male students' learning outcome is not affected by their inattentiveness score as much as female students. Although some studies suggest no differences between students with attention problems and their achievement based on gender [182], one explanation for this might be that females generally display lower calculation fluency than males as well as no noticeable growth in math skills over time when displaying symptoms of Inattentiveness [35].

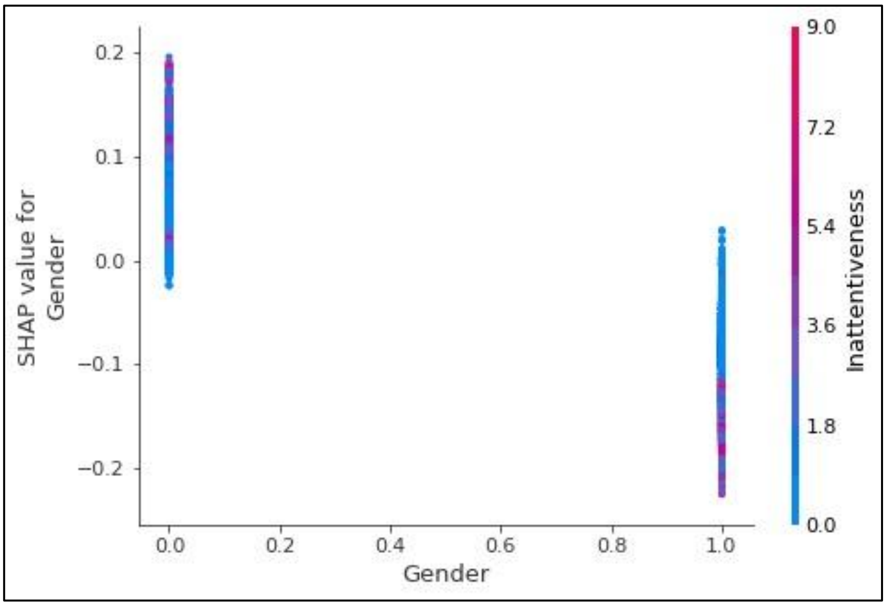

Figure 14. Dependence plot of gender and Inattentiveness

Figure 15 also shows some interesting outcomes, where Start Math associates with inattentiveness. We can see that a noticeably large number of students have low initial knowledge in math and a high inattentiveness score, showing that this disruptive behaviour may affect students' further performance as both low Start Math and inattentiveness are associated with lower learning outcomes (Figure 13).

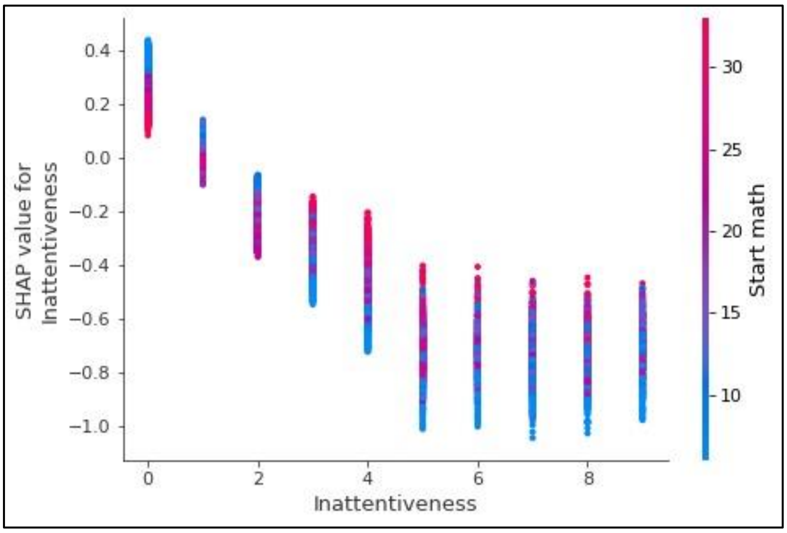

Figure 15. Dependence plot of Start Math and inattentiveness feature

Another interesting observation comes from Figure 16 where IDACI is associated with start read scores. We can see mostly a positive relationship between these values and the predicted learning outcome, but we can also see a clear negative SHAP value for students with a high start read score, but lower IDACI, which suggests that the impact of having a lower IDACI is higher than the initial knowledge, for this group of students. To elaborate, the data seems to point to the situation where students who have high initial knowledge (Start read) would naturally have high end knowledge (End Read) but some students with low IDACI would not improve well, showing low learning outcomes, despite their high initial knowledge, possibly due to poverty.

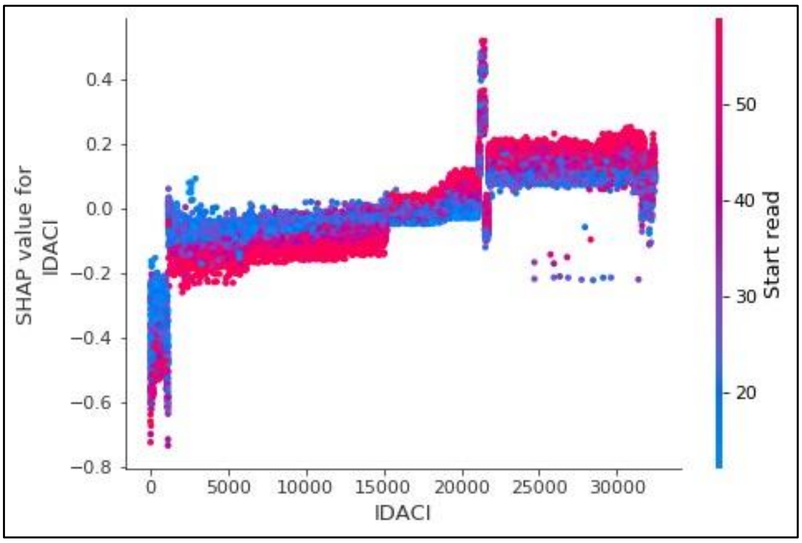

Figure 16. Dependence plot of IDACI and Start read

SHAP provided a ranked list of the most influential features in the model's predictions. For example, it highlighted that Start_Math had the strongest positive impact on the prediction of End_Math scores, while Inattentiveness and Hyperactivity had negative influences. This ranking helped in understanding which factors were most critical in predicting student performance. It also allowed for both global (for the entire dataset) and local (for individual predictions) explanations. This made it possible to see the overall patterns, as well as specific instances where disruptive behaviour significantly impacted individual student outcomes. The summary plots, dependence plots, and feature importance diagrams made it easier to visualise and interpret the effects of disruptive features. This was useful for model understanding, as it provided a clearer picture of how different characteristics interacted, to produce the model's predictions.

5.6 Epilogue

In this chapter, I leveraged ML techniques to explore three disruptive behaviours, specifically, Inattentiveness, Hyperactivity and Impulsiveness, since these behaviours were known to impact student learning outcomes in the classroom [11][18][36][138][188] (see Chapter 2; section 2.1). This study generated new knowledge by revealing variations of how these behaviours impact academic performance, as well as powerful trends, including the way Intensiveness impacted male and female students distinctively. I was able to show an effect of certain socioeconomic factors, like the IDACI on classroom level academic outcomes, above and beyond the individual level measures such as initial knowledge. Moreover, I tuned an XGBoost model and applied natural breaks in the data, resulting in a high performing predictive model of the involved nature of disruptive behaviour with academic attainment.

The use of SHAP has made the ML findings more transparent, which is a great help in educational AI research. It enables researchers and decision-makers to make clear, informed choices about where interventions are needed. By showing how disruptive behaviours affect students differently based on gender and socio-economic factors, this chapter establishes the groundwork for targeted, data-driven approaches to support students dealing with these challenges.

This chapter addresses RQ2 by demonstrating how interpretable ML can shed light on complex behaviours that influence learning. In the following chapter, I'll build on these results by blending ML capabilities with ABM to tackle RQ3. Through this combined ML-ABM approach, Chapter 6 shows how machine learning can improve ABM's ability to predict and explore disruptive behaviour, creating a fuller, more effective way to understand classroom dynamics.

CHAPTER 6

6 Improved version of the ABM with ML

6.1 Prologue

Building on the exploration of factors affecting student performance through ML techniques in the previous chapter (Chapter 5), this chapter advances the thesis, by integrating ML into an ABM. Chapter 5 reveals key insights into the predictive role of disruptive behaviours, specifically as defined via three parameters, inattentiveness, hyperactivity, and impulsiveness, on student learning outcomes, using XAI techniques. These findings provided a foundation for assessing student performance and highlighted the broader potential of ML in understanding educational dynamics in term of disruptive behaviour. Chapter 4 discusses the challenges of simulating realistic classroom environments and it underscored the limitations of traditional ABM in accurately predicting performance outcomes (see section 4.6). Thus, in the current chapter, I attempt to answer the following research question:

RQ3: How can Machine Learning (ML) be integrated into an agent-based model (ABM) to improve the simulation of classroom disruptive behaviour, and what parameters of ML prediction yield realistic results in this hybrid ML-ABM approach?

In this chapter, I refine the initial design of the ABM, by introducing ML to address two parts of the research question. First, I investigate how ML can enhance the ABM's ability to simulate disruptive classroom behaviours and interactions. Second, I assess which ML parameters yield realistic results when predicting student performance within this hybrid ML-ABM model. This hybrid model addresses realism challenges identified in Chapter 4 and leverages ML insights gained in Chapter 5. Moreover, this approach not only strengthens the model’s predictive accuracy, but also aligns with the overall research objective of this thesis to develop a data-driven hybrid model that aids in educational decision-making. This chapter's contribution lies in its demonstration of how ML and ABM can offer a richer and more dynamic representation of classroom behaviours and outcomes when combined.

Thus, here I extend my work on a classroom ABM from Chapter 4, by developing an ABM & ML hybrid model that simulates classroom disruptive interactions during a school year, and outputs predicted learning outcomes. This chapter is based on my paper [5].

6.2 Approach

In this work, a hybrid ABM and machine learning model is designed that simulates pupils' interactions during classroom lessons for one academic UK school year. As this model is based on data from UK primary schools, it has a potential of reproducibility for primary school system that is similar that of UK schools - with similar 3-terms structure, similar 45-minute lesson structure and similarly teaching mathematics (abbreviated to 'Math' below) every day. ABM outputs disruptive state minutes that are mapped onto the range of End Math score, as previously introduced in Chapter 4; and ML outputs predicted End Math score from data; then, both models' output is combined to represent the final End Math score, as shown in Figure 17.

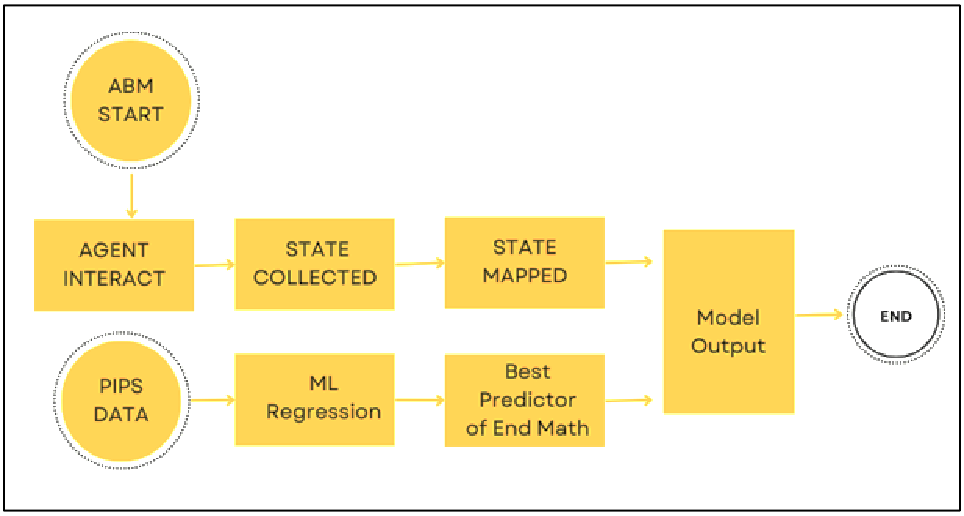

Figure 17 ABM-ML output process flowchart

6.2.1 Data

The main source of data was obtained from the Performance Indicators in Primary Schools (PIPS) monitoring system [197][9] mentioned in section 3.4.2. The full list of variables used in the model from the dataset is explained in this section:

- *Start Math score*: is the initial math score of the student from the baseline assessment (0-63 range);
- *Start read(ing) score*: is the initial reading score of the student from the baseline assessment (0-169 range);
- *End Math score*: is the final score in math from end year assessment (0-69 range);
- *End Read (ing) score*: ix the final score in reading from end-year assessment (0-178 range).
- *Start Vocabulary*: The total number of initial words the student knows.
- *Gender*: Boolean value 0 for male and 1 for female.
- *Student ID*: a unique number that distinguishes each student in the dataset.

[9] RR344_-_Performance_Indicators_in_Primary_Schools.pdf (publishing.service.gov.uk)

6.2.2 Simulation Overview and Scheduling

Here I summarise the model design: the model runs in time steps called *ticks* that represent 1 minute of classroom time. The run stops when it reaches the end of an UK school year, i.e. 190[10] teaching days in total, with delivery of one Math represented by 45 ticks per lesson per day. The simulation has 8,550 ticks in total, which accounts for the whole duration of all math classes that occur during the entire school year. More details on the design can be found in Chapter 4. The updated model presented in this chapter introduces several new sub-models that were not part of the ABM model in Chapter 4, along with the integration of machine learning (ML) predictions. The following sections provide a comprehensive explanation of the updated model, detailing the new components and improvements. Where relevant, references are made to the original ABM model in Chapter 4, to highlight the similarities between the two models.

6.2.3 Design Concepts

Basic Principles: I base my model calculations on two sub-models: the *agent state model* and the *prediction model*. The agent simulates a pupil in a classroom, similar to the work in Chapter 4, and the prediction model extends the agent state model, and predicts the learning outcomes (End_Math score) of a pupil at the end of the year. I use the sigmoid function to create the probability distribution used to determine the agent's next state serving as a thresholding function similar to the work of [66]. Based on this probability, a pupil is in one of the states {*learning*, *passive*, *disruptive*}. Similar to the ABM in Chapter 4, the decision is updated with every time step, unless manually modified by the education stakeholder (e.g. teacher) (see Equation 6.1).

[10] NASUWT | Working Days (England)

Emergence: In alignment with the ABM framework presented in Chapter 4, the emerging result from the model is the pupil's End_Math score at the end of the simulated school year, considering their characteristics and those of their neighbours and of the whole classroom. Emerging patterns can be seen in the performance of a class and the level of disruptive pupils within that class. Another output is the state of the agents on every step of the model.

Adaptation: Agents (simulating the pupils) are given some predefined rules, and they adapt their behaviour accordingly—much like the ABM model described in Chapter 4. For example, if a pupil with a moderate disruptive level sits next to low disruptive neighbours, they are more likely to be in a learning state than other states. Further explanations about the agent rules can be found in Chapter 4.

Interaction: Similar to the agent-based model discussed in Chapter 4, the model assumes that interactions occur between neighbours. If two agents are placed on the grid next to one another, the level of disruptiveness of both will either increase or decrease the chances of being disruptive on their next state.

Stochasticity: The movement of agents within the classroom follows a random pattern. Each day, pupils are assigned new seating positions with different neighbours, unlike the ABM in Chapter 4, where students remained in fixed positions. This approach allows for an exploration of how changing neighbours influence learning outcomes after initially examining the effects of a fixed seating arrangement [10]. This will be further explained in the *Seating_Update*

Seating_Update sub-model.

Observation: Data is collected with every run of the model at every time step and includes all variables explained in section 6.2.7 for visualisation and analysis of the model output. Additionally, the average score of Math of all agents is displayed in real-time during the simulation run.

6.2.4 Simulating Classroom Dynamics: Agent-Based Model with Spatial Units and Environmental Variables

Agents/Pupils: The agent in the model represents first grade pupils; they are represented via characteristics like age, gender, academic performance scores and more [7]. The full list of agent attributes and their description of is found in Table 13 with a newly added variable “Space Seat”.

Spatial Units: Each grid cell in the model represents a seating or non-seating area in a classroom.

Environment: The simulated world has a two-dimensional space made up of grid cells. Each cell represents a seating area of a classroom. Agents’ adjacent cells are called neighbours. The effect of agents on each other are affected by their position on the cells. Only neighbours can have an effect. The location of agents changes every 45 ticks to indicate another day of school. If two agents with high disruptive scores become neighbours, there is a high possibility that they will both become disruptive and increase the effect of disrupting their neighbours (see Equation 6.1).

Environmental variables within the model include Grid size, Number of Minutes, Number of Days, and average math scores of all agents. Each time step in the model represents one minute in a lesson of math. Every forty-five minutes is considered one school day [7]. The number of days is tracked to stop the model at 190 days (as this represents the number of school days in one academic year according to gov.uk).

6.2.5 Initialisation of Agent Attributes

The initial values of agent attributes (Start_Math, Start_Read, Inattentiveness, Hyperactivity, Impulsiveness, Age, Gender) is loaded from a sample taken from the PIPS dataset.

6.2.6 Creating the world: Agent Initialisation and Seating Arrangements

As I do not have real-world data on how the model initiates, I initiate agents with random states in random seating positions. For the agent's attributes, the model either loads pre-existing data or can create a random population. For the purpose of testing and validation for this thesis, I am using pre-existing data, as stated in the previous section. Agents are then placed on the 'Seating' grids. 'Seating' grids are specified to have at most two neighbours for each group of 'Seating' grids. This will reflect both regular seating in classrooms in columns and rows as well as in a groups, as both seating arrangements have their advantages [76], to improve the design in Chapter 4. Therefore, the maximum number of neighbours for all agents is two and the lowest is one. A screenshot of the front page of the model is shown in Figure 18.

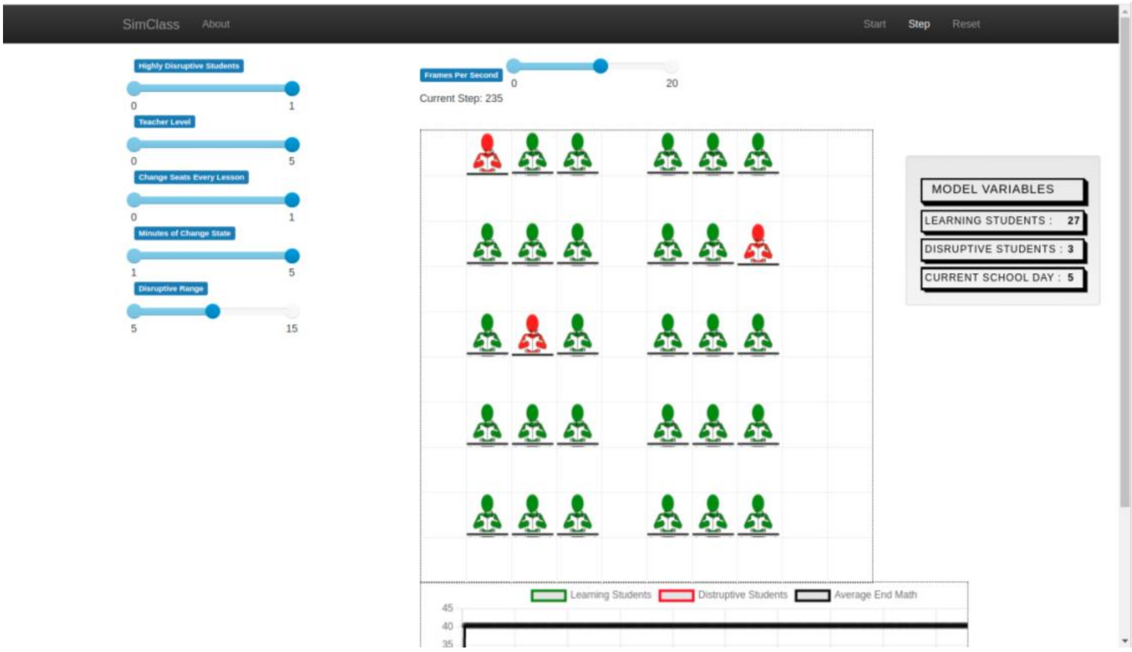

Figure 18 Screenshot of ML-ABM approach interface

6.2.7 *Configuring Simulation Parameters*

There are several parameters that need to be set before the start of the simulation. These parameters are either predefined and built into the model or can be altered by the educational stakeholder (teacher) to explore their effect on the simulation run results.

Table 13 Description of Hybrid ML-ABM components

Variable	Description
ID	Is a unique integer assigned by the model to identify each agent
Pos	The coordinates of the agent in the simulation environment
Gender	0 for male 1 for female
Start_Math	Integer between 1 and 70 for initial score in math
Start_Read	Integer between 1 and 160 for initial score in reading
Inattentiveness	Score between 0 and 9 for level of inattentiveness
Hyperactivity	Score between 0 and 6 for level of Hyperactivity
Impulsiveness	Score between 0 and 3 for level of Impulsiveness
Age	Age of agent at the beginning of the simulation run
State	The state of agent 1 for learning 0 for passive -1 for disruptive
Data	The records that will be used as an input to the model's agents, as explained in section 6.2.5
Space Seat	A set of patches or grid cells that represent the gap between seats in the simulated classroom
Minutes	The minutes of the lesson 1-45.
Days	The number of school day 1-190
Average End_Math	The average of End_Math score over all agents each minute

6.2.8 Simulation Framework for Classroom Dynamics

Input Data: The model uses input data from a sample of PIPS data (defined in section 3.4.2) that feed into the model a total of 30 pupils (representing the maximum size of an UK classroom), with their general characteristics (gender: Gender, age: Age, initial reading: Start_Read, and math Start_Math) as well as disruptive behaviour attributes (Inattentiveness, Hyperactivity, and Impulsiveness). Each pupil is assigned an ID by the model for tracing as well as a state and an initial position. For multiple classrooms, the model first groups the input data records file based on School and Class as index, to define unique classrooms within the data. The model then extracts each group/classroom and uses it as input into the simulation model, using the batch runner to be able to compare the results of classrooms under the same parameters.

Sub-models: The following sub-models are used to simulate the disruptiveness and learning outcome simulation. Some models are used in every step, and some are used when certain conditions are met. For example, the clock model is used when a full day passed, according to simulation time.

Time: the time model is called to keep track of time during the simulation. It contains a minute counter that increases by one for each tick of the model and once it reaches 45 minutes, the day counter is updated, to indicate a new day (as the model simulates a lesson of math every day). The minute counter is then reset to start a new count.

Seating_Update: The seating update sub-model updates the location of agents, for every day of the simulation time. With every update, it removes the agents from their designated current position first, before relocating all of them. As the initiation is simultaneous for all agents, they are placed back on specific seating locations in the simulated classroom, randomly, at the same time. Before placing agents in a location, it checks whether it is not already allocated to a classroom seating. If it is a seating location the agent is placed there, If the location on the simulated classroom space is set to be a space area, the sub-model searches for the next empty location, until a seating location is found.

Data_Extraction: this sub-model takes multiple classroom data within a file and splits it, so that each classroom is input into the model as a separate run/iteration. It uses integers that represent keys to each grouped records that represents a classroom.

State_Determination: When an agent spends a tick in one of the defined agent types, it changes the type on the next tick based on the result of the State Determination Test. This uses the sigmoid function to provide a range that defines the next state of the agent, based on their own characteristics and those of their neighbours. For example, if the agent is suitable to disruption (high disruptive behaviour values) and surrounded by a high proportion of agents in a disruptive state, the agent will be disrupted, and their state will change to disruptive (red).

State_Counter: A counter that keeps track of the agent's state. Two counters are included in the model; learning and disrupt counters. If a student is in a learning state, the learning counter will increase in every step the agent remains in that state. Similarly, the disrupt counter increases with agent's disruptive state. The counter resents once the agent switches to a different state.

6.2.9 Mathematical Learning Model

In the simulation model, each pupil is to be in a state S. Every state is represented by a number; learning is 1, passive is 0 and disruptive is -1. The pupil's learning though the school year is influenced by the pupil i's total disruptive score D_i and neighbour's state $N_{(t,i)}$ adjacent to pupil i. The current state $S_{(t,i)}$= f(P(X)) with X given by:

7. $X = \sum_{j=1}^{j}(D_{(t-1,j)} * N_{(t-1,i,j)} + D_{(t-1,i)} * S_{(t-1,i)})$ (6. 1)

Where $D_{(t-1,j)}$ is the disruptive score of pupil j at time $t-1$ and $N_{(t-1,i,j)}$ is the state of neighbour j adjacent to pupil i at time $t-1$. $D_{(t-1,i)}$ is the pupil's own disruptive score at time $t-1$ and $S_{(t-1,i)}$ is the pupil's own state at time $t-1$.

The probability of changing to a disruptive state is calculated using the sigmoid function:

8. $P(X) = \frac{1}{1+e^{-X}}$ (6. 2)

The higher the resulted value, the higher the probability that the pupil will be in a disruptive state. The general look of the model can be seen in Figure 21, showing the two phases used to update the agent state and compute the agent's End_Math score. X is then tested by The function *f* determines the new state based on *P(X)* and an arbitrary threshold value *n*:

if *X* < *n* then the pupil will be in a *learning state*;

elseif the pupil == impulsive, they will be in a *disruptive state*;

else they will be in a *passive state*

The value of *n* was determined experimentally. The output *L* of the model is then calculated using the formula:

$$L_{(i,end-time)} = P_{(i,end-time)} - \sum_{t=1}^{t(end-time)} DS_{(t,i)} \quad (6.\ 3)$$

Where $P_{(i,end-time)}$ is the prediction of the ML model and $DS_{(t,i)}$ is the disruptive state of student *i* at time *t* after mapping disruptive state minutes into score [6] and $L_{(i,end-time)}$ is the *final learning score*, which adjusts *P* by subtracting the cumulative disruption score $DS_{(t,i)}$. The results of running the model with different values of *n* examining the density of End_Math scores between PIPS and simulated End_Math for each run showed that the closest results of Pearson correlation between End_Math and other features to PIPS is with *n=9*, therefore this value of *n* was chosen for the design of the model.

6.2.10 Prediction with Machine Learning

The role of machine learning is to be trained on the original data to predict the initial student learning score; then the best performing algorithm is chosen for incorporation into the ABM model. First, I removed all missing values in PIPS from all features that are used in the prediction process. Out of 65,385 records, the resulted records after missing values removal are: 36,844, of which 18754 are male; and 18275 are female, creating a balance in numbers between genders. The majority of the pupils do not receive a free school meal FSM, with 32242 compared to 5033 who receive a free school meal FSM. The distribution of the predicted value, End_Math score, is shown in Figure 19, to provide a visual guide to the data.

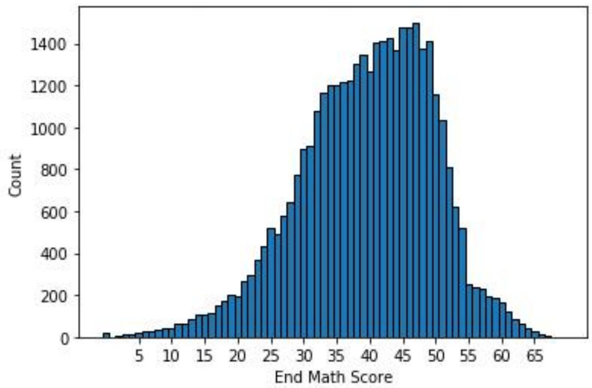

Figure 19 Distribution of End_Math scores

It can be seen from the figure that the majority of pupils' End_Math scores in all classes from the PIPS dataset fall between 30 to 55 from the range 0 to 69, with the highest number of pupils between 45 and 50. Next, I analyse the Pearson correlation of features with the predicted value (End_Math score), as shown in Table 14.

Table 14 Features correlation with End_Math

Feature	Correlation coefficient
Start_Math	0.70
Start_Reading	0.56
Start vocabulary	0.47
Inattentiveness	-0.36
End Age	0.25
Hyperactivity	-0.19
Start Age	0.18
Impulsiveness	-0.07

The highest correlated feature is Start_Math, followed by Start Reading. While the highest correlated feature among the disruptive features is Inattentiveness with -0.36, the negative correlation is to be expected in this case. The listed features in Table 14 are used for predictions of End_Math scores. For the measurement of our results, I apply the most widely used metrics for evaluating the performance of regression models, MAE and RMSE, as well as the Pearson correlation coefficient [157] while classification models were applied in Chapter 5 where evaluated using Accuracy, F-score [181]. We see from Figure 20 that the best performing algorithm for prediction of continuous values for the End_Math scores is Linear Regression, with a score of 5.41 of MAE. The highest score is to be expected, considering the nature of the dataset and limitation of the features. Based on these results, Linear Regression is chosen to be incorporated within the ABM model, for its higher performance among other models.

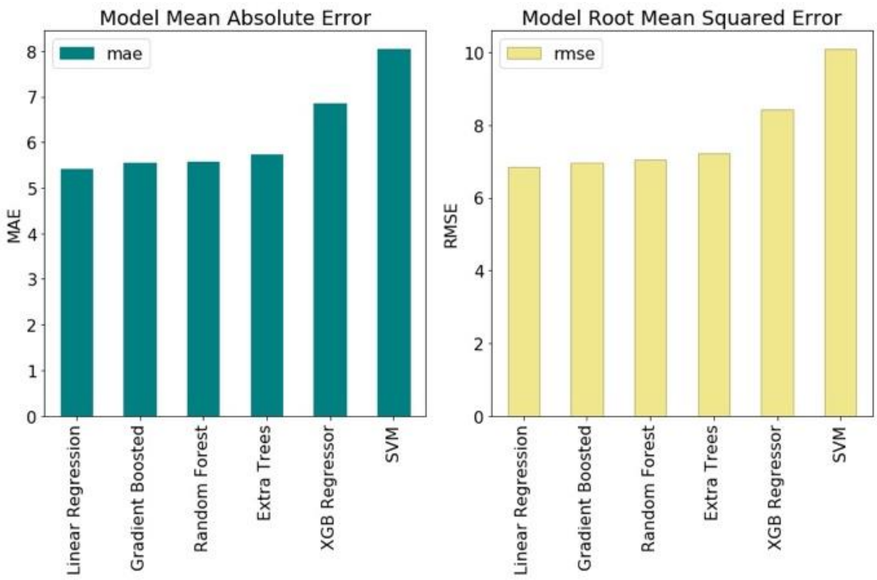

Figure 20 Machine Learning model performance

6.2.11 ML-ABM Hybrid Model Integration

The aim of a hybrid model is to improve the original model, by using both machine- and agent-based models, and benefitting from their synergetic effect. Thus, the model uses ML to predict outcomes such as students' End_Math scores, based on historical data. Moreover, the introduction of ML allows the model to handle complex and non-linear relationships between input variables. Thus, the hybrid approach is expected to allow the model to align more closely with real-world data, ultimately providing a more accurate prediction of outcomes. This is to replace the previous ABM model in Chapter 4, which primarily simulates classroom interactions and outcomes (End_Math) of the simulation based on predefined rules and student behaviour during the simulation. The simulator was run by testing the data of 30 pupils, 12 of whom were disruptive. Figure 21 depicts the flowchart of the ML-ABM approach that involves setting variables through both the simulation environment and the PIPS data shown in Table 13, deploying agents to the environment through the grid, and starting the model. Finally, the agents will be updated using the state determination model (*State_Determination* submodel see section 6.3.8*)*, moving states from the initial state to a new one. If the state of the pupil is disruptive, the disrupt counter will be updated to reflect their state (*State_Counter* submodel see section 6.3.8). If the state of the pupil is disruptive, the disrupt counter will be updated; otherwise, the learning counter is updated. Next, the model checks if the lesson is over (*Time* submodel see section 6.3.8); if not, it calls the state determination model for agent states. If the last school day is reached, the pre-trained ML model is activated, and final results are calculated, before the model stops. Table 15 highlights the key differences between the ABMs in Chapters 4 and Chapter 6, emphasising the impact of incorporating ML in Chapter 6.

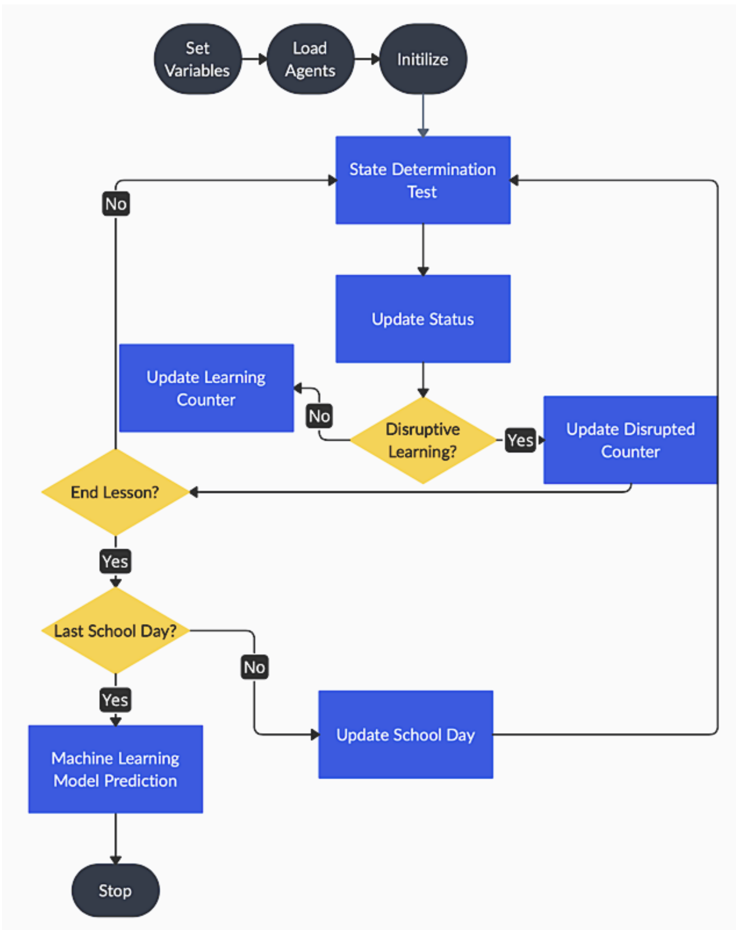

Figure 21 Flowchart of ML-ABM approach

Table 15 Comparison of Agent-Based Models (ABMs) between Chapters 4 and 6

Feature	Chapter 4 ABM	Chapter 6 ABM (with ML)	Improvement
Model **Basis**	Pure ABM.	Hybrid ABM with ML.	ML was added to enhance prediction accuracy of learning outcomes.
Data Source	PIPS data for initial conditions and variable ranges.	PIPS data with additional ML training data.	Expanded data source supports realistic outcome predictions.
State Changes	Based on predefined rules.	Enhanced with state change probabilities.	Refines state transitions to introduces stochasticity, mirroring real-life variabilities.
Sub-models	Limited to agent behaviour and classroom interactions.	Additional sub-models, including seating updates, state tracking, and time management.	Added to simulate more complex, realistic classroom dynamics.
Outcome Predictions	Determined by ABM rules and simulation time.	Determined by both ABM interactions and ML-predicted scores.	ML integrates non-linear relationships, increasing predictive realism.

6.3 Results and Discussions

The model is run in a scenario based on the parameter: Seating. The model runs with the Seating parameter set to 1 for a changed seating every new simulation day. The average End_Math scores are 32.63 while the average End_Math Scores of real data are 33.21. Figure 22 shows the density plot of the resulted simulated End Math scores and the PIPS End Math scores.

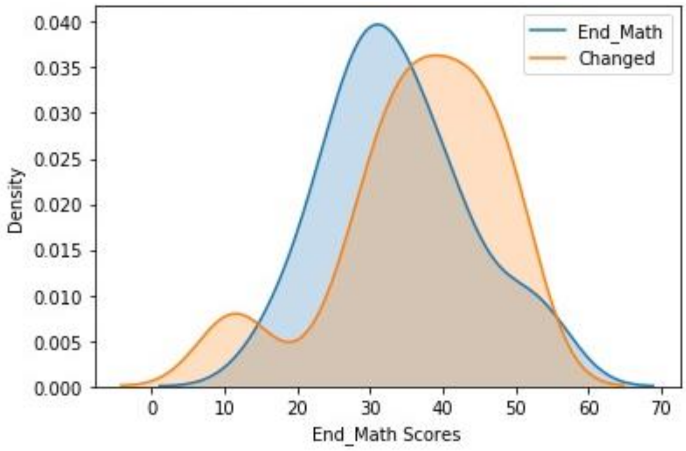

Figure 22 Density plot of End Math scores by changed seating condition

To establish the consistency of the model results, the model was run for 100 runs. Figure 23 shows the resulted mean End_Math score of all runs. We can see that the model runs yielded a mean of End_Math scores that falls mostly between 35 and 32, with only one run that yielded a mean around 30. The results of the multiple runs indicate that the model is consistent.

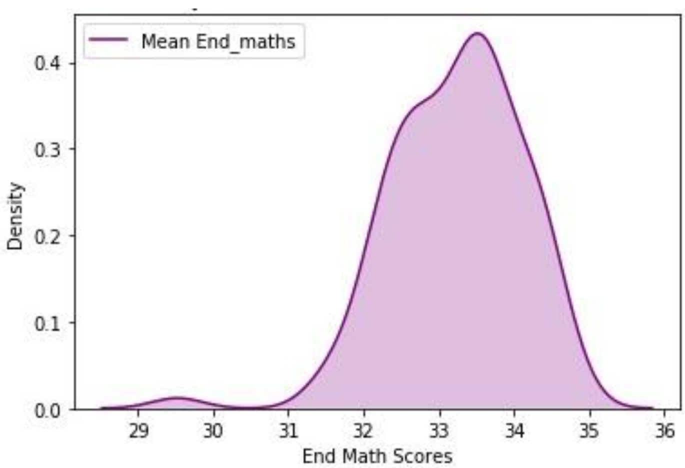

Figure 23 Mean of End_Math scores for 100 simulation runs

For the validation of our results, I use Pearson correlation coefficient that describes how well two variables tend to move in the same direction. All values are listed in Table 16.

Table 16 Pearson correlation between features and observed End_Math and simulated End_Math

Feature	End Math	Simulated End Math
Start Math	0.61	0.68
Start Reading	0.69	0.45
Inattentiveness	-0.41	-0.59
Hyperactivity	-0.30	-0.63
Impulsiveness	-0.26	-0.49
Start Vocabulary	0.31	0.36

We can see that the correlation coefficient between Start_Math and End_Math is 61.48 and with Simulated End_Math 71.32; both are an indication of a strong correlation between the two values.

6.3 Epilogue

In this chapter, I developed and tested a hybrid ML-ABM model to improve the accuracy of simulating classroom interaction, especially the prediction of students' academic performance influenced by peer interaction over a school year. The study was able to capture the impact of students' characteristics, as well as their dynamic interactions on final academic outcome, by integrating ML into the ABM framework. Results from this simulation were verified by actual End_Math scores (of real-life students), demonstrating that the model captures the observed learning patterns in a classroom. This work has shown that the hybrid model can be a practical tool for educational decision makers. This provides insight into students' interactions, informing targeted interventions and policy changes. In addition, the model offers a platform for researchers interested in educational interventions to evaluate and improve their strategies in a controlled, but realistic, simulated environment. Since this model is designed using data from UK primary schools (three-term structure and a 45-minute lesson duration), comparable data from the target educational context would be required to apply it in a different educational context. This includes data on student behaviours, initial and final performance scores, and contextual factors like classroom size and academic schedule.

The next chapter (Chapter 7) then explores further potential enhancements to this model from a usability point of view, by incorporating gamification strategies. The gamification strategies are to be explored first in a classical context, to gauge student engagement in intelligent learning system. Specifically, the next chapter analyses how gamified elements can be included and evaluated within educational systems to improve motivation and support engagement. Then, Chapter 8 explores how these gamification strategies can be used in the context of a teacher-facing tool.

CHAPTER 7

7 Enhancing Educational Dynamics: Evaluating Gamification within Intelligent Learning Systems

7.1 Prologue

In this chapter, which was published as my paper [4], I investigate the gamification in large-scale online learning systems by examining the CamaleOn platform, a Brazilian MOOC that aims to increase the chances of students from disadvantaged backgrounds to get into university. Building on insights from the previous chapter (Chapter 6), where a hybrid ML-ABM was developed to simulate classroom behaviours and improve predictive accuracy, this chapter expands the scope by analysing student engagement patterns in a real-world educational context. Study in this chapter uses CamaleOn data to investigate how gamification can sustain engagement on a broad scale, thereby supporting effective learning environments.

In this way, the focus now turns to a data-first approach, moving away from theoretical assumptions about gamification and towards practical analysis from grassroots data. The objective of this approach is to determine which gamification elements (such as badges, points, medals) enhance engagement and to evaluate how such elements might be further optimised, in order to encourage sustained user interaction and performance.

In this chapter, I partially answer the following research question:

RQ4: How can gamification strategies be implemented to increase engagement in an educational setting, and which gamification elements have the most significant impact on engagement, both in student-oriented systems and teacher-oriented systems?

To answer the student-related part of the question, I explore a large, gamified learning environment's log data for students with various optional game elements. Here, I use machine learning classifiers to predict their effect over engagement. Overall, this chapter is important for the thesis, since it is an analysis for the thesis's larger goal of building engaging educational systems, as well as a first step towards building engaging teacher-facing tools, such as the ML-ABM. Thus, this work is an important initial step to understand how gamification insights can be applied to ML-ABM models to enhance the productivity of learning environments, by increasing engagement through data-informed educational tools.

7.2 Approach

To understand how to improve gamification based on existing or expanding theories, I have analysed user behaviour in a given system of e-learning, and based my improvement suggestions on the existing user behaviour. In my case, this system is CamaleOn (see 1 section). This is a realistic approach, as many educational online systems are available and in use, and it is costly and often problematic to change them completely. Instead, a more gradual approach to this change is proposed, based initially on available data, and subsequently informed by gamification theories.

7.2.1 Data

The data used in this chapter is based in CameleOn that was introduced in Chapter 3. In this chapter, I provide further details about this dataset. There is a particular focus on providing students from public schools the resources needed to attend a Brazilian university. To motivate the user to continue with the website, CamaleOn uses different aspects of gamification (e.g., elements such as experience points (XP), badges, etc.) as methods for motivation explained as follows:

Level: Students start at level 1 and move up to higher levels through gaining XP points. These points can be gained through answering questions correctly and interacting with the system in different ways.

Badges: Students can earn from a range of 13 badges available in the system. These badges have a different avatar and can be earned when the student performs the action related to the said badge such as: learning a subject in 3 days, staying one hour in the system or getting a silver medal for the first time.

Figure 24 presents the design of CamaleOn's webpage, where points (XP) are displayed on the top of the screen at all times to provide a visualisation of the student's advancement via a gamified progress bar. Trophies are greyed out until earned; each holding a label explaining how it can be earned. Additionally, a progress map at the bottom of the screenshot visualises the student progress through the subjects of the curriculum.

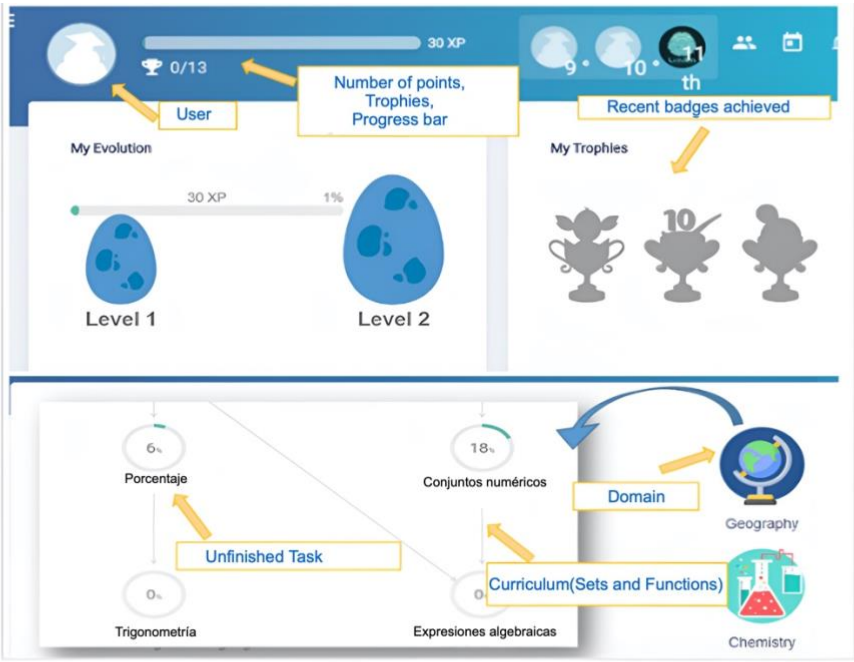

Figure 24 CameleOn's user dashboard

7.2.2 Defining the Effect of Gamification Elements on Student Engagement in CameleOn dataset

Students receive points by advancing in the subject level which means point gain cannot be linked directly to the effect of gamification. Badges, however, require specific actions performed by the student that are not generally gained by simply using the system and solving questions. Therefore, I choose them as an indicator for student gamification element engagement. Trophies and badges as well as peer emotion feedback were among the most engaging gamification mechanics in MOOCs as suggested by the findings of [38].

To have a general understanding of the data, I will calculate the number of badges for each student then perform a correlation test between the number of badges for a student and student's interaction with course material.

Provided that the correlation test shows a positive correlation between the number of badges and the number of student activity, I will use both features as predictors for student advancement in subject levels using machine learning algorithms.

Based on these observations, Engagement was defined as:

Engagement: is students' system interactions as well as course material interactions as students can advance through the system without accessing all resources but accessing such resources will win the student a reward.

7.2.3 CameleOn Dataset pre-processing

This section introduces the description of the CameleOn dataset used for this chapter as well as the steps I followed to prepare it. One of the of the objectives of our research is to find the effect of gamification elements and social interactions on students learning in an online environment. Several steps were taken for preparing and extracting the dataset that is explained as follows:

The dataset was received as a large dump file which required extraction of the data. To open the dump file, a local server was installed using Apache and XMAAP. Then, a database was created to load the dump file into. However, due to the large size of the file, the process of importing to the database faced multiple failures. Therefore, Durham University's MySQL service was used to create an online database and the dump file was then successfully imported into the database. The database, however, suffered from inconsistencies where some tables had a large number of records (180,104 record) and several tables had significantly smaller number of records (800~22 rows) while a large proportion of the overall existing tables had zero records. Every dataset requires a document that comprehensively explains the purpose of every table and the relationship between different tables in the database as well a detailed description of the system the data was collected from. Unfortunately, none of which was available in this dataset; therefore, a primarily step was to create a documentation which included the mentioned information in order to use as a reference while analysing and processing the data.

7.2.4 Processed CameleOn Dataset

I include the description of features provided in the CameleOn dataset in Table 17 and dataset statistics in Table 18.

Table 17 CamaleOn dataset features description

Feature	Explanation
Activity Loops	The description of the activity archived
type	The type of reward gained by
Domain	The domain of the subject accessed by student

Curriculum	The levels in each subject domain
Topic	Files provided as resources in each level
Resource type	The type of used resource i.e. video or pdf
Interaction Time	Time stamp of the interaction
View Time	Length of time for resource interaction measured in milli seconds
Current Points	Number of total points archived through system interaction
Current Level	The highest gamification level reached
Type Avatar	Type of avatar based on level
User_ID	Unique number for identifying users in the system
Login Time	Time stamp of login to the system

Table 18 Statistics of CamaleOn dataset

Category	Type	Number
Material	Video interactions	827
	Pdf interactions	304
	Mean view Time	136818.76ms
Questions	Correct Answers	177235
	Wrong Answers	130579
Activity	Total students' activities	236346
	Type of reward for activities	3
	Highest reward	points
	Highest Domain accessed	History
Users	Total number of users	5809

7.2.5 Data Preparation

To prepare CameleOn dataset for analysis, the following steps were applied:

Step 1: I checked for the existence of abnormal values in feature instances. These abnormalities can form due to wrong inputs or character encoding errors. Based on the relevance to the problem, I try to answer, I either remove the whole record that contains the irregular value or replace the value with the mean of this feature values.

Step 2: A check was carried out for null observations in selected features for analysis as they may cause errors in the programming code or affect results. Null observations can be the cause of combining two tables that hold different log data for example: gamification interactions and material interactions where a student might have advanced in the system without accessing any material which resulted in a number of empty values when these two kinds of interaction were combined for analysis.

Step 3: Large float numbers such as percentages of wrong and correct answer to questions, were rounded to two decimal points to reduce run time.

7.2.6 Matching Data to Research Questions

The first step in the data-driven approach is to extract refined research questions from the data, based, on the overall aims of the research. In Table 19, the bold words in the “Data Subset” column indicate which dataset the subset of data originated from. The list of attributes, following the dataset, are the attributes which were selected from that subset. Analysing the attributes and data available from CamaleOn, I need to first extract the gamification elements used; here, they are badges, points, medals. For student engagement, frequency of interaction can be used (e.g., number of logins) and lack of dropout (thus involvement in the higher levels of the course).

__Table 19. Matching data sets to research questions.__

Data subset	Research items
Students: Number of Points, Number of Badges, Number of Medals, Number of Problems Solved, Number of Mistakes and Number of Correct Answers Logs: Log Type (equal to "Problem Solving"), Problem Correctly Done	Investigate performance of students versus engagement

The purpose is to find out if existent gamification features are useful, and if more gamification features need to be introduced, to address engagement. It is important to note here that further analysis is possible, and that this work only illustrates how existent data may be used to improve the design of an extant system.

7.2.7 Definitions and Measures

For my research question, I chose to define engagement by both the number of logins and the total number of question attempts (following research such as [43]). This is due to the fact that students' academic performance is not a necessary indication of engagement[132]. Here, I set the threshold for the highly engaged group of students as consisting of students $u \in St$ from the student cohort, where:

$$G_{HE}=\{u \in St \mid \#login_u \geq avg(\#login_{u;\,\forall u \in St}) \text{ AND } \#questions_u \geq avg(\#questions_{u;\,\forall u \in St})\} \quad (7.1)$$

Where G_{HE} represents the set of highly engaged students and $u \in St$ refers to individual students in the student cohort St. $\#login_u$ is the number of times a student u has logged into the system and $\#questions_u$ is the total number of question attempts by student u. $\#login_{u;\,\forall u \in St}$ is the average number of logins across all students in St while $\#questions_{u;\,\forall u \in St}$ is the average number of question attempts across all students in St.

$\#x$ refers to the number of x and $avg(y)$ computes the average value of y. This corresponds, for our data, to students who have logged into the system more than 8 times and attempted to answer at least 304 questions, which are the mean values for number of logins and question attempts, respectively. This resulted in 1058 highly engaged students, and 7212 less engaged students. The gamification elements in the system are:

Points: points are earned by answering low level questions.

Medals: medals are earned by showing high skills in questions answering, such as answering all questions in a topic correctly or solving side assignments.

Badges: They are earned by interacting with the system in a specific way such as: spending one hour in the system or learning a sub-assignment 3 days in a row.

I defined a measure for the gamification elements introduce into in the system by the variable "Reward Count" RC_u, as the sum of Points p_{qu}, Badges b_{qu}, Medals m_{qu}earned by a student u:

$$RC_u = \sum_{q=0}^{\#no\_que} p_{qu} + \sum_{q=0}^{\#no\_que} m_{qu} + \sum_{i=0}^{\#no\_int} b_{qu} \qquad (7.2)$$

Where RC_u is the total reward count for a student u and $p_{qu}$is the total points earned by student u, from answering question q; and $\#no\_que$ are all the questions. Similarly, m_{qu} is the number of medals earned by student u for high performance in a given question or assignment q, while b_{qu} is the number of badges earned by student u for specific interaction q, for instance, a badge for a certain amount of time spent on the platform; $\#no\_int$ are all possible interactions. I first answered the research question using correlation analysis, based on the Pearson coefficient. Next, I use both shallow and deep learning methods to further answer the questions in more depth. For shallow methods, I use and compare a number of ML models for classification: Linear

Regression (LR), Linear Discriminant Analysis (LDA), K-Nearest Neighbours (KNN), Classification and Regression Trees (CART) and Naive Bayes (NB). Then, I apply two deep learning algorithms to compare the performances of Machine Learning (ML) against Deep Learning (DL) models for numerical data with a low number of predictors, namely Multilayer Perceptron (MLP) and Convolutional Neural Network (CNN), which are recommended for numerical, non-sequential data. Figure 25 provide a general view to our methodology.

Figure 25 General view of methodology followed in answering the research question

7.3 Results

7.3.1 Normality Test

For the normality test of high and low engagement for students, I applied the Kolmogorov–Smirnov test, rather than Shapiro Wilk, due to the large data size that exceeded 5000 instances. Results indicate a non-normal distribution for each group ($p \leq .00$).

7.3.2 Data Visualisation: Higher/Lower Engagement versus Gamification Use

I next visualise the two groups to analyse visual differences in gamification elements' use via the total number of earned rewards for each group (see Figure 26).

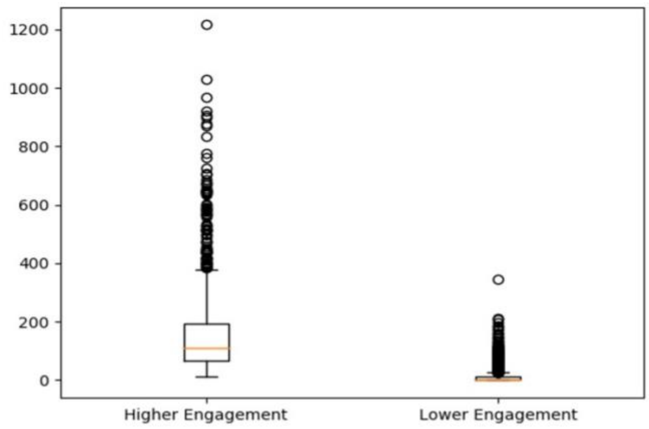

Figure 26 Box plot of higher and lower engagement groups versus gamification

7.3.3 Data Correlation: Engagement versus Gamification

Table 20 shows the correlation between engagement and gamification. For instance, it indicates a strong positive association between students' number of logins and the number of rewards they earn. The highest correlation value is noticed between badges and engagement status. The lowest value is seen between the number of earned medals and that of logins, possibly due to fact that medals are questions and curricula related. However, the engagement variable "Is Engaged" shows a positive association with all of the gamification elements represented by Reward Count, .

Table 20 Correlation test results between engagement indicators and gamification elements.

	Reward Count	Points	Medals	Badges
High login	0.531	0.482	0.373	0.631
High question attempts	0.660	0.656	0.604	0.671
Is Engaged	0.660	0.656	0.604	0.682

7.3.4 Engagement Prediction based on Gamification

Following the correlation test results, I used the gamification elements and the additional aggregate parameter "Reward Count", and their combination, as inputs of different dimensions, to classify high and low engagement with various classification models. The output of the classifier would either be the learner is engaged (1) or not engaged (0). These results show that the CamaleOn gamification elements are a strong predictor for students' engagement, with all accuracies > 0.924. I.e., the number of rewards students earn is strongly linked to the number of logins and general advancement through the system. The accuracy of CNN and MLP exceed the traditional ML models, suggesting that ML and DL classifiers perform slightly better - but similarly, for problems with a small number of features, such as this. MLP was the clear overall winner in terms of prediction model comparison. The highest score is observed (mostly) with the combination of all elements. What is interesting is the similarity of individual elements' score, despite the differences between them in functionality and purposes. I.e., Medals reward curricula advancement, while Badges reward defined system actions.

Table 21 Classifiers' results for engagement level based on gamification elements.

	Inputs	Acc	Low-engagement (0)			High-engagement (1)		
			P	R	F1	P	R	F1
LR	Reward Count	.951	.97	.98	.98	.87	.78	.82
	Points	.950	.97	.98	.98	.87	.77	.82
	Medals	.937	.95	.98	.97	.85	.65	.74
	Badges	.938	.96	.97	.97	.80	.72	.76
	All Elements	.954	.97	.98	.98	.88	.80	.84
LDA	Reward Count	.924	.93	.99	.96	.98	.46	.63
	Points	.950	.93	.98	.96	.98	.45	.62
	Medals	.937	.92	.98	.96	.96	.42	.59
	Badges	.938	.92	.97	.97	.80	.72	.76
	All Elements	.954	.96	.99	.97	.91	.72	.80
KNN	Reward Count	.947	.97	.97	.97	.81	.82	.81
	Points	.950	.97	.98	.97	.83	.78	.81

	Medals	.937	.96	.97	.96	.76	.75	.75
	Badges	.938	.96	.95	.96	.71	.75	.73
	All Elements	.954	.98	.98	.98	.84	.84	.84
CART	Reward Count	.944	.96	.98	.97	.81	.73	.77
	Points	.950	.96	.98	.97	.83	.69	.76
	Medals	.937	.95	.98	.97	.82	.67	.74
	Badges	.938	.96	.97	.97	.80	.72	.76
	All Elements	.954	.97	.97	.97	.81	.81	.81
NB	Reward Count	.954	.98	.97	.98	.83	.84	.83
	Points	.950	.98	.97	.98	.83	.69	.76
	Medals	.937	.96	.97	.97	.78	.74	.76
	Badges	.938	.96	.97	.97	.80	.72	.76
	All Elements	.954	.98	.96	.97	.78	.86	.82
MLP	Reward Count	.958	.98	.97	.98	.83	.84	.84
	Points	.957	.98	.97	.98	.83	.86	.83
	Medals	.942	.96	.98	.97	.82	.71	.76
	Badges	.941	.96	.97	.97	.80	.72	.76
	All Elements	.964	.98	.98	.98	.87	.86	.86
CNN	Reward Count	.957	.97	.98	.98	.85	.81	.83
	Points	.956	.98	.97	.98	.81	.87	.84
	Medals	.941	.96	.98	.97	.82	.69	.75
	Badges	.931	.93	.99	.96	.92	.51	.66

The models' performance in this chapter differs from those in Chapter 5 due to multiple reasons. First, the data type and structure are different as data from PIPS is collected through structured assessments and teacher observations, capturing direct academic performance and specific disruptive behaviours. This structured format leads to data that is relatively clean and targeted, with a focus on academic metrics that can be well-suited to traditional machine learning models due to their ability to handle discrete, non-sequential inputs and low-dimensional spaces [28, 113].

CameleOn's data, on the other hand, is less structured, as it records real-time digital interactions across various gamification elements. It includes frequent, repeated measures of engagement-related behaviours like frequency of logins and badges earned, making it larger, richer, and more complex but also noisier than PIPS. This requires models that can handle time-series or sequential data, which are more dynamic than the static assessment data in PIPS[105]. Secondly, the size of the data is another elemet where PIPS is vast but has a more controlled data scope, focusing on academic achievement and a limited set of behavioural metrics. The uniformity of PIPS data supports models that rely on structured inputs, such as tree-based models and simpler ML techniques. CameleOn's dataset, while extensive, is more variable, with 8270 students and millions of interactions across different gamification activities. This variation provides a wider feature range but also requires complex model tuning to account for the various engagement activities. Deep learning models, such as Convolutional Neural Networks (CNN) and Multilayer Perceptions (MLP), which handle high-dimensional data effectively, might thus outperform simpler models in this chapter [109].

7.4 Epilogue

In this chapter, I presented a data-driven approach to understanding how gamification affects student engagement and performance on a Brazilian MOOC aimed at high school students preparing for higher education. The study analysed how some gamification elements, including badges, points and medals, led to increased student engagement and interaction on the system. This aligns with findings in broader literature, which confirm the effectiveness of gamification in promoting engagement and motivation across varied educational settings (see section 2.6, Chapter 2). The application of gamification in educational settings, such as MOOCs, is illustrated through this research, showing its potential to support educational outcomes in large-scale data rich environments. The adaptability of gamification strategies can extend beyond specific platforms or cultural contexts. By relying on general principles, such as rewards and recognition, gamification can be tailored to address the needs of different educational levels and disciplines. Although the model is oriented towards a Brazilian context, the insights it provides for gamification strategies could be applied to a variety of different age groups and educational levels. By using data-driven techniques and machine learning, the study provides a robust framework to study and predict how gamification affects student engagement. However, the data collected using these techniques may reflect inherent biases, whether from demographic factors, user behaviour patterns, or sampling methods. These biases can limit the generalisability of the findings, making them more relevant to certain user groups than others.

In the next chapter (Chapter 8), these findings can be extended, by including gamification elements in an ML-ABM framework for teachers. In this next stage, we apply insights from student engagement to the teaching context, to investigate how gamification can improve teacher engagement and interaction in intelligent learning systems in general, and our specific ML-ABM in particular. This integration not only strengthens the study’s overarching aim of developing dynamic, data-driven educational tools, but brings together experiences of both students and educators in a gamified framework.

CHAPTER 8

8 Enhancing Educational Dynamics: Integrating and Evaluating Gamification within the ABM System

8.1 Prologue

This chapter extends the work initiated in Chapter 7, where the impact of gamification elements on student engagement was explored through the Brazilian CamaleOn platform. Chapter 7 covered how gamification features like badges, points, medals, played a role in potential user interaction, and supports thus in general the inclusion of gamification features within learning systems, including potentially in our ABM for teachers. This chapter thus explores to what extent the implementation of gamification can enhance teachers' interaction, pedagogical efficacy and classroom management, especially in dealing with disruptive student behaviours. This thesis aims to develop a data driven hybrid ML-ABM model that can incorporate student and teacher dynamics to improve educational outcomes. The aim of this chapter is to contribute to this goal by analysing how the validated gamification strategies, as presented in Chapter 7, can be utilised for a teacher-focused ABM.

Parallel to empirical studies like Chapter 7, this chapter draws on theoretical frameworks, such as Toda's Elements for Educational Environments (TGEEE)[192], which offers a structured approach to applying gamification in educational contexts. By combining these theoretical insights with the data-driven findings from Chapter 7, this chapter seeks to validate how gamification can be effectively integrated into the hybrid ML-ABM supporting teacher engagement. Additionally, this chapter seeks to further explore and validate the answer to research question RQ3 of elements affecting student performance and the teacher-related part of RQ4. Thus, this chapter aims to answer the following research questions:

RQ3: How can Machine Learning (ML) be integrated into an agent-based model (ABM) to improve the simulation of classroom disruptive behaviour, and what parameters of ML prediction yield realistic results in this hybrid ML-ABM approach?

RQ4: How can gamification strategies be implemented to increase engagement in an educational setting, and which gamification elements have the most significant impact on engagement, both in student-oriented systems and teacher-oriented systems?

This synergy between the empirical research presented in Chapter 7 and the theoretical framework outlined in this chapter facilitates a holistic approach to designing a gamified environment that is aimed to be both engaging and effective. This means enabling the synergy between theory and data-driven practice, where the empirical findings from student engagement are enriched by the structured, theoretical understanding of gamification principles.

Frequently, findings from one educational context cannot be directly transferred to another without relevant attention given to the unique needs of the new context. In this thesis, the CamaleOn study provided valuable insights into student engagement, but its applicability to teacher engagement is to be used with much care. The challenge thus addressed by this chapter combining empirical findings with theoretical principles that facilitate the systematic integration of gamification elements into the hybrid ML-ABM. The empirical findings from CamaleOn highlight which gamification elements - including badges, points and medals - enhance engagement, while theoretical frameworks like TGEEE offer a structured understanding of how these elements can be applied in educational contexts. It is not just to increase teacher engagement, but, ultimately and more importantly, to reduce disruptive behaviours and offer a practical, data-driven teacher classroom management approach.

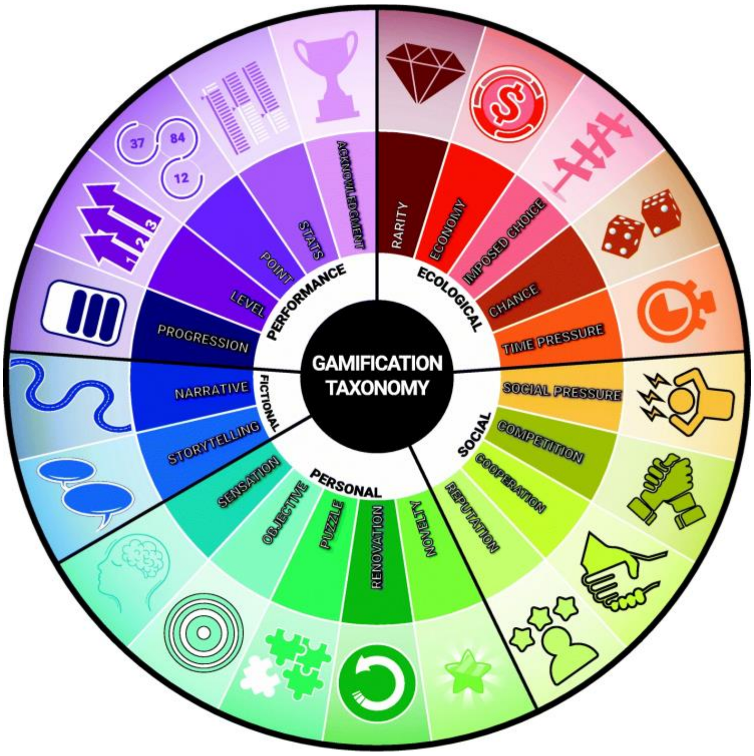

Figure 27 Toda's TGEEE Wheel: Taxonomy of Gamification Elements for Education Contexts *[192]*

8.2 Methodology

8.3.1 Initial Analysis and Application Framework

The initial phase involved a detailed examination of the CamaleOn system using advanced machine learning techniques to ascertain the impact of specific gamification elements such as badges, points, and leaderboards on student engagement levels. This empirical investigation utilized a blend of correlation analysis alongside both shallow and deep learning methodologies to pinpoint elements that significantly boosted student interaction and sustained engagement.

Subsequently, these validated elements are to be adapted and integrated into a hybrid ML-ABM approach designed for teachers. This integration is informed by Toda’s TGEEE, which provided a structured taxonomy of gamification strategies tailored specifically for educational settings. This taxonomy facilitated the strategic incorporation of elements like 'Statistics', 'Time', and 'Progression', ensuring that the gamification enhancements were both educationally relevant and empirically supported.

Toda’s TGEE Wheel [192] illustrated in Figure 27 introduces a structured way to classify gamification elements in education. It divides these elements into five key dimensions, each focusing on a different aspect of the learning experience. The Performance / Measurement dimension includes points, levels, progression, stats, and acknowledgement. These elements provide feedback to learners, helping them track their achievements. Without them, learners may feel lost or unmotivated. The Ecological dimension relates to the properties of the learning environment. It includes chance, imposed choice, economy, rarity, and time pressure. These elements shape interactions and ensure engagement. Without them, the environment may feel dull and uninspiring.

The Social dimension focuses on learner interactions through competition, cooperation, reputation, and social pressure. These elements encourage collaboration or rivalry, making the learning experience more dynamic. Without them, learners may feel isolated. The Personal dimension considers the learner’s experience and includes sensation, objectives, puzzles, novelty, and renovation. These elements ensure engagement and keep the learning process meaningful. Without them, learners may lose interest. Lastly, the Fictional dimension connects the learner to the learning environment through narrative and storytelling. These elements add context and immersion, making learning more engaging. Without them, the system may feel unmotivating and disconnected from the learner's experience. Toda's TGEE Wheel provides a structured approach to designing and evaluating gamified educational systems, ensuring a more effective and engaging learning environment.

8.2.1 Participants

To thoroughly validate and explore the potential expansion of gamification strategies for teachers within the ML-ABM approach, an extensive mixed-method study was designed and conducted with teachers teaching in schools directly. This study involved the structured interviewing of *twelve teachers*, carefully selected to represent a diverse range of teaching experiences and pedagogical backgrounds from various public schools in Riyadh, Kingdom of Saudi Arabia (KSA). Of these twelve participants, 58.3% (seven teachers) were teaching mathematics, while 41.7% (five teachers) were teaching regular reading. The study was conducted in public schools, which more readily provided permission for research, in contrast to private schools that are fewer in number. Given the restrictions and challenges posed by the COVID-19 pandemic, a smaller, convenient sample of twelve participants was chosen to ensure the safety and feasibility of the research. The pandemic necessitated limiting face-to-face interactions and minimising group sizes, which influenced the decision to use a smaller sample size. This is consistent with experiments with experts, in prior research; and here, teachers can be considered to be teaching experts, albeit with different amount of expertise, as follows. The selection of participants was strategically made to maintain diversity in teaching experiences while adhering to safety protocols. Among the participants, 33.3% (four teachers) were male, and the remaining 66.7% (eight teachers) were female. This gender distribution reflects an effort to include a balanced perspective on teaching practices and experiences in the study.

The research was structured to gather both qualitative insights and quantitative data, thereby providing a holistic view of the teachers' perceptions and the effectiveness of various gamification elements in the ABM system as well as the results of its improved version.

8.2.2 Procedure

The procedure was structured to ensure that participants interacted meaningfully with the ML-ABM approach and gamification elements, while providing feedback based on their experience. Each participant was first given a detailed introduction to the ML-ABM, followed by a brief tutorial on how to navigate the platform. Participants received an introduction to the ML-ABM system before starting the task. The session began with an overview of the ML-ABM's purpose, highlighting its purpose to support teachers in classroom management. Key system elements, including its functions and gamification features, were also presented. The task was designed to last approximately 40 minutes, during which participants were asked to engage with the simulation, using the gamified features, as they explored a virtual classroom scenario. The participants' main task was to explore the gamified ML-ABM in managing classroom interactions, including helping them in addressing disruptive student behaviours. Participants explored the system, by running approximately 2-3 simulations involving different classroom management decisions. The system generated outcomes based on their choices. After completing the task, participants were interviewed using a structured interview procedure with both close- and open-ended questions, to assess their perceptions of the system's effectiveness, the usability of the gamification features, and their overall engagement with the model. The interview took approximately 20 minutes to complete (see section 8.2.3.2 below).

8.2.3 Materials

The study utilised a combination of digital tools and materials, which are described in the following sub-sections.

8.2.3.1 The ML-ABM Simulation

This was the central material of the study, designed to simulate classroom management scenarios integrated with gamified features. The hybrid ML-ABM (see Chapter 6) allowed teachers to interact with gamification elements such as ‘Statistics’, ‘Time’, and ‘Progression’, integrated into the system. Moreover, participants were asked to show their perceptions on the potential of integrating three more gamification elements, including ‘Badges’, ‘Points’, and ‘Medals’ into the system in the future. Participants were able to see the results of their decisions in real-time, with outcomes displayed.

8.2.3.2 Interviews

As mentioned above, participants were interviewed at the end of their interaction with the hybrid ML-ABM. The interview included both closed and open-ended questions listed in the Table 23, aimed at capturing participants' perceptions of the effectiveness of the gamification elements, and the perceived impact on teacher engagement.

Table 22 List of questions used in teacher interviews

#	Interview Question	Answers/Options	Type

Q1	From your perspective as a teacher, which of the following features has the highest effect over student performance?	• Initial Knowledge (Start Math Start Read) • Disruptive behaviour score: Inattentiveness Hyperactivity and Impulsiveness. • Financial status (Free School Meal).	Multiple Choice Question
Q2	"Do you find these results to be consistent with your expectations, and could you please explain whether you agree or disagree with the findings, and why?"		Open Ended
Q3	Do you believe that incorporating various gamification elements into the ABS system could enhance your motivation to utilize it effectively in your teaching practice?	Yes/No/May Be	Multiple Choice Question
Q4	Reflecting on the gamification elements currently embedded in the system, which specific gamification element do you find most engaging or valuable for your instructional objectives?	Statistics. Time. Progression.	Multiple Choice Question
Q5	Among the following suggested gamification elements proposed for inclusion in the system, which one do you believe would be the most engaging and beneficial for enhancing your teaching experience and student participation?	Leader Boards. Points. Badges.	

Q6	Based on your experience and understanding of gamification, which additional gamification elements would you recommend incorporating into the system to further enhance its effectiveness and engagement for you?		Open Ended

Through this comprehensive approach, the study aimed to critically assess the current gamification implementations and identify opportunities for enhancing the system to better meet the needs of teachers. The following section presents the six steps involved in this study

8.2.4 Study Steps

Step 1: Feature Impact Assessment

To understand their take on the underlying principles of the ML-ABM approach and answer RQ3, teachers were initially presented with a multiple-choice questionnaire that probed which student-related features they perceived as most influential on educational outcomes. The choices included variables such as 'Initial Knowledge' (Start Math, Start Read), 'Disruptive Behaviour Score', 'Financial Status' (Free School Meal), and 'Age'.

Step 2: Predictive Outcomes Evaluation

Following their selections, teachers were shown the predictive impacts on student performance both with and without the selected feature. To deepen the insight, the influence of these features was also demonstrated at doubled values for each of the above feature, facilitating an understanding of their scalability and effect magnitude. Teachers subsequently voted on the most engaging features based on the presented outcomes.

Upon examining the student performance predictions with and without the incorporation of the top two features identified by the teachers as most influential and observing the effects of doubling the values of these features, a structured inquiry was directed towards the teachers. They were asked to reflect on whether the resultant data aligned with their initial expectations. Specifically, the teachers were requested to articulate their agreement or disagreement with the findings, providing detailed justifications for their perspectives. This approach was designed to elicit insights into the perceived accuracy and relevance of the predictive models in relation to actual educational outcomes.

Step 3: Gamification Element Motivation Inquiry

At this step, teachers were asked whether different gamification elements employed in this study could improve their motivation and effectiveness while using the ML-ABM approach. The study explored whether gamification elements could promote higher engagement with teaching tools, inspire active classroom management, and contribute to better pedagogic results. It was critical to understand what teachers thought of gamification, in order to see if these elements might be practical enough to actually help teachers feel more successful, or if they would just make teaching more interactive and more rewarding. This inquiry aimed to determine how far gamification could be utilised as a significant intervention within contemporary educational settings.

Step 4: Engagement with Existing Gamification Elements

Teachers evaluated the existing gamification elements embedded within the system specifically 'Stat', 'Time', and 'Progression'. They identified which of these elements they found most engaging and valuable for achieving their instructional goals.

Step 5: Proposal of New Gamification Elements

The further exploration involved teachers selecting from the proposed new gamification elements including 'Leaderboards', 'Points', 'Badges'. They were asked to choose which they believed would most significantly enhance their teaching experience and engagement.

Step 6: Open-ended Feedback on Additional Elements

Finally, teachers provided descriptive feedback on any other gamification elements they would recommend incorporating into the system to boost its overall effectiveness and engagement, explicitly excluding any previously used or suggested elements.

8.3.2 Analytical Approach and Data Synthesis

The collected data were analysed using a mixed-method strategy with a combination of qualitative and quantitative methods. This dual approach ensured a comprehensive analysis of the feedback, allowing for detailed insights into how various gamification elements influence teacher perceptions. Through this methodology, the research aims to deliver a robust, evidence-based enhancement of the ABM system, ensuring that gamification strategies are effectively tailored to meet educational needs and improve teaching practices.

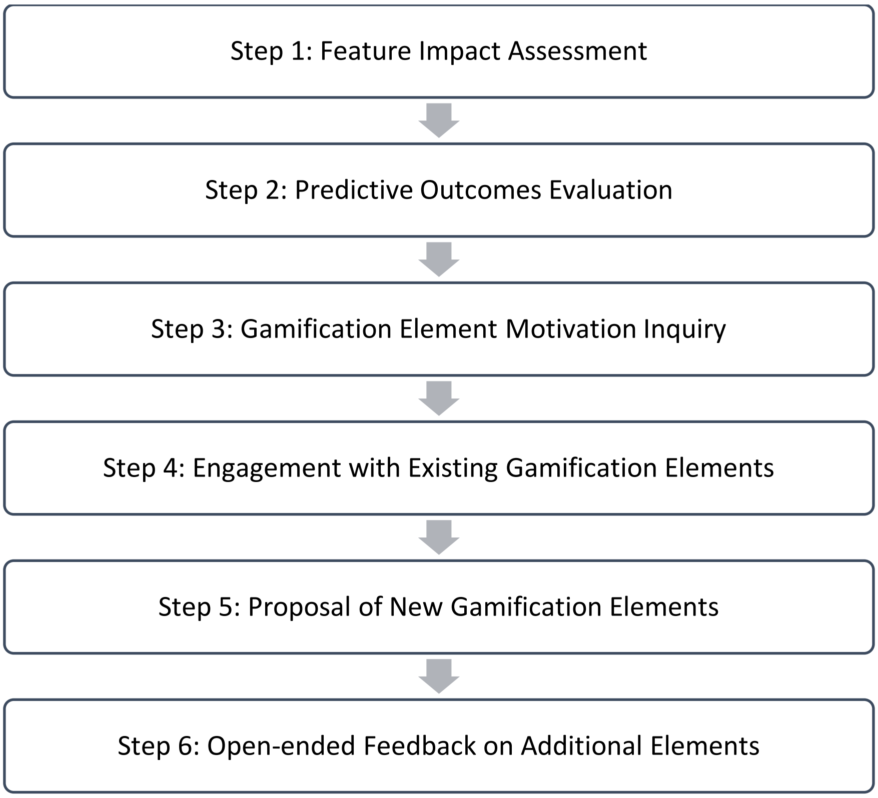

Figure 28 Flow chart of mixed method study steps

8.3 Result Analysis

8.3.1 Feature Impact Assessment

To evaluate the impact of various student-related features on educational outcomes, teachers were asked to identify which feature they believed most significantly influenced student performance. The features presented for selection were 'Initial Knowledge', 'Disruptive Behaviour Score', 'Financial Status', and 'Age'. The teachers were particularly asked the following question during the interview-

"*From your perspective as a teacher, which of the following features has the highest effect over student performance?*"

Based on the responses from the participating teachers, Figure 29 illustrates the findings. Despite being a foundational aspect of learning, only 2 out of 12 teachers identified initial knowledge in mathematics and reading as the most crucial factor affecting student performance. This response may indicate a perception that while foundational knowledge is important, other factors play a more significant role in the day-to-day educational achievements and challenges faced by students. It suggests that teachers might see the potential for students to overcome initial shortcomings through effective teaching strategies and support.

A substantial majority, 9 out of 12 teachers, pointed to disruptive behaviour, encompassing inattentiveness, hyperactivity, and impulsiveness, as the most influential factor on student performance. This overwhelming consensus highlights a critical concern within classroom management, suggesting that disruptive behaviours might significantly hinder student engagement and learning potential. The teachers' responses underscore the importance of behavioural management strategies and interventions in educational settings, reflecting a need to address these behaviours proactively to enhance student academic outcomes.

Interestingly, no teachers chose financial status, represented by eligibility for Free School Meals, as the most influential factor. This lack of selection could imply that, within this group of teachers, economic factors are not perceived as directly impactful on student performance as behavioural and cognitive factors. Yet, such a finding must be embedded in the KSA cultural setting because financial aid to students might look different from the UK. Considering the systemic difference in resource distribution and social status in different counties. It might also reflect a belief that while financial status may affect resources available to students, its direct impact on daily academic performance is less significant compared to behavioural issues.

Only one teacher selected age as the predominant factor influencing student performance, which accounts for 8.33% of the responses. This suggests that age, while being a factor in cognitive and emotional development, it is considered less decisive in influencing academic performance compared to behavioural and foundational knowledge factors. The teacher who chose this might have observed age-related maturity impacting learning outcomes in their educational environment.

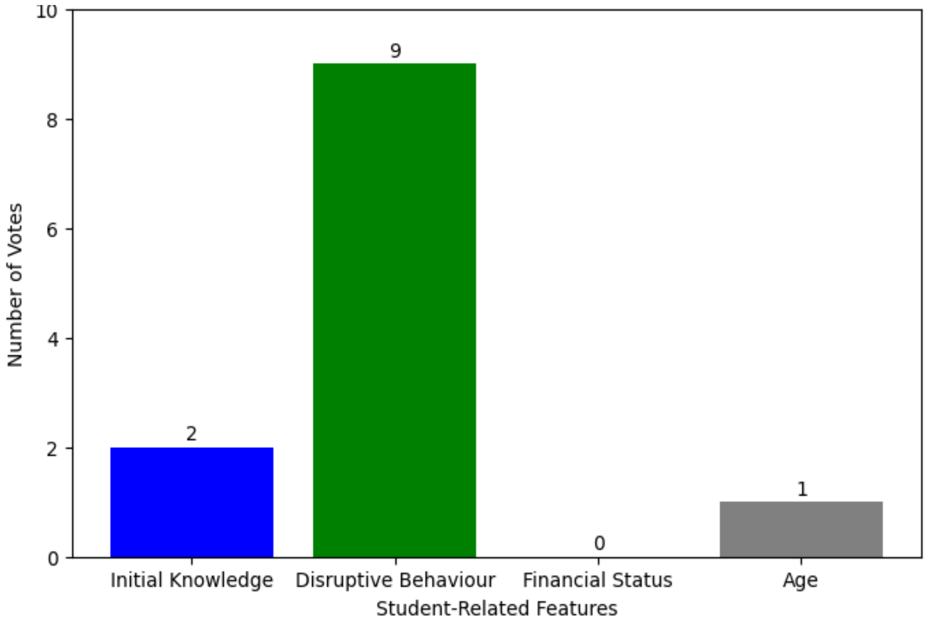

Figure 29 Teacher Perceptions of Factors Affecting Student Performance

8.3.2 Teacher Responses to Student Performance Predictions

Having reviewed the student performance predictions both with and without the top two features the teachers deemed most influential, and the outcomes when the value of these features is doubled (Table 23), teachers were asked:

> *"Do you find these results to be consistent with your expectations, and could you please explain whether you agree or disagree with the findings, and why?"*

Table 23 The variations in End Math average scores with alterations in feature values (Doubled)

Feature	**End Math Average Score**

No Initial Knowledge	21.6
Double Initial Knowledge (Start Math Start Read)	53.63
No Disruptive Behaviour	42.23
Double Disruptive Behaviour	33.51

Out of 11 teachers surveyed, 7 agreed that the results were consistent with their expectations, while 4 disagreed. Responses from the participating teachers are presented in Table 24 and Table 25. Based on a thematic-analysis of qualitative feedback, the following 3 common themes were identified for each category of agreement and disagreement.

Table 24 Feedback of teachers who agreed that the results of the study aligned with their expectations. The comments highlight key observations such as the significance of initial knowledge, the impact of disruptive behaviours, and the role of data-driven approach

Teacher No.	**Feedback**
1	*"The dramatic improvement in scores when initial knowledge is doubled clearly shows how crucial early learning foundations are. It is evident that robust early education can set the stage for future academic success."*
3	*"Seeing the decline in performance with increased disruptive behaviour resonates with my experience. Managing such behaviours is vital; without addressing them, even the brightest students struggle."*
4	*"These results validate what many of us have observed: solid initial knowledge and a stable learning environment significantly boost student outcomes."*
5	*"The findings make perfect sense. Students who start strong tend to stay strong. We need to focus more on building a firm foundation from the start."*

6	*“The negative impact of doubling disruptive behaviour on performance is not surprising. It underlines the need for effective behavioural management strategies in our schools.”*
8	*“I agree with the results because they align with the data-driven approach we aim for in education today. Understanding these relationships helps us tailor our teaching methods more effectively.”*
10	*“These predictive outcomes offer a clear rationale for prioritizing initial academic readiness and creating supportive, focused learning environments.”*

Table 25 Feedback of teachers who disagreed with the study’s results. The responses reflect concerns about the overemphasis on initial knowledge, the perceived oversimplification of disruptive behaviours, and the methodology used in scaling features.

Teacher No.	**Feedback**
2	*“While initial knowledge is important, this overemphasis might lead us to neglect the holistic development of the child, which is equally crucial for learning.”*
7	*“Doubling the numbers does not necessarily reflect real-world scenarios. Education is not just about mathematics; it is about understanding each student's unique context and needs.”*
9	*“The model simplifies disruptive behaviour too much. We need to look deeper into why students behave disruptively rather than just how it affects their scores.”*
11	*“I'm sceptical of the methodology used to double the features. Real educational change does not happen through such straightforward scaling but through nuanced, gradual improvements.”*

Common Reasons for Agreement

- *Increased Performance with Enhanced Initial Knowledge:* The significant rise in scores from 21.6 to 53.63 when initial knowledge was doubled underscored the foundational importance of initial literacy and numeracy skills in mathematical achievement. Teachers who agreed felt this affirmed the critical role of early education fundamentals in subsequent academic success.

- *Decreased Performance with Increased Disruptive Behaviour:* The reduction in scores from 42.23 to 33.51 with a doubling of disruptive behaviour validated the teachers' observations that behavioural issues significantly hinder learning processes. This was seen as a confirmation that managing student behaviour is crucial for educational attainment.
- *Empirical Support for Pedagogical Interventions:* Teachers who agreed with the findings viewed them as empirical support for targeted interventions focusing on reinforcing initial knowledge and managing disruptive behaviours. They believed that these strategies could be effectively used to enhance educational outcomes.

Common Reasons for Disagreement

- *Overemphasis on Initial Knowledge:* Some teachers felt that while initial knowledge is important, the doubling effect depicted an unrealistic and overly deterministic view of its impact on learning outcomes. They argued that student engagement, teaching quality, and other socio-emotional factors also play significant roles.
- *Concerns Over Simplistic Interpretation of Disruptive Behaviour:* Disagreeing teachers were concerned that the study might oversimplify the complex nature of disruptive behaviour by directly linking it to poorer academic performance without considering underlying causes such as socio-economic factors, learning disabilities, or home environment.

- *Methodological Scepticism:* A few teachers questioned the methodology of doubling feature values, suggesting that such a mechanical increase might not accurately represent real-life scenarios where incremental changes and nuanced interventions are more common.

8.3.3 Incorporation of Gamification Elements

In a decisive response to the query regarding the potential benefits of incorporating gamification elements into the ABM system, all twelve participating teachers expressed a positive outlook for the following question-

"Do you believe that incorporating various gamification elements into the ABM system could enhance your motivation to utilise it effectively in your teaching practice?"

The unanimous response indicates a broad acknowledgment of the potential benefits that gamification can bring to educational environments. These benefits likely include increased engagement, improved motivation, and a more interactive learning experience for both teachers and students. Teachers appear to be aware that gamification can transform traditional teaching methods, making learning processes more enjoyable and dynamic. The fact that all respondents see value in gamification also suggests a widespread perception that current educational tools and methods could be significantly enhanced.

8.3.4 Teacher Preferences for Gamification Elements

To further understand the effectiveness and appeal of specific gamification elements within my ABM system, twelve teachers were asked the following question:

"Reflecting on the gamification elements currently embedded in the system, which specific gamification element do you find most engaging?"

This question was asked to reflect on the current gamification elements embedded in the system and identify which they found most engaging or valuable for achieving their instructional objectives. Their choices included 'Statistics', 'Time', and 'Progression'. Teachers' preferences regarding the most engaging element from the three existing gamification elements are illustrated in Figure 30 while Figure 31 shows the gamification elements in the system encompassing 'Statistics', 'Time', and 'Progression'.

The overwhelming majority of teachers (10 out of 12 teachers) identified 'Statistics' as the most engaging gamification element. This preference suggests that teachers value the ability to access real-time data and feedback on student performance and learning patterns. Statistics likely help teachers track student progress more effectively, adjust teaching strategies based on empirical data, and potentially predict student outcomes. However, this might point to cultural and contextual variations. For example, as seen in the earlier study described in Chapter 7, preferences among gamification elements leaned towards elements that encourage direct student engagement, like badges or points. In this study conducted in KSA, we observe a teaching culture that emphasises measurable outcomes and data in performance assessment. That contrast highlights the role of cultural factors and local pedagogical practises in determining which gamification features are most useful. The high vote count for this feature underscores its perceived utility in enhancing instructional efficacy and facilitating a data-driven teaching approach.

Interestingly, none of the teachers favoured the 'Time' element, which often introduces time pressure into teaching activities. This unanimous disfavour indicates a possible concern that time constraints could induce stress among teachers, detracting from the productivity rather than enhancing it. Teachers might believe that productivity should be self-paced to accommodate individual differences in working speeds. This response could reflect a preference for fostering a more relaxed and thoughtful working environment over one that prioritises speed.

However, a smaller segment of the group (2 out of 12 teachers) found 'Progression' to be a valuable gamification element. This feature, which often involves visualising advancement through content via levels or milestones, was seen as beneficial by those who perhaps value a structured working path. Progression can motivate teachers by clearly marking their achievements and providing explicit goals to work towards. However, the relatively low number of votes suggests that while seen as beneficial, it may not be as crucial or universally appealing as real-time statistics.

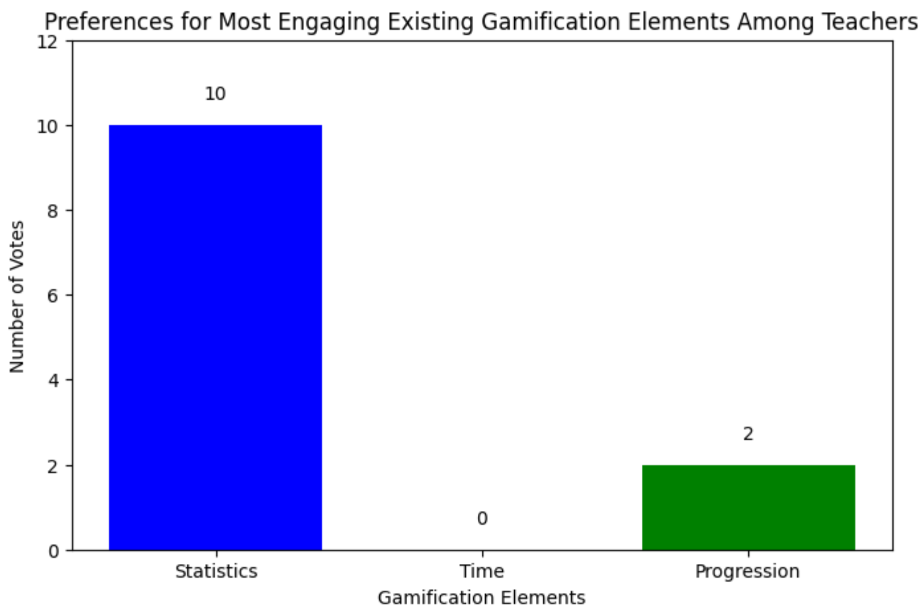

Figure 30 preferences for most engaging existing gamification elements among teachers

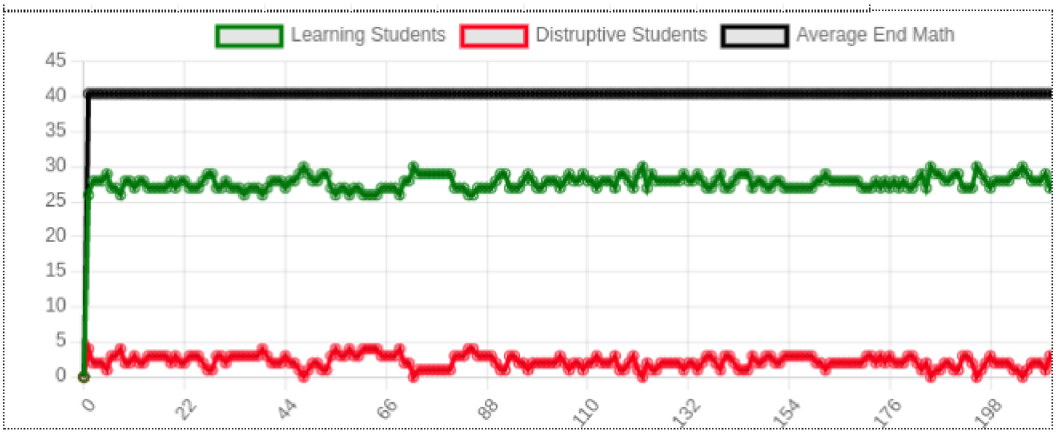

Figure 31 Gamification elements in the system 'Statistics', 'Time', and 'Progression'

8.3.5 Teacher Preferences for Proposed New Gamification Elements

In response to the question,

"Among the following suggested gamification elements proposed for inclusion in the system, which one do you believe would be the most engaging and beneficial for enhancing your experience?",

Twelve teachers were asked to ascertain their preferences: Leaderboards, 'Points', and 'Badges'. This analysis aims to unpack the implications of their choices and understand the perceived value of these gamification elements for teachers. Findings from the teachers' feedback are visualised in Figure 32.

The majority of teachers (7 out of 12 teachers) selected ' leaderboards the most engaging gamification element to be included in the ABM system. This preference indicates a significant recognition of the competitive aspect that leaderboards bring to the working environment. Teachers likely perceive that such an element could drive motivation and engagement among them by publicly acknowledging their achievements, thereby fostering a competitive spirit that encourages teachers to excel. Leaderboards could also facilitate a transparent metric for teachers to track performance progress over time.

In addition, 'Points' received a substantial number of votes (4 out of 12), suggesting that teachers appreciate the incremental and immediate feedback that points offer. This gamification element allows teachers to receive instant recognition for their efforts, potentially increasing their engagement and persistence in working activities.

Moreover, the low preference (only 1 vote) for 'Badges' might indicate that teachers feel this element provides less immediate or clear value compared to the other options. While badges are a form of recognition and achievement, they may be perceived as less influential in motivating performance or might be seen as too sporadic or arbitrary to effectively drive consistent engagement.

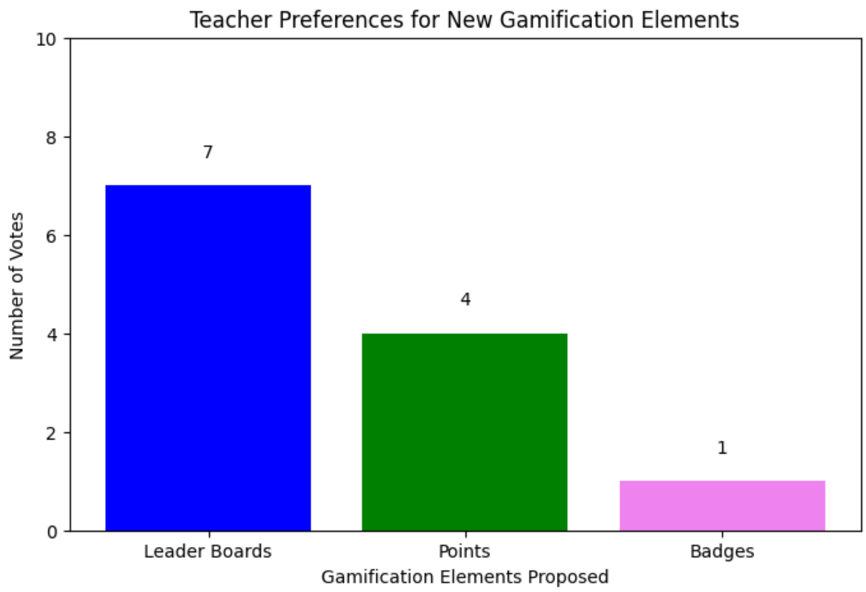

Figure 32 Teacher preferences for new gamification elements

8.3.6 Open-ended Feedback on Additional Gamification Elements

To gain a deeper understanding of the participating teachers' perceptions and recommendations for the inclusion of more gamification elements, teachers were asked to provide their responses to the following open-ended question:

"Based on your experience and understanding of gamification, which additional gamification elements would you recommend incorporating into the system to further enhance its effectiveness and engagement for you?"

Twelve teachers provided their feedback, which is documented in Table 26.

This analysis delves into the qualitative responses from twelve teachers who were asked to recommend additional gamification elements for enhancing the effectiveness and engagement of the ABM system. Their suggestions reflect a deep understanding of gamification and its potential to foster a more dynamic and supportive teaching environment. This thematic analysis identifies and discusses the five key themes (gamification elements) emerging from the responses: 'Social Competition', 'Social Cooperation', 'Social Reputation', 'Performance Acknowledgement', and 'Performance Points'.

The recommendations provided by the teachers suggest a strong belief in the power of gamification to transform teaching environments. The themes of Social Competition, Social Cooperation, Social Reputation, Performance Acknowledgement, and Performance Points capture a comprehensive view of how these elements can be strategically integrated to enhance both teaching effectiveness and engagement. This analysis underscores the need for educational systems to adopt a more nuanced and teacher-centric approach to gamification, ensuring that these elements align with the overarching goals of education and foster an environment conducive to teaching and learning and professional development.

8.3.6.1 Social Competition

Two teachers highlighted the benefits of integrating social competition elements into the teaching environment. The theme revolves around the idea of using competitive frameworks to stimulate teacher engagement and professional growth. Teacher 5 emphasised the motivational aspect of social competition, suggesting that competition could enhance creativity and instructional quality. The notion is that a competitive environment encourages continual improvement and the adoption of innovative teaching methods. Teacher 9 advocated for a leaderboard system to recognise and share best practices. This approach not only fosters professional growth but also promotes a culture of excellence and collaboration, where teachers can benchmark their methods against peers, enhancing overall educational standards.

8.3.6.2 Social Cooperation

Another recurring theme involved the endorsement of social cooperation mechanisms by two participants, highlighting the potential of collaborative efforts in teaching settings. One teacher proposed enhancing collaborative projects across departments, which would help in building a supportive teaching community and enrich the collective educational experience. Another teacher focused on the practical benefits of joint lesson planning and collaborative grading frameworks, especially in managing disruptive behaviours. This cooperative approach allows for pooling insights and strategies, thus optimising the learning environment and behaviour management.

8.3.6.3 Social Reputation

Three teachers expressed a preference for implementing a social reputation system to acknowledge and motivate innovative teaching methods. A teacher saw value in a system that rewards creativity and pedagogical innovation, suggesting that such recognition would boost teacher confidence and inspire continued pedagogical experimentation. Two teachers both highlighted the motivational impact of publicly acknowledging teaching strengths, suggesting that a reputation framework could enhance knowledge sharing and collaborative engagement across disciplines.

8.3.6.4 Performance Acknowledgement

The theme of performance acknowledgement was underscored by two teachers who appreciated regular and personalised feedback on their teaching effectiveness. Some teachers advocated for a feedback loop that recognises lesson delivery and student engagement, which could foster a reflective teaching culture cantered around growth and improvement. Others emphasised the importance of acknowledging incremental achievements, which helps teachers assess their progress and encourages continuous innovation.

8.3.6.5 Performance Points

Three teachers recommended incorporating a performance points system that rewards various professional development activities. These teachers supported the idea of a points system that acknowledges and rewards efforts in professional development, mentorship, and innovative teaching activities. Such a system would not only motivate teachers but also facilitate their professional growth and engagement by providing tangible rewards for their contributions to the educational community.

Table 26 Various gamification elements recommended by teachers

Gamification Element	**Teacher No.**	**Feedback**
Social Competition	5	*"Incorporating social competition would create a lively environment that motivates me to achieve my teaching objectives more efficiently. Competing with fellow teachers will stimulate creativity and allow me to learn new techniques. This sense of healthy competition pushes us to continuously improve and enhances overall instructional quality."*

	9	*"A leaderboard system, where teachers earn recognition for best practices, will encourage a culture of sharing and continuous development. The opportunity to benchmark my own teaching methods against others will foster professional growth and highlight areas where I can refine my strategies."*
Social Cooperation	1	*"Collaborative projects across departments could be enhanced through social cooperation elements. Working together with colleagues on interdisciplinary projects would foster a supportive teaching community and enable me to gain valuable insights from peers. This cooperation helps us build a united learning environment, benefiting not only teachers but also the entire school culture."*
	4	*"Incorporating joint lesson planning or collaborative grading frameworks would be immensely beneficial, particularly for addressing and managing disruptive student behaviours. This allows me to consult and brainstorm with colleagues, making the teaching process a collective effort rather than an isolated one. By sharing strategies and insights, we leverage our collective expertise to develop more effective approaches to behaviour management, thereby building a stronger, more supportive teaching environment."*
Social Reputation	3	*"A social reputation system that recognises innovative teaching methods would be motivating. Being acknowledged for developing creative lesson plans or adopting new technologies would bolster my confidence. This recognition reinforces the idea that thoughtful pedagogy matters and inspires me to continue experimenting with new approaches."*
	6	*"Highlighting teachers' strengths through a public reputation framework can be a strong motivator. If my accomplishments are highlighted, it would inspire me to continue working diligently while providing a template for others to follow."*
	11	*"A reputation board that showcases expertise across different subjects would encourage us to collaborate more effectively. If I know my specific strengths are recognised, I'm more likely to*

		engage in knowledge sharing and learn from other colleagues who excel in complementary skills."
Performance Acknowledgement	2	*"Regular recognition of lesson delivery and student engagement metrics would be highly motivating. This feedback loop acknowledges our efforts and allows me to identify what's working and where I can improve. The acknowledgment fosters a culture of reflective teaching that prioritises growth."*
	7	*"A personalised acknowledgment system that celebrates incremental achievements would be beneficial. Receiving specific feedback on new initiatives helps me gauge my progress and solidifies the value of continuous innovation."*
Performance Points	8	*"A performance points system that rewards participation in professional development activities would be invaluable. It encourages me to stay updated on the latest teaching strategies and curriculum trends. Accumulating points could translate into bonuses or classroom resources, which would further motivate me to excel."*
	10	*"A structured points system that acknowledges participation in mentorship or peer coaching programs can be highly engaging. If my efforts to support new teachers are recognised with points that contribute to professional development opportunities, I would be inclined to actively participate and grow alongside my colleagues."*
	12	*"A points-based system that encourages lesson innovation and extra-curricular involvement would be ideal. It ensures my creative efforts are recognised, giving me tangible rewards while reinforcing my commitment to holistic teaching."*

8.4 Discussion

The current findings resonate significantly with those presented in Chapter 7 of this thesis, where the efficacy of various gamification elements like badges, points, and medals was explored within the context of the CamaleOn system for students. Both studies underscore the potent impact of gamification on enhancing user engagement and educational outcomes. However, a deeper comparison reveals similarities and differences arising from the contrasting educational and cultural contexts of the studies. Chapter 7 concentrates on gamification in a Brazilian online secondary educational environment (CamaleOn), where badges, points, and medals were particularly useful to promote student engagement. However, unlike the Brazilian study, the current study takes place in Saudi Arabia and is focused on gamification in a primary (face to face) educational setting, where the teachers are the users. Educators here preferred data-driven strategies and competitive frameworks, and so elements like 'Statistics' and 'Leader Boards' emerged as popular. This divergence in preferences points to the importance of cultural and contextual factors. In the Brazilian context, the emphasis on student-oriented gamification elements is consistent with a learning culture that values individualised learning and motivational rewards for students. On the contrary, the strong preference for 'Statistics' in Saudi Arabia implies a teaching culture that favours measurable outcomes and structured pedagogical strategies. Both 'Leader Boards' in both contexts reflect its wide appeal as a motivational tool, but used in different ways: in Brazil, to encourage student competition, and in Saudi Arabia, to lift the engagement of teachers through display of comparative performance metrics.

The application of Toda’s TGEEE has proven instrumental in structuring the gamification strategies within the ABM system. The taxonomy's emphasis on dynamics, mechanics, and aesthetics allows for a nuanced integration of gamification elements that cater specifically to educational needs. User feedback and engagement metrics from the current study indicate that elements categorised under mechanics, such as 'Statistics' and 'Progression', are particularly effective. These elements not only align with the pedagogical goals but also enhance the educational interface, making learning and teaching both engaging and efficient. The introduction of new gamification elements such as 'Leaderboards, 'Points', and 'Badges' was received positively, as evidenced by the enthusiastic teacher responses. The adoption of 'Leaderboards, which gained the highest approval, marks a notable success, reflecting a significant increase in both teacher engagement in this study and student participation in the prior study (see section 7.3). This success can be attributed to the competitive yet motivational environment fostered by leaderboards.

Integrating the results from various stages of the study, it is evident that certain gamification elements have a more pronounced influence on engagement than others for teachers. 'Statistics' and 'Leaderboards' stand out as particularly impactful, which can be theoretically backed by their classification within the TGEEE Wheel under mechanics and dynamics respectively. These elements enhance the learning and teaching experience by providing clear goals, immediate feedback, and a sense of achievement, which are crucial for maintaining high engagement levels. While the study provides valuable insights, it is not without limitations. One of the primary constraints was the relatively small sample size of teachers, which may not fully capture the diverse range of perspectives and experiences within the educational sector. However, qualitative data was also collected to better support the findings. Additionally, the reliance on self-reported data can introduce biases that might affect the accuracy of the findings. Further research involving a larger cohort and perhaps a more varied array of gamification elements could help in validating and expanding upon these results.

8.5 Epilogue

This chapter demonstrates the integration of gamification into a hybrid ML-ABM approach tailored for teachers, building on the findings from Chapter 7. By incorporating gamified elements, such as ‘Statistics’, ‘Time’, and ‘Progression’, the study explored how these tools can enhance teacher engagement, improve classroom management, to address disruptive behaviours. The feedback from educators highlighted the potential of gamification to motivate teachers, and consequently supporting data-driven decision-making in educational settings. A key contribution of this chapter is its demonstration of how gamification strategies can transition from focusing on student engagement, as seen in Chapter 7, to empowering teachers in managing classroom dynamics. This chapter also showcases the importance of aligning gamification elements with pedagogical objectives, guided by Toda’s TGEEE framework. By bridging theoretical principles with empirical findings, the study created a systematic approach to designing effective gamified tools. These findings contribute to the methodology in this thesis, by establishing a framework for integrating gamification into the hybrid ML-ABM. The limitation of these findings include the relatively small sample size of 12 teachers, which may not fully capture the diversity of perspectives and teaching practices within broader educational contexts. Additionally, the study was conducted in Saudi Arabia, and the cultural and educational practices unique to this setting may limit the generalisability of the findings to other regions. Finally, the reliance on self-reported data could introduce bias, as participants’ responses may not fully reflect their actual experiences or behaviors when using gamification elements. Future work should explore further validation across varied educational environments and cultures, refining the model to ensure its sustainability and effectiveness. This chapter sets the stage for the final synthesis of the research findings, positioning gamification as a transformative element in both teaching and learning. Building on the findings presented in the previous chapters, Chapter 9 critically examines their implications, situates them within the broader literature, and explores their significance for research field.

CHAPTER 9

9 Discussion

9.1 Prologue

In this thesis, a simulation model is proposed and investigated, to aid teachers in the management and mitigation of disruptive student behaviour within classroom and online learning environments. The simulation is based on ABM, designating teachers and students as agents within a model that emulates a classroom environment. Agents interact and simulate various classroom characteristics, encompassing disruptive behaviours and potential mitigation strategies. The simulation enables teachers to trial scenarios of disruptive behaviours and solutions before implementation, minimising the need for trial-and-error approaches. To understand teachers' requirements for managing disruptive behaviour, semi-structured interviews were conducted, exploring their strategies and classroom management techniques. Teachers' responses and observations were analysed and integrated into the simulation model through co-design, ensuring it effectively mimicked a classroom setting.

The ABM was combined with ML models to bolster the effectiveness of the strategies. ML was utilised in three stages: pre-processing, agent behaviour and decision-making, and post-simulation output processing. During pre-processing, ML algorithms trained datasets to model the anticipated student and teacher behaviour under varying circumstances. The trained models forecasted changes in student behaviour and their impact on the learning process. ML contributed by modelling the expected behaviour based on circumstances and enabled teachers to benefit from their own or their colleagues' experiences. Moreover, it facilitated the incorporation of gamification in controlling disruptive behaviour. Throughout the agent behaviour and decision-making process, ML algorithms determined suitable strategies based on the simulated situation. This thesis presents a novel integration of Agent Based Modelling (ABM), Machine Learning (ML), and gamification to explore and mitigate disruptive student behaviour in educational environments. Unlike previous studies that separately examined these techniques, this research uniquely combines them into a comprehensive framework, providing a data-driven approach to classroom management.

By addressing the eight formulated RQs, this thesis advances the understanding of how ABM is utilised, to explore disruptive behaviours' impact on learning, how ML can enhance classroom simulations, and how gamification strategies optimise engagement. This chapter synthesises the findings, highlighting their theoretical and practical implications as well as limitations.

Initially, this chapter provides an overview of fundamental aspects pertinent to this thesis. It includes a demonstration of the significance and impact of disruptive behaviour in both conventional classroom-based and online learning environments, discusses issues concerning peers and disruptive behaviours, and explores the utilisation of gamification to mitigate such behaviours. Subsequently, it examines the summary of findings presented in the literature review and related works. This comprises an overview of the thesis's outcomes, together with its overarching conclusions and contributions. Additionally, it considers the limitations associated with ABM and ML in education, as well as their exploration into student performance in relation to disruptive behaviour and gamification tactics. Finally, potential avenues for future research are identified, with broader research implications.

9.2 General Overview

9.2.1 Disruptive Behaviour and Student Performance in Educational Settings

Disruptive behaviour in educational settings represents a significant barrier to effective teaching and learning. This discussion explores the interplay between student disruptive behaviour and academic performance, emphasising the role of teachers in managing these challenges. Teachers are pivotal in either mitigating or exacerbating student disruptions. Teaching strategies that engage students and relate content to their lives can significantly reduce disruptive behaviour [56]. Conversely, monotonous and irrelevant teaching methods can increase boredom and disruptive behaviour, leading to a decline in academic performance [143].

Disruptive behaviour stems from both internal desires, such as the need for attention, and external pressures like peer influence and societal expectations [127]. These factors often interact, making it challenging for educators to address the root causes of disruption. Educational interventions must therefore be multifaceted, addressing both the psychological needs of students and the environmental factors influencing their behaviour. Disruptive behaviours have far-reaching consequences, affecting not only the psychological well-being of students and teachers but also academic outcomes. Research indicates a direct negative correlation between disruptive behaviour and academic performance, particularly in subjects like mathematics [6]. This underscores the importance of addressing disruptive behaviour to enhance educational outcomes.

To combat disruptive behaviour, an integrative approach involving both educational and psychological strategies is essential. Strengthening the perceived cognitive ability and the importance of schoolwork can result in enhanced student involvement and fewer disruptions [58, 196]. Furthermore, establishing a positive teacher-student rapport and implementing proactive management strategies can significantly mitigate disruptive behaviours [129]. Overall, the intersection of disruptive behaviour and student performance in educational settings is complex, influenced by psychological, pedagogical, and social factors. Effective management of disruptive behaviour in schools requires a holistic approach that considers the cognitive and emotional needs of students, as well as the pedagogical approaches employed by educators. By fostering an engaging, relevant, and supportive learning environment, educators can significantly improve both student behaviour and academic performance.

9.2.2 Predictive Modelling with Machine Learning for Disruptive Behaviour Management in Classroom Settings

In the realm of educational technology, the use of predictive modelling and ML to manage disruptive behaviours in classroom settings has emerged as a transformative approach. This methodology not only facilitates the identification of at-risk students but also enhances the development of tailored interventions that can significantly improve both academic outcomes and classroom dynamics.

ML models are adept at processing vast and varied data sets, including student behavioural patterns and academic records [94]. These models, which employ techniques ranging from decision trees and random forests to neural networks, allow for a nuanced analysis of how specific disruptive behaviours such as inattentiveness, hyperactivity, and impulsiveness correlate with academic achievements [12]. The predictive power of these models lies in their ability to uncover non-linear relationships and intricate patterns that traditional statistical methods might overlook.

For instance, regression analysis can quantitatively assess the impact of disruptive behaviours on students' academic performance, providing educators with clear metrics on which behaviours most strongly predict negative educational outcomes [17]. This insight is invaluable for developing targeted strategies that address these specific behaviours. One of the most significant advantages of employing ML in educational settings is the capability to develop customised intervention strategies. By analysing patterns in student behaviour and academic performance, ML models can predict which students are likely to encounter academic difficulties due to their disruptive behaviours. This early identification allows for timely and specific interventions, such as one-on-one tutoring or behavioural therapies, which are tailored to the individual needs of each student [12].

Moreover, these predictive models are instrumental in optimising resource allocation within educational institutions. By identifying classes or grade levels with a higher prevalence of disruptive behaviours, schools can strategically allocate resources like additional counsellors or special education services, ensuring that support is provided where it is most needed [12]. Predictive modelling also plays a critical role in shaping educational policies. Insights derived from these models can inform more effective strategies for managing disruptive behaviours in schools. For example, data-driven policies might include mandatory training for teachers in behavioural management techniques, thus fostering a more conducive learning environment for all students [12]. However, the use of predictive models and ML in education raises several ethical considerations, including concerns over the protection of data privacy and the possibility of biases in algorithmic recommendations. It is imperative that these technologies are implemented in a transparent and ethically responsible manner to maintain trust among all stakeholders including students, parents, educators, and policymakers.

In addition to typical ML models, XAI plays a crucial role in the application of predictive modelling within educational settings, particularly in managing disruptive behaviours and monitoring student performance. XAI techniques [144], such as LIME and SHAP, make the outcomes of ML models transparent and comprehensible to human users. This explainability is essential for educators and stakeholders to trust and effectively utilise AI predictions. For instance, by understanding the specific contributions of behavioural factors like inattentiveness or hyperactivity to predicted academic outcomes, educators can tailor interventions that address these issues with precision. Furthermore, the clarity provided by XAI fosters a deeper understanding of how various disruptive behaviours influence learning outcomes, thereby enabling more targeted and effective educational strategies. This approach not only enhances the efficacy of interventions but also builds trust among educators, students, and parents, ensuring that the decisions based on AI are both justified and beneficial to the educational process.

In summary, integrating ML and predictive modelling into educational practices offers a robust tool for managing disruptive behaviours in classroom settings. By enabling a deep understanding of the relationships between student behaviours and learning outcomes, these technologies facilitate the creation of a more personalised, supportive, and effective educational environment. As these models continue to evolve, their potential to transform educational strategies grows, promising significant advancements in the way educational institutions address the challenges of disruptive behaviour.

9.2.3 ABM and Machine Learning Combination for Disruptive Behaviour Management in Classroom Settings

The integration of ML with ABM represents a forward-thinking approach in addressing disruptive behaviour in educational settings. This synergy harnesses the strengths of both disciplines to create dynamic, adaptive models that provide deep insights into student interactions and the effectiveness of classroom management strategies. ML excels in identifying patterns and making predictions from large datasets, thus enhancing the capability of ABMs to simulate complex systems such as classrooms. These simulations can include varying student behaviours, interaction dynamics, and the effect of different educational strategies [94]. For instance, ABMs can detail the emergent phenomena that occur from student interactions, while ML can predict the outcomes of these phenomena on learning processes and behaviour management [87].

Incorporating ML algorithms within ABMs allows the models to not just operate under predefined rules but to adapt based on ongoing data analysis. This aspect is particularly crucial in educational settings where student behaviour and learning outcomes can fluctuate significantly due to various factors including teaching methods, peer interactions, and individual psychological states [94]. Such models can simulate how changes in classroom strategies might improve engagement and reduce disruptive behaviours, offering a potent tool for educational research and policy-making. The combined use of ABM and ML not only increases the predictive accuracy of educational models but also enhances their adaptability. This is critical for developing interventions that are responsive to the evolving dynamics within a classroom. Models can forecast how students might respond to different teaching styles or classroom configurations, helping educators and administrators to make informed decisions that promote a conducive learning environment [94]. Integrating ML techniques such as reinforcement learning within ABMs can help simulate how agents (students and teachers) learn and adapt over time, reflecting more accurately the real-world learning processes. Predictive analytics further aid in forecasting and planning, ensuring that educational strategies are both effective and timely [121].

However, there is a notable gap in research in the application of ML techniques to predict and manage disruptive behaviours in classrooms. Traditional methods often rely on retrospective analyses and static intervention strategies, which may not account for the dynamic nature of classroom interactions. Current research frequently overlooks the potential of real-time data analytics to identify patterns and trends in disruptive student behaviour, which could inform more timely and effective interventions. This thesis leverages ML algorithms (see Chapter 0) to predict disruptive behaviours based on historical and real-time data, providing educators with actionable insights to mitigate disruptions before they escalate. Furthermore, another notable gap in the utilisation of ABM to simulate complex classroom environments and interactions. Existing studies have predominantly focused on linear models and static analyses, which fail to capture the nuanced and emergent behaviours that arise from the interactions between students, teachers, and their environment. ABM offers a powerful tool to model these interactions dynamically, allowing for the exploration of various scenarios and the identification of effective strategies for managing classroom behaviour and enhancing student performance. This thesis focuses on the influence of disruptive students, teacher, and peer characteristics on learning outcomes. Disruptive behaviours, such as inattentiveness and hyperactivity, can significantly hinder the learning environment. In a novel approach towards mitigating disruptive behaviour, the research in this thesis aims to quantify the extent to which these behaviours affect other students' performance and classroom dynamics. By identifying the specific impacts of different types of disruptive behaviours using ML techniques combined with ABM simulation (see Chapter 0), this research can inform targeted interventions to mitigate their negative effects.

9.2.4 *Gamification* and Predictive Modelling Combination for Disruptive Behaviour Mitigation in Educational Settings

Combining gamification, predictive analytics, and predictive modelling offers a compelling strategy for addressing disruptive behaviour in educational environments. This multi-faceted strategy leverages the motivational allure of gamification, the predictive power of analytics, and the foresight of modelling to create an engaging and harmonious learning environment. Gamification involves introducing game elements such as points, badges, and leaderboards into educational contexts to increase student engagement and motivation. Early research by Deterding et al. [51] and further studies by Kapp [97] underline the potential of gamification to transform learning from a passive activity into an interactive, rewarding experience. These elements, when well-implemented, do not merely captivate students' attention; they enhance intrinsic motivation, making the learning process both enjoyable and challenging. For instance, Kapp [97] empirically demonstrated that such strategies could significantly boost student participation and sustain interaction with course materials, suggesting a direct correlation between gamification and enhanced learning outcomes.

Despite the expansive reach of MOOCs, they often struggle with high dropout rates and low student engagement. Gamification has been identified as a solution to these issues, providing a more dynamic and interactive educational experience that can help sustain student interest and participation over time. For example, Carrera and Ramírez-Hernández [37] highlighted that the impersonal scale of MOOCs often dilutes student motivation, an issue that gamification can address by fostering a sense of community and achievement through interactive and competitive elements. Predictive analytics and modelling further extend the capabilities of gamification by allowing educators to anticipate student behaviours and tailor interventions accordingly. Predictive models use historical data and real-time interactions to forecast outcomes and inform the design of gamification elements that best meet students' needs. Jordan and Mitchell [94] noted the efficacy of such models in educational settings.

Recently conducted systematic literature reviews, such as those conducted by Freitas and Silva [91], provide empirical evidence supporting the effectiveness of gamification in reducing dropout rates and enhancing student satisfaction in MOOCs. These studies often reveal that gamified courses see higher completion rates and more positive feedback compared to non-gamified counterparts, underscoring the practical benefits of integrating game mechanics into educational platforms. While gamification can significantly improve engagement, it is crucial that it does not overshadow the educational content. Kapp [97] emphasised the importance of balancing gamification elements with the curriculum to ensure that learning objectives are met without compromising the integrity of the educational material. This balance is vital for ensuring that the educational benefits of gamification translate into genuine learning outcomes and not just higher engagement metrics.

Addressing the diverse needs of learners is another critical aspect of effectively implementing gamification. Personalised gamification strategies that cater to different learning preferences and cultural backgrounds can maximise inclusivity and effectiveness. Yujia et. al. [85] suggested, based on a systematic literature review, that customising gamification elements to suit varied learner profiles could enhance the inclusivity and effectiveness of educational interventions, making them more engaging and educationally valuable for a global audience. Beyond immediate engagement, predictive models can be used to maintain interest and participation throughout the course duration. These models analyse patterns of engagement and performance to adjust gamification strategies dynamically, ensuring they remain effective and relevant as student needs and behaviours evolve. In summary, gamification has been widely adopted in various educational contexts, to enhance motivation and engagement, as stated. However, empirical evidence on its specific impact within online learning, such as MOOCs, was limited at the time this research was conducted. In my work on gamification in education for MOOCs, I tackle engagement in a gamified online learning, in Chapter 0. As the design of gamified learning systems is usually theory-driven, there is a lack of *runtime feedback, non-gamified scaffolding,* and *under-exploitation of interaction data*. Whilst the theoretical basis is very important in designing purpose-fit gamified systems, in the context of large-scale online learning like MOOCs, it is not feasible to propose a one-size-fits-all design of gamification. For this reason, it is very important to take into account the data generated from the system, in order to better understand the users' interactions, and refine the offering. This thesis employs ML to investigate which gamification elements are most effective in increasing engagement and how these effects can be monitored in real-time, as understanding the real-time impact of gamification elements can enable educators to dynamically adapt their strategies to maximise student engagement (see Chapter 0 , [4]).

Additionally, while the available related scholarly works have explored the integration of gamification across various educational disciplines, there has been a noticeable gap in research specifically focusing on gamification for teachers. Compared to the research in this thesis, which examines how gamification can be utilised to enhance teacher engagement and effectiveness, most studies primarily focus on student outcomes. In Chapter 0, this research addresses this gap, by providing an in-depth analysis of the impact of gamification on teachers, in the context of the ABM model proposed, demonstrating how elements like real-time statistics and leaderboards can significantly improve teaching practices and teacher motivation. This focus on teachers is essential, as their engagement is crucial for the successful implementation of gamified learning environments, and it aligns with the broader aim of enhancing educational outcomes through innovative strategies.

9.3 Findings from the Thesis

9.3.1 RQ1: How can Agent-Based Models be utilised to explore the influence of disruptive students on their peers and the roles of teaching quality, teacher control in a disruptive classroom?

This research investigates the role of ABM in understanding and mitigating disruptive behaviour by simulating classroom interactions integrating real-world behavioural data and predictive analytics to simulate and analyse the impact of disruptive students on their peers and the learning environment. In exploring the impact of disruptive students on learning environments, particularly inattentive or hyperactive students, and the roles of teaching quality and peer dynamics, research available in this thesis employs ABM to simulate classroom interactions and scrutinise these effects over time. This detailed analysis serves to address the critical research questions outlined, drawing on the simulated classroom dynamics to gauge the impact of disruption and the effectiveness of instructional control.

The simulation results in Chapter 4 reveal that the presence of disruptive students, especially those exhibiting inattentiveness or hyperactivity, has a significant and generally negative impact on the learning outcomes of their peers. This is evidenced by the correlation between higher rates of disruptive behaviour and lower average End Math scores in simulated classroom scenarios. Disruptive behaviour tends to interrupt the flow of lessons and can lead to a reduction in effective instructional time, thereby compromising the learning achievements of all students in the environment. This result aligns with traditional educational theories which posit that classroom disruptions diminish the overall educational climate, making it difficult for students to maintain focus and absorb instructional content effectively.

The ABM results shown in section 4.3 indicate that classrooms characterised by higher instances of disruptive behaviours, particularly inattentiveness and hyperactivity, see a significant downturn in academic performance. For instance, the simulation reveals that classes with elevated levels of disruption recorded lower End Math scores, with a notable negative correlation of -0.16 between the percentage of disruptive students and average End Math scores. This suggests that disruption has a tangible, detrimental effect on the learning environment, impeding the academic progress of not only the disruptive students but their peers as well. The detrimental impact of hyperactivity and inattentiveness is further underscored by correlations of -

0.18 and -0.33, respectively, with End Math scores, highlighting the extensive reach of such behaviours on classroom achievement.

Teaching quality and teacher control emerge as significant factors in mitigating the effects of disruptive behaviour. High-quality teaching, characterised by engaging instructional methods and strong classroom management skills, tends to buffer the negative impacts of disruptive behaviours. The results shown in section 4.5 suggest that strong teacher control can significantly mitigate the disruptive tendencies of students, leading to improved academic outcomes. Furthermore, peer characteristics also play a crucial role; supportive peer interactions and positive classroom dynamics contribute to an environment where students can achieve better despite the presence of disruptive behaviours. This finding supports the idea that both teacher efficacy and peer support are pivotal in shaping the learning outcomes in environments challenged by disruptive behaviours.

Teaching quality and teacher control are pivotal in mitigating the adverse effects posed by disruptive behaviours. Findings in section 4.5 illustrate that higher teacher control and better teaching quality can significantly improve academic outcomes, even in challenging environments. For example, the simulation run with maximum values for teacher control and teaching quality resulted in substantially better End Math scores (average score increase from initial 27.43 to 48.56), demonstrating the effectiveness of strong educational leadership in overcoming disruptions. This suggests that proactive teaching strategies and firm classroom management are crucial in cultivating an environment conducive to learning, particularly in settings prone to disruptions.

Teacher control, when effectively combined with positive peer characteristics, significantly enhances the resilience of the classroom environment against disruptive influences. Classrooms where teachers maintain high levels of control and where peers exhibit supportive behaviours demonstrate higher academic performance and less overall disruption. This dual approach not only directly counters the effects of disruptive behaviour but also fosters a supportive learning atmosphere that can buffer the potential academic risks associated with such behaviours.

The combination of teacher control and positive peer characteristics plays a significant role in student achievement. Where teacher control was high, the model showed an improvement in the learning states of students, thereby enhancing overall academic performance. This effect was amplified in environments where peer interactions were also positive, highlighting the symbiotic relationship between teacher-led interventions and peer dynamics. For instance, in simulations where teacher control was paired with constructive peer influence, there was a notable improvement in both behaviour and academic outcomes, illustrating the dual importance of strong leadership and a supportive peer environment in educational success. This research question was thus answered by providing quantitative validation of relationships between disruptive students, teacher control and teaching quality and lower classroom performance using ABM.

9.3.2 RQ2: How can we predict and explore students' learning outcomes based on disruption-related features (Inattentiveness, Hyperactivity, Impulsiveness), using ML models and Explainable Artificial Intelligence (XAI)?

This research explores the potential of ML models in predicting student performance and the application of XAI techniques (SHAP values) to explain the predictions. The discoveries of this thesis (in Chapter 5) identified the ability to predict students' learning outcomes based on features related to disruptive behaviour, such as inattentiveness, hyperactivity, and impulsiveness. Additionally, the use of XAI to explain these relationships offers valuable insights for educational interventions targeting disruptive behaviours. The results are obtained through a methodical approach that combines statistical approaches and ML models to clearly discuss the RQ2.

The methodology employed involves the Jenks' natural breaks method to categorise the intensity of disruptive behaviours and learning outcomes. This classification facilitated the application of various predictive models, with the XGBoost classifier emerging as the most effective, demonstrating superior handling of complex datasets and resistance to overfitting. The study reports that using the XGBoost, Gradient Boosting and AdaBoost algorithms, the average F1-score of 0.91 across the three groups. F1-score presents the harmonic mean of precision and recall and it was notably high in these three algorithms, indicating a robust model performance.

The predictive model incorporated features such as inattentiveness, hyperactivity, and impulsiveness, each weighted according to their influence on learning outcomes. The analysis revealed that inattentiveness was the most significant predictor of lower learning outcomes, followed by impulsiveness and hyperactivity.

This result was quantified using improvement scores calculated from the differences between students' Start Math and End Math scores, effectively demonstrating how each type of disruptive behaviour influenced academic improvement over time. Quantitatively, the study showcases how each disruptive behaviour feature contributes to the predictive accuracy of the learning outcomes. Using under-sampling to address class imbalances, the analysis moves forward by quantifying the improvement in students' scores from the beginning to the end of the school year. This quantification, denoted as an improvement score, is a pivotal part of the analysis, highlighting the practical impact of the educational interventions on students' performance.

Next, to elucidate the relationship between disruptive behaviours and learning outcomes, the study applied SHAP values to the best-performing model, XGBoost. SHAP values provided a detailed decomposition of each feature's contribution to the predictive outcome, offering insights into the magnitude and direction of each behaviour's impact. The application of SHAP values in explaining the predictions made by ML models offers a profound insight into the significance of each feature in the learning outcomes. XAI allows for a granular understanding of how features such as inattentiveness, hyperactivity, and impulsiveness impact the model's predictions, providing educators and researchers with a powerful tool to interpret complex model behaviours.

For instance, the SHAP values indicated that inattentiveness had the most significant negative impact on learning outcomes, followed by impulsiveness and hyperactivity. This finding is crucial as it not only validates the predictive models but also offers a nuanced view into the specific aspects of disruptive behaviour that most strongly affect student learning. By pinpointing the exact nature of these impacts, educational strategies can be better tailored to address specific disruptive behaviours, potentially leading to more effective interventions. Similarly, impulsiveness contributed negatively, though to a lesser extent than inattentiveness, while hyperactivity had the least impact among the three studied behaviours. Thus, the findings in this thesis, as summarised in this section largely answered the research question, by providing a validated framework for using ML in education, ensuring interpretability and actionable insights for educators.

9.3.3 RQ3: How can Machine Learning (ML) be integrated into an agent-based model (ABM) to improve the simulation of classroom disruptive behaviour, and what parameters of ML prediction yield realistic results in this hybrid ML-ABM approach?

This research investigates how ML can refine and improve the results of ABM simulation of classroom disruptive behaviour. The integration of ML into the ABM framework is designed to enhance the model's predictive capabilities regarding the impact of disruptive behaviours on learning outcomes. The methodology involves using ML to analyse historical data from the PIPS system to predict the End Math scores of students based on their initial assessments and behavioural characteristics. These ML predictions are then fed into the ABM, which simulates daily classroom interactions and adjusts the learning outcomes based on simulated behaviours and interactions.

This incorporation allows the ABM to not only simulate the direct effects of observed behaviours but also to adjust predictions based on the complex, dynamic interactions typical of classroom environments. For instance, ML outputs are used to set initial conditions within the ABM, providing a baseline against which the impact of simulated interventions and interactions is measured. The final End Math scores are a combination of the ML predictions and the outcomes of the ABM simulations, thus offering a nuanced view of how disruptions influence academic performance.

In the hybrid ML-ABM approach described in this thesis, ML plays a crucial role not only in setting initial conditions for the ABM but also in continuously feeding into it to refine and enhance its accuracy and adaptability. This integration showcases a strategic deployment of ML to compensate for the limitations in available features within the PIPS dataset, particularly when either simulation or prediction alone might not suffice due to these limitations. The ML component's utility begins with data pre-processing, where it analyses the available features such as 'Start Math' and 'Start Reading' scores alongside behavioural indicators like 'Inattentiveness', 'Hyperactivity', and 'Impulsiveness'. Through robust data analysis, ML algorithms discover patterns and relationships that might not be immediately apparent but are crucial for setting accurate initial conditions in the simulation. For example, understanding how initial assessments correlate with end academic performance enables the model to simulate more realistic outcomes.

Moreover, ML significantly contributes to predictive modelling within this hybrid framework. By predicting which students are likely to exhibit disruptive behaviours based on historical data and current classroom dynamics, ML sets the stage for more targeted and effective interventions within the ABM. This predictive capability is pivotal in managing classroom dynamics proactively rather than reactively. A unique concept of the employed approach is the establishment of a feedback loop between ABM and ML. As the ABM simulates classroom interactions and academic outcomes, it generates new data that feeds back into the ML model. This data is then used to refine future predictions and adjustments in the simulation. This iterative process enhances the model's accuracy over time, allowing it to adapt to new information and changing classroom dynamics. It essentially creates a dynamic system where both components (ABM and ML) benefit from ongoing interaction, leading to continuous improvement in both the realism of the simulation and the effectiveness of the predictive models. This sophisticated use of ML to support ABM not only addresses the lack of sufficient features in the initial dataset but also leverages the strengths of both approaches to produce results that are as close as possible to real-world scenarios. By integrating ML and ABM in this manner, this thesis research provides a compelling example of how complex educational dynamics can be effectively modelled and understood, leading to more informed and effective educational strategies.

Next, to ensure that the ML component of the hybrid model yields realistic predictions, various parameters were analysed. The primary data included variables such as Start Math and 'Start Reading' scores, as well as behavioural indicators like 'Inattentiveness', 'Hyperactivity', and 'Impulsiveness'. The ML algorithms particularly focused on how these initial assessments and behaviours correlated with the End Math scores. LR was identified as the most effective algorithm for this purpose due to its ability to handle linear relationships between variables effectively. The model's performance was evaluated using MAE and the Pearson correlation coefficient, with results indicating strong predictive accuracy. For instance, the correlation between the predicted End Math scores and the actual scores from the PIPS dataset (for academic year 2007/2008 as well as the following year 2008/2009) was notably high, suggesting that the ML model could reliably predict academic outcomes based on early-year assessments and observed behaviours.

ML’s role in data pre-processing primarily centres on the identification of patterns and correlations that are not immediately obvious. The available data features such as baseline academic scores, behavioural indicators including inattentiveness and hyperactivity and demographic factors provide a rich dataset from which meaningful insights can be derived. For example, understanding how early academic performance might correlate with behaviours such as hyperactivity could help in tailoring interventions that are both timely and effective. The use of various ML algorithms in this exploratory phase allows for the examination of linear and non-linear relationships between features. Techniques such as LR can highlight direct correlations between variables, such as Start Math scores and End Math scores. However, more complex relationships, particularly those involving behavioural data, often require sophisticated ML approaches like decision trees or ensemble methods that can handle the multi-dimensional nature of educational data [60, 150].

The hybrid ML-ABM approach was tested by simulating an academic year with various configurations of student behaviours and classroom interventions. The simulations revealed that changing seating arrangements and implementing targeted behavioural interventions could significantly affect students' academic outcomes. The average End Math scores from the simulations closely matched the actual data, with Pearson correlation coefficients ranging from 0.61 for Start Math to -0.59 for 'Inattentiveness', indicating that both academic and behavioural variables were accurately reflected in the simulation outcomes. Moreover, the model's ability to adapt to different classroom settings and its robustness in handling various behavioural dynamics were affirmed through multiple runs, which consistently produced realistic End Math scores. This consistency underscores the model's utility in exploring the effects of educational strategies and classroom management techniques on student performance.

However, traditional ABMs are often static and do not adapt to changing conditions without manual intervention. With ML integration, ABMs can dynamically adapt to changing behaviours and conditions by updating the agent states based on real-time data analysis. This makes the simulations more responsive to future additions of interventions or changes in the classroom environment. Besides the parameters, by understanding individual student behaviours and their impacts on the classroom environment, educators can tailor interventions more effectively. ML-driven insights can help identify which students need support and what type of interventions are most likely to be effective, enabling more personalised and effective educational strategies. This research in this thesis, as summarised in this section, answers the research question, by demonstrating that ML enhances ABM's adaptability and predictive power, providing a data-driven approach to classroom behaviour modelling.

9.3.4 RQ4: How can gamification strategies be implemented to increase engagement in an educational setting, and which gamification elements have the most significant impact on engagement, both in student-oriented systems and teacher-oriented systems?)

This section assesses the role of gamification in student and teacher engagement, using ML and real-world student data. The detailed analysis of the data in this thesis (Chapter 7) explores the efficacy of gamification elements in increasing engagement within MOOCs, specifically using a Brazilian gamified intelligent learning software known as CamaleOn. The study primarily focuses on real-time detection of which specific gamification elements most significantly impact student engagement. Key findings from the study indicate a strong positive correlation between the gamification rewards earned by students and their engagement levels, measured in terms of login frequency and interaction intensity. This positive association suggests that gamification significantly enhances student engagement, highlighting the effectiveness of elements like badges, which showed the highest correlation with engagement metrics. To elaborate, the thesis presents in section 7.3 a piece of evidence that gamification elements such as XPs, badges, and medals significantly enhance student engagement in MOOC environments. The quantitative data collected from CamaleOn includes interactions from 8,270 students, showcasing a robust dataset well-suited for statistical analysis. For example, during the period studied, students solved 307,814 problems, watched 1,131 videos, received 236,345 badges, and logged in 67,752 times, demonstrating a substantial interaction with the platform's gamified elements.

A crucial aspect of the analysis involves correlating the number of logins and various gamification rewards with student engagement levels. The results from the correlation tests indicated a strong positive relationship between the number of rewards (badges, points, medals) a student earned and their engagement metrics, such as login frequency and interaction with course content. For instance, the Pearson correlation coefficients showed particularly high values between the number of badges earned and login frequency (0.631), and between total reward count and question attempts (0.660).

To determine in real-time which gamification elements most impact student engagement, the study employed both traditional ML models and advanced deep learning (DL) methods to analyse the data. ML techniques like LR, KNN, and more complex algorithms like MLP and CNN were used. These models classified students into high and low engagement categories based on the gamification rewards they interacted with. Interestingly, the MLP model emerged as the most effective, suggesting that DL techniques might be particularly adept at interpreting the complex patterns associated with gamified learning environments. This insight is critical for the real-time analysis as it suggests that deploying models like MLP in live educational platforms could help in dynamically adjusting gamification strategies to enhance student engagement continuously.

The teacher-related part of the RQ4 provides a rich exploration of the efficacy of integrating gamification strategies within an ABM system tailored for teachers. Building upon the foundational work of the CamaleOn study and guided by Toda's TGEEE, this RQ delves into the practical application and iterative refinement of gamification elements to enhance teacher interaction with ABM system. A key finding from Chapter 07 for the student-related part of RQ4 is the significant correlation between specific gamification components including badges, points, and medals and increased student interaction metrics, such as frequency of logins and completion rates. This correlation underscores the practical impact of gamification in fostering student engagement and retention in educational settings. Moreover, the chapter highlights the role of data-driven approaches in facilitating the adaptive integration of gamification elements, ensuring their alignment with teacher needs. The synergy between Toda's theoretical frameworks and the empirical data from Chapter 7 crystallises into a comprehensive understanding of how gamification can be optimised within educational systems for teachers. The chapter illustrates how theoretical constructs cannot only underpin practical implementations but also guide the strategic enhancement of educational technologies, thereby ensuring that gamification remains a dynamic and effective tool in educational settings. Therefore, the research in this thesis, as summarised here, fully answers the research question, by providing empirical validation of gamification elements and offering insights for integrating gamification in diverse educational settings.

9.4 Limitations

Despite its promise, the integration of ABM, ML, and gamification faces numerous challenges. These include ensuring the privacy and ethical use of student data, addressing the technical complexities associated with integrating sophisticated models, and catering to the diverse needs of a global student population. Additionally, the effectiveness of these methodologies can vary significantly depending on several contextual factors, such as classroom size, available resources, and the specific needs of students and educators.

9.4.1 *Limitations on the impact of disruptive behaviours in classrooms and how teacher control, teaching quality, and peer characteristics influence academic achievement (RQ1)*

The ABM approach applied in the doctoral thesis offers invaluable insights into the dynamics of classroom behaviour and its impact on learning outcomes. However, the application of ABM in this context is not without limitations and challenges. First, the model relies on generalised behavioural categories (inattentiveness, hyperactivity, impulsiveness) to represent disruptive behaviours. While these categories capture a broad spectrum of disruptive activities, they do not encompass all possible forms of classroom disruptions. For instance, emotional disturbances or external environmental factors that might influence student behaviour are not accounted for in the simulation. This simplification might lead to an underestimation of the complexity of student interactions and the variety of disruptions that can occur in a real classroom setting.

Moreover, the simulation presumes a consistent effect of disruptive behaviours across different demographic and psychological profiles of students, which may not hold true universally. The diversity in student resilience and susceptibility to disruption based on personal or socio-economic backgrounds is not considered in the current model. This lack of granularity could affect the applicability of the findings to diverse educational settings, potentially limiting the generalisability of the intervention strategies proposed by the study.

Another challenge is the model's static nature regarding the teaching quality and teacher control parameters. In reality, these factors are dynamic and can change over time based on numerous variables, including teacher training, experience, and the specific classroom context. The ABM does not account for the evolution of teaching strategies in response to real-time classroom dynamics, which could lead to discrepancies between the simulated outcomes and actual classroom scenarios. The simulation's outcomes are impacted by this disparity because it is limited to the input data. In the real world, several factors are actually affected such as the performance of students. The spatial component of the model, while innovative, presents another limitation. The impact of disruptive students on their immediate neighbours is considered, but the broader classroom environment and the possible diffusions of disruption are not fully explored. This might overlook the subtler influences that a disruptive student could have on the entire class, beyond their immediate neighbours.

In addition, the ABM's scalability and adaptability to different educational contexts pose a challenge. Transferring the findings from this controlled simulation to real-world educational settings requires a careful consideration of contextual factors that were not simulated. Moreover, the computational demands of scaling the ABM to larger educational settings or more complex student interactions could limit its practical applicability.

The exploration of how teacher control, teaching quality, and peer characteristics influence academic achievement in classrooms with disruptive behaviours presents a complex array of challenges and limitations. Understanding these factors within such a dynamic environment requires meticulous consideration of the interplay between various educational elements and student behaviours. One significant limitation is the inherent complexity of isolating the effects of teacher control, teaching quality, and peer interactions. These factors are deeply intertwined and influenced by a multitude of external and internal classroom dynamics. For instance, the effectiveness of teacher control may not solely depend on the teacher's ability or strategy but also on the nature of the disruptive behaviours and the specific needs and backgrounds of the students involved. Additionally, teaching quality is a multifaceted attribute that includes not only the delivery and content expertise of the teacher but also their ability to engage students and manage the classroom effectively under varying conditions.

Another challenge is the subjective nature of measuring teaching quality and teacher control. These concepts can be highly subjective and vary widely among educators and observers. What constitutes effective teaching or adequate control in one classroom or cultural context may not translate directly to another. This variability makes it difficult to standardise measures of teaching quality and control for the purposes of broad research or application. Peer characteristics also present a complex variable to quantify. The influence of peers can be both positive and negative, and the impact of these interactions on individual students can vary greatly depending on personal characteristics such as resilience, susceptibility to peer pressure, and pre-existing academic abilities. Additionally, the shifting dynamics within a group of students over a school year can alter the influence of peer characteristics on academic achievement, making it a moving target for educational researchers and practitioners.

Moreover, the data collection methods used to assess the impact of these factors are often limited by practical constraints such as time, access, and the willingness of schools to participate in such studies. Observational studies, while rich in detail, can be time-consuming and may not always capture the full range of behaviours and interactions within a classroom. Surveys and self-reports, on the other hand, can introduce biases that skew the data, particularly when respondents are asked to evaluate subjective aspects such as teaching quality or peer influence.

9.4.1 Limitations on the use of disruptiveness features in machine learning models to predict student outcomes, and XAI's interpretation of these predictions (RQ2)

The thesis study provides a comprehensive examination of how disruptive behaviours, such as inattention, hyperactivity, and impulsivity, can be quantified using ML models to forecast students' academic performance, with subsequent interpretations facilitated by XAI. This novel method of prediction and interpretation gives useful information, but it also has certain drawbacks and limits.

One primary limitation in using disruptiveness features to predict educational outcomes is the risk of oversimplification. Disruptive behaviours are complex and multi-faceted, often influenced by a range of environmental, psychological, and physiological factors that may not be fully captured by the three features used. This simplification can lead to models that do not account for the nuanced realities of individual student experiences, potentially resulting in predictions that are not universally applicable or that misrepresent the underlying dynamics. Additionally, the reliance on quantitative measures from structured datasets to predict outcomes introduces potential biases. These biases can stem from the way data is collected, the representativeness of the sample, and the inherent assumptions made during the modelling process. For instance, if the data primarily originates from a specific population or a particular type of educational institution, the predictive model's applicability to other groups or environments may be limited.

On the other hand, the use of XAI to interpret these predictions also presents several challenges. While XAI provides a means to understand the contribution of each feature to the model's predictions, it relies heavily on the accuracy and reliability of the underlying ML model. If the model is biased or based on incomplete data, the explanations generated by XAI will be correspondingly flawed. Moreover, XAI interpretations can sometimes be complex and require a high level of expertise to understand, which may not be accessible to all educational stakeholders. Furthermore, the application of SHAP values and other XAI techniques often assumes that the model's feature-importance rankings are stable across different instances and scenarios. However, in practice, these importance rankings can vary, leading to different interpretations under different circumstances. This variability can confuse stakeholders or lead to inconsistent educational interventions.

9.4.2 Limitations for enhancing machine learning-based agent simulations of disruptive behaviour in the classroom in ABM, as well as the parameters needed to make reliable predictions in this hybrid model (RQ3)

One of the primary limitations in enhancing ML-based simulations within an ABM framework is the dependency on high-quality, comprehensive data. The effectiveness of ML algorithms relies heavily on the availability of large, diverse, and accurate datasets that represent a wide range of student behaviours and outcomes. In many educational settings, collecting such detailed and extensive data can be challenging due to privacy concerns, logistical issues, and the variability in data collection methods across schools. Moreover, the data may not capture all the nuanced factors that influence disruptive behaviour, such as psychological conditions, home environment, and other socio-economic factors, leading to incomplete or biased models. As the complexity of a hybrid ML-ABM approach increases, so does the difficulty in interpreting the results. While ML can provide detailed predictions and ABM can simulate complex interactions, the integration of these two approaches can result in a model that is hard to understand and analyse. This complexity can hinder the ability of educators and policymakers to derive actionable insights from the simulations. Furthermore, complex models require extensive computational resources, which can limit their accessibility for many educational institutions.

Another significant challenge is the generalisability of the hybrid model across different educational contexts. Disruptive behaviour and its impacts can vary greatly depending on cultural, regional, and institutional factors. A model trained on data from one specific educational system or demographic may not perform well when applied to another, limiting its utility across diverse settings. Additionally, scaling the model to accommodate larger or more diverse student populations without losing accuracy or increasing computational costs remains a daunting task. While ML models excel at making predictions based on historical data, their ability to adapt to real-time changes in the classroom is still limited. In dynamic classroom environments, where student interactions and behaviours can change rapidly, the model might not respond swiftly enough to be effectively used as a real-time management tool. This lag can diminish the usefulness of the model in providing immediate support or interventions.

Integrating ML into educational simulations raises ethical questions, particularly regarding student privacy and the potential consequences of predictive modelling. There is a risk of stigmatising students based on predicted behaviours, which could influence teacher perceptions and treatment of students. Ensuring that the use of such models adheres to ethical standards and respects student confidentiality is paramount.

9.4.3 Limitations on the methods in which gamification elements improve MOOC engagement and the real-time metrics measuring these elements' effects on student engagement (RQ4)

The thesis on the impact of gamification on MOOC engagement, particularly on the CamaleOn platform, provides valuable insights, but it also has several limitations and challenges that require further discussion. One significant limitation is the generalizability of the findings. While CamaleOn provides a robust dataset from a

Brazilian context, extending these results to other cultural or educational environments may not be straightforward. Different student populations may interact with gamification elements in diverse ways, influenced by varying educational backgrounds, learning preferences, and cultural factors.

Additionally, the reliance on engagement metrics such as login frequency and interaction data to measure the efficacy of gamification might not fully capture the depth of student engagement or learning outcomes. Engagement is a multi-dimensional construct that includes emotional, cognitive, and behavioural aspects. The current metrics predominantly address the behavioural dimension, potentially overlooking deeper cognitive and emotional engagements that contribute significantly to effective learning.

Another challenge is the evolution of technology and gamification elements themselves. As digital learning environments rapidly evolve, so too do the technologies and methodologies used to implement gamification. The study's findings are based on data-driven approaches that may need continual updates to remain relevant, involving ongoing adaptation of the models to incorporate new gamification trends and technologies that could affect student engagement. The data-driven approach, while innovative, also implies dependence on the availability and quality of the data collected. Any biases in data collection or the specific gamification elements tracked could skew the results. Furthermore, the analysis heavily depends on ML models, which require careful calibration and validation to avoid overfitting or underestimating the complexity of real-world behaviours.

Moreover, the practical implementation of real-time adjustments based on gamification feedback loops is technically challenging. It requires sophisticated systems capable of processing large datasets swiftly and accurately, which can be resource-intensive. Ensuring privacy and ethical handling of student data while implementing such systems is another critical concern that must be addressed to maintain trust and integrity in educational settings.

9.4.4 Limitations for engagement of Toda's gamification elements in the ABM system (RQ8)

Despite the insightful findings based on RQ8, several limitations merit consideration. Firstly, the study relies heavily on quantitative metrics to gauge the effectiveness of gamification elements, which may not fully capture the qualitative aspects of student engagement and learning experiences. This reliance on quantitative data could overlook nuanced behavioural changes and emotional responses that are crucial for a holistic understanding of gamification's impact. Another significant limitation is the sample size and diversity of the teacher participants involved in the study. With only twelve teachers, the findings may not be generalizable across broader educational contexts or diverse student populations. Furthermore, the teachers' perceptions and experiences might reflect a biased view towards innovative educational technologies, thereby skewing the results towards more positive outcomes.

The study also assumes a uniform application of gamification elements across different subjects and educational levels, which may not be effective given the varied nature of subject matter and student demographics. This lack of differentiation could lead to suboptimal engagement strategies in subjects that require different pedagogical approaches. Lastly, the research methodology employed in the study, while robust, does not account for long-term impacts of gamification on educational outcomes. The short-term nature of the study limits the ability to assess the sustainability and evolution of gamification strategies over time, which is critical for validating the enduring benefits of these interventions.

The research model designed for this study, initially based on data from UK schools, faced significant challenges due to the COVID-19 pandemic. This required unforeseen adaptations, including conducting crucial demonstrations and interviews predominantly in Saudi Arabia. This geographical shift introduces a range of limitations affecting the study's applicability and generalizability. Key challenges stem from inherent differences in educational cultures, practices, and technological infrastructures between the UK and Saudi Arabia. These variances may alter the implementation and outcomes of gamification strategies, causing potential data skewness when applied in non-UK contexts. Furthermore, educational objectives, curriculum standards, and engagement metrics in Saudi schools differ significantly from those in the UK, potentially diminishing the effectiveness and impact of the gamification elements originally tailored for the UK. This could lead to research findings that are not universally applicable, thereby constraining the ability to draw broad conclusions about the model's efficacy across diverse educational systems.

Additionally, logistical and communicational hurdles associated with international research during a pandemic add further complexity to conducting and validating the study effectively. These challenges include but are not limited to, time zone differences, language barriers, and the varying degrees of pandemic impact, all of which could affect the quality and consistency of data collection.

9.5 Domain-Specific vs. Generic Applicability of Research Insights

This thesis contributes to understanding student disruptive behaviour, gamification in education, and hybrid ML-ABM systems in a variety of educational contexts. While some findings and methodologies are domain-specific, tailored to the unique characteristics of the datasets and contexts used, others have broader implications and can inform educational strategies in diverse settings.

9.5.1 Domain-specific Insights

Several findings from this research are tightly coupled with the specific contexts of UK primary schools, Brazilian MOOCs, and Saudi Arabian teacher experiences. For instance, the PIPS dataset provided unique behavioural and academic performance metrics that shaped the ABM and ML models, making these tools particularly relevant to traditional classroom settings. Similarly, the use of the CamaleOn dataset highlighted gamification's impact on secondary students in an online context, where cultural and technological factors played a significant role.

These domain-specific insights underline the importance of tailoring interventions to the unique characteristics of each educational environment. For example, cultural attitudes toward classroom management in Saudi Arabia influenced teachers' feedback on gamification elements, demonstrating the need for sensitivity to local norms when deploying such systems.

9.5.2 Generic and Broadly Applicable Findings

Despite the contextual specificity of some results, the research also reveals principles that surpass individual domains. The foundational concept of integrating ML and ABM to simulate and predict educational outcomes can be adapted across various settings. Likewise, the efficacy of gamification in enhancing engagement, supported by clear metrics and predictive modelling, provides a framework applicable to diverse educational platforms.

The use of XAI techniques, such as SHAP values, to interpret the impact of behavioural factors on learning outcomes also offers a universally relevant approach. Educators and policymakers in different contexts can leverage these methodologies to gain deeper insights into student behaviours and optimise interventions.

9.5.3 Balancing Context and Generality

A key challenge in this research has been striking a balance between leveraging domain-specific data and extracting insights that can inform broader applications. The hybrid ML-ABM approach exemplifies this balance, combining the granularity of context-specific simulations with the flexibility of machine learning predictions. While the immediate applications of this work align closely with the studied contexts, the underlying methodologies and conceptual frameworks provide a blueprint for adaptation to other educational systems.

9.5.4 Future Directions

To expand the broader applicability of these findings, future research should focus on testing and validating these models in varied educational environments. This includes exploring how socio-economic, technological, and cultural differences influence the generalisability of interventions. Cross-contextual studies could provide valuable insights into the universal and context-specific elements of disruptive behaviour management and engagement strategies.

9.6 Epilogue

This thesis has ventured deep into the complex interplay of educational practices, student behaviour, and technological advancements. The research has underscored the pivotal role of ABM in simulating and managing classroom dynamics, particularly in addressing disruptive student behaviours. By integrating these technologies with pedagogical strategies and psychological insights, this research has endeavoured to craft a more nuanced understanding of classroom interactions and to develop interventions that are both effective and empathetic. Throughout this research, the significance of high-quality, comprehensive data has been continually emphasised. It is the lifeblood that fuels the efficacy of the employed ML models. As these advances, there is a pressing need to enhance the data collection methodologies to ensure broader coverage and deeper insights. This will not only refine the predictive accuracy of the ML models but also enable them to adapt dynamically to the nuanced needs of diverse educational environments.

Moreover, the generalisability of the findings across different cultural and institutional contexts presents a formidable challenge, necessitating rigorous cross-validation studies to ensure that the models are robust and universally applicable. This endeavour will involve a concerted effort to test and adapt these models across various educational settings, thus paving the way for their widespread adoption and implementation. The ethical dimensions of employing ML in education, particularly concerning student privacy and the risk of stigmatisation, call for a thoughtful examination. Future research must strive to propose and build stringent frameworks and guidelines with ethics that govern the development and utilization of these technologies, ensuring that they enhance rather than compromise the educational experience.

The promise of interdisciplinary research has never been more evident. The fruitful integration of insights from psychology, education, and computer science has unveiled new horizons in understanding and improving student engagement and behaviour. This collaborative approach not only enriches the models but also ensures that they resonate more profoundly with the real-world complexities of educational environments. As we look to the future, the potential of real-time ML applications in dynamically shaping educational interventions offers a thrilling prospect. The development of streaming ML algorithms and the adoption of edge computing can revolutionise the way educational interventions are crafted and delivered, ensuring they are as timely as they are effective.

CHAPTER 10

10 Conclusion

This thesis has rigorously explored the integration of ABM, ML, and gamification within educational settings. Aimed at addressing and mitigating disruptive behaviours while enhancing both student and teacher performance, this research utilises advanced technological methodologies to revolutionise educational practices, offering substantial improvements in engagement and behavioural management.

The initial chapter sets the stage for this investigation by outlining the prevalent issues of disruptive behaviours in educational settings and the potential of technological interventions to address these challenges. Disruptive behaviours, which significantly impede learning processes, necessitate innovative solutions beyond conventional disciplinary actions. The research questions were established to explore how ABM and ML, enhanced with gamification techniques, could be utilised to simulate educational environments and manage classroom dynamics effectively. This foundational chapter emphasised the need for a multidisciplinary approach to understand and improve the interactions within classrooms, setting a clear trajectory for the research.

Chapter 2 delves into the existing literature surrounding ABM, ML, and gamification in education, providing a robust theoretical framework for the study. This comprehensive review highlighted previous studies that demonstrated the efficacy of these technologies in various educational contexts. By examining these precedents, the chapter underscores the gaps in current research, particularly in integrating these three technologies to create a cohesive system that addresses both student and teacher needs. This background established the academic justification for the thesis, linking established theories with the proposed innovative approaches.

The methodology described in Chapter 3 provides details about the technical execution of integrating ABM and ML to create predictive models and simulations of classroom behaviour and learning outcomes. This chapter is crucial in outlining the sophisticated tools and techniques employed, such as the development of specific algorithms for behaviour prediction and the construction of simulation environments for testing different intervention strategies. The methodological rigour ensures that the research was grounded in reliable, scientifically valid techniques, providing a clear pathway from theoretical models to practical applications. These interviews were crucial in refining the models based on first-hand insights, ensuring that the simulations were not only technically robust but also resonated authentically with the practical realities of classroom environments. The qualitative interviews also served as a platform for demonstrating the simulation models to the teachers, a process that was key to validating the models' effectiveness and relevance. Teachers' feedback from these demonstrations was invaluable, providing critical insights that guided further refinement of the simulation models to better meet the needs of schools.

In Chapter 4, the focus shifts to the implementation of the initial simulation models and the preliminary findings. This chapter presents the first iteration of the ABM framework, detailing the setup, and the initial results from the simulations. The findings highlighted how different variables, such as teacher characteristics and student backgrounds, affected the dynamics within the simulated classroom. These initial results were critical in identifying the strengths and limitations of the early models, providing a foundation for subsequent refinements. One significant insight from the model is the impact of peer influences on student behaviour. The simulations revealed that students surrounded by highly disruptive peers tend to exhibit lower academic performance. This finding suggests that interventions should not only focus on individual students but also consider the broader classroom environment. The model also demonstrated that hyperactivity, while a disruptive behaviour, has a less pronounced effect on academic performance compared to inattentiveness. This distinction is important for developing targeted interventions that address the specific types of disruptive behaviours most detrimental to learning.

Chapter 5 has identified the varying effects of three disruptiveness-related features, namely inattentiveness, hyperactivity and impulsiveness on students' learning outcomes through prediction explanations. In this chapter, it was demonstrated how disruptive behaviour can influence both initial and future knowledge, with financial features of students showing a higher impact at the classroom level than at the individual level. Notably, the study revealed that the Students' IDACI has a generally higher effect on learning outcomes, sometimes even surpassing the effect of initial knowledge, contrasting with existing research that typically emphasises initial knowledge as the primary determinant of learning success. Using advanced ML techniques, including hyperparameter tuning and the Jenks' natural breaks method, the performance of predictive models like XGBoost was significantly enhanced. These models, even with a limited number of features, provided transparent predictions through XAI, offering visual explanations for teachers on the predicted learning outcomes using disruptiveness-related features.

Further advancements to ABM simulations were detailed in Chapter 6, where the design and implementation of a hybrid ML-ABM approach were thoroughly examined to enhance the simulation of classroom interactions and disruptive behaviours. The research in this chapter aims to integrate the predictive power of ML with the dynamic capabilities of ABM by creating a robust framework for understanding and improving educational outcomes. This chapter is pivotal in demonstrating the practical applicability and effectiveness of the integrated system, showcasing significant improvements in managing disruptive behaviours and enhancing teaching methodologies. The empirical trials also provided valuable feedback from teachers, which was instrumental in fine-tuning the system to better meet the actual needs of schools. By leveraging advanced computational techniques, this research contributes to the development of more effective and data-driven educational strategies.

Moreover, Chapter 7 presents a grassroots approach to understanding the gamification needs of students and analysing how gamification elements impact student engagement and performance. Specifically, this research investigates how gamification can be linked to student engagement within CamaleOn, a Brazilian MOOC for high school students preparing for higher education. The study revealed that integrating gamification elements such as badges, points, and medals significantly enhances student engagement metrics, including frequency of logins and course completion rates. The results highlighted that the existing gamification features in CamaleOn, such as points, badges, and medals, have a strong positive correlation with student engagement. The data analysis demonstrated that badges, in particular, showed the highest correlation with engagement, indicating that students responded well to the recognition and rewards associated with their activities. The study also showed that these gamification elements could effectively predict student engagement levels, with machine learning models achieving high accuracy rates in their predictions.

In Chapter 8, the integration of gamification elements within the ABM for educational settings is systematically evaluated, drawing on empirical insights and theoretical frameworks. Building on the foundational findings from previous chapters, particularly the efficacy of various gamification elements explored within the CamaleOn system, this chapter aims to validate and enhance the gamification strategies to improve teacher engagement and pedagogical effectiveness. Empirical validation through teacher feedback and engagement metrics indicated that gamification elements categorised under mechanics, like 'Statistics' and 'Progression', had a pronounced impact on engagement. These elements not only aligned with pedagogical goals but also enhanced the educational interface, making learning and teaching more interactive and effective. These interviews were crucial in refining the models based on first-hand insights, ensuring that the simulations were not only technically robust but also resonated authentically with the practical realities of classroom environments. The qualitative interviews also served as a platform for demonstrating the simulation models to the teachers, a process that was key to validating the models' effectiveness and relevance. Teachers' feedback from these demonstrations was invaluable, providing critical insights that guided further refinement of the simulation models to better meet the needs of schools.

In Chapter 9, the discussion expands on the implications of the findings, considering the impact on educational theory and practice. The chapter acknowledges the limitations encountered during the research, such as the variability in educational settings and potential biases in data collection. It also proposes future research directions, emphasising the need for ongoing development and adaptation of the models to encompass a wider range of behavioural dynamics and educational scenarios.

In conclusion, this thesis presents a compelling case for the integration of gamification, ML, and ABM in education. It lays a solid foundation for future research and development in educational technologies, aiming to enhance the learning experience and address challenges within educational settings effectively. The synergy between theoretical frameworks and empirical data observed in this study not only validates the effectiveness of the proposed models but also enriches the academic discourse surrounding educational innovation. The potential for this integrated approach to transform educational practices is immense, promising a future where education is more engaging, adaptive, and effective.

Summarising, the major contributions of this research are as follows:

- For the first time, a comprehensive model was proposed that seamlessly integrates ABM, ML, and gamification techniques within educational settings. This integration is designed to effectively mitigate disruptive behaviours while simultaneously enhancing both student and teacher performance.
- To the best of my knowledge, this is the first study to develop a hybrid ML-ABM approach specifically tailored to simulate classroom interactions and predict disruptive behaviours. This model leverages the predictive power of ML and the dynamic capabilities of ABM to create a robust model for understanding and improving educational outcomes.(Chapter 6)
- A novel method was developed to quantify and analyse the impact of gamification elements on student engagement and performance within a Brazilian MOOC (CamaleOn) for high school students preparing for higher education. The research revealed that integrating gamification elements significantly enhances student engagement metrics, including frequency of logins and course completion rates. (Chapter 7)

- An original model was introduced for systematically evaluating the integration of gamification elements within an ABM for educational settings. I proposed a new method for empirical validation through teacher feedback and engagement metrics, indicating that gamification elements categorised under mechanics, like 'Statistics' and 'Progression', have had a pronounced impact on engagement. (Chapter 8)

The following could be addressed in future research and applications in the field of education. First, future research should explore more sophisticated gamification strategies that align closely with educational objectives and cater to a diverse range of learners. Investigating the long-term effects of gamification on learning outcomes and student retention in MOOCs can provide deeper insights into gamification in enhancing education outcomes [97].

Second, future research should consider more comprehensive behavioural interventions. There is a need for more comprehensive studies on the effectiveness of various behavioural interventions in managing disruptive classroom behaviours. Research should also focus on the scalability and sustainability of these interventions in different educational settings [54].

Third, future research should enhance predictive models with AI. Integrating artificial intelligence and advanced machine learning techniques in predictive modelling offers vast potential. Future research should develop more accurate, reliable, and ethical predictive models, considering the complexities of student behaviours and educational contexts [94]. Furthermore, to enhance the generalisability of ML models, future research should focus on cross-validation studies that apply these models to different educational settings. This involves testing the models across various schools, regions, and cultures to assess their efficacy and adaptability. Diversifying the range of input data can help improve the model's generalisability.

Fourth, further studies are required to understand the complex dynamics of peer influence on academic achievement. Research should focus on how educational policies and classroom practices can harness positive peer interactions while mitigating negative influences [68].

Fifth, future research should consider cross-disciplinary approaches: The research should adopt a cross-disciplinary approach, integrating insights from psychology, education, data science, and behavioural economics. This approach can lead to more holistic and effective educational strategies and interventions.

Sixth, future research could focus on developing more sophisticated data collection methods that ensure broader coverage and deeper insights into student behaviour. This might include longitudinal studies that track student behaviour over several years or the incorporation of qualitative data that can offer richer contextual understanding. Enhancing data collection techniques will aid in training more robust ML models that can accurately predict and adapt to a range of disruptive behaviours.

Therefore, while significant strides have been made in understanding and enhancing educational processes and outcomes, continuous research and innovation are essential. By addressing these recommendations, future research can contribute to more effective, engaging, and inclusive educational practices and policies.

References

[1] Muhammad Adnan, Asad Habib, Jawad Ashraf, Shafaq Mussadiq, Arsalan Ali Raza, Muhammad Abid, Maryam Bashir, and Sana Ullah Khan. 2021. Predicting at-Risk Students at Different Percentages of Course Length for Early Intervention Using Machine Learning Models. *IEEE Access* 9, (2021), 7519–7539. https://doi.org/10.1109/ACCESS.2021.3049446

[2] Lalitha Agnihotri, Ani Aghababyan, Shirin Mojarad, Mark Riedesel, and Alfred Essa. 2015. Mining Login Data For Actionable Student Insight. *Proceedings of the 8th International Conference on Educational Data Mining (EDM)* (2015), 472–475.

[3] Luis Alfaro, Claudia Rivera, and Jorge Luna-Urquizo. 2019. Using Project-based Learning in a Hybrid e-Learning System Model. *International Journal of Advanced Computer Science and Applications* 10, 10 (2019). https://doi.org/10.14569/IJACSA.2019.0101059

[4] Khulood Alharbi, Laila Alrajhi, Alexandra I. Cristea, Ig Ibert Bittencourt, Seiji Isotani, and Annie James. 2020. Data-Driven Analysis of Engagement in Gamified Learning Environments: A Methodology for Real-Time Measurement of MOOCs. . 142–151. https://doi.org/10.1007/978-3-030-49663-0_18

[5] Khulood Alharbi and Alexandra I Cristea. 2024. Hybrid Agent-Based Machine Learning Simulation of a Classroom Disruption Model. In *38th ECMS International Conference on Modelling and Simulation*, 2024. International Conference on Modelling and Simulation. ECMS 2024.

[6] Khulood Alharbi, Alexandra I Cristea, Lei Shi, Peter Tymms, and Chris Brown. 2021. Agent-Based Classroom Environment Simulation: The Effect of Disruptive Schoolchildren’s Behaviour Versus Teacher Control over Neighbours BT - Artificial Intelligence in Education. 2021. Springer International Publishing, Cham, 48–53.

[7] Khulood Alharbi, Alexandra.I Cristea, Lei Shi, Peter Tymms, and Chris Brown. 2021. Agent-based Simulation of the Classroom Environment to Gauge the Effect of Inattentive or Disruptive Students. In *17th International Conference on Intelligent Tutoring Systems*, 2021. .

[8] Raed S. Alsawaier. 2018. The effect of gamification on motivation and engagement. *The International Journal of Information and Learning Technology* 35, 1 (January 2018), 56–79. https://doi.org/10.1108/IJILT-02-2017-0009

[9] A Atif. 2018. Analytics to Determine and Improve Student Engagement and Academic Success in a Unit: Student and Teacher Perspectives. (2018).

[10] Amy H. Auchincloss, Rick L. Riolo, Daniel G. Brown, Jeremy Cook, and Ana V. Diez Roux. 2011. An agent-based model of income inequalities in diet in the context of residential segregation. *American Journal of Preventive Medicine* 40, 3 (2011), 303–311. https://doi.org/10.1016/j.amepre.2010.10.033

[11] Ryan S. Baker. 2016. Stupid Tutoring Systems, Intelligent Humans. *International Journal of Artificial Intelligence in Education* 26, 2 (2016), 600–614. https://doi.org/10.1007/s40593-016-0105-0

[12] Ryan Shaun Baker and Paul Salvador Inventado. 2014. Educational Data Mining and Learning Analytics. In *Learning Analytics*. Springer New York, New York, NY, 61–75. https://doi.org/10.1007/978-1-4614-3305-7_4

[13] Marcelo Balbino, Caroline Jandre, Débora De Miranda, and Cristiane Nobre. 2022. Predictions of Academic Performance of Children and Adolescents with ADHD Using the SHAP Approach. *Studies in Health Technology and Informatics* 290, (2022), 655–659. https://doi.org/10.3233/SHTI220159

[14] Yavuz Selim Balcioğlu and Melike Artar. 2023. Predicting academic performance of students with machine learning. *Information Development* (November 2023). https://doi.org/10.1177/02666669231213023

[15] J. Baldeón, M. Lopez-Sanchez, I. Rodríguez, and A. Puig. 2016. Gamification of multi-agent systems theory classes. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 10003 LNAI, September (2016), 172–183. https://doi.org/10.1007/978-3-319-46840-2_11

[16] J. Baldeón, M. Lopez-Sanchez, I. Rodríguez, and A. Puig. 2016. Gamification design framework to support multi-agent systems theory classes. *Communications in Computer and Information Science* 677, January (2016), 136–155. https://doi.org/10.1007/978-3-319-52039-1_9

[17] Elizabeth F Barkley. 2017. *Student Engagement Techniques A Handbook for College Faculty*.

[18] Russell A Barkley. 2015. *Attention-Deficit Hyperactivity Disorder Fourth Edition A Handbook for Diagnosis and Treatment*. The Guilford Press.

[19] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador Garcia, Sergio Gil-Lopez, Daniel Molina, Richard Benjamins, Raja Chatila, and Francisco Herrera. 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward

responsible AI. *Information Fusion* 58, October 2019 (2020), 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

[20] Carmel Blank and Yossi Shavit. 2016. The Association Between Student Reports of Classmates' Disruptive Behavior and Student Achievement. *AERA Open* 2, 3 (2016), 233285841665392. https://doi.org/10.1177/2332858416653921

[21] Carmel Blank and Yossi Shavit. 2016. The Association Between Student Reports of Classmates' Disruptive Behavior and Student Achievement. *AERA Open* 2, 3 (July 2016), 233285841665392. https://doi.org/10.1177/2332858416653921

[22] Paulo Blikstein and Marcelo Worsley. 2016. Multimodal Learning Analytics and Education Data Mining: using computational technologies to measure complex learning tasks. *Journal of Learning Analytics* 3, 2 (2016), 220–238. https://doi.org/10.18608/jla.2016.32.11

[23] Srricharan Bolisetti, Pranith Reddy Sankepally, Sai Rohith Reddy Kunta, Rohith Reddy Lakkireddy, Manoj Kumar Vemula, Sankepally Pranith Reddy, Kunta Sai Rohith Reddy, and Lakkireddy Rohith Reddy. 2021. Student Performance Analyser using Supervised Learning Algorithms. (2021).

[24] Steven Bramhall, Hayley Horn, Michael Tieu, Steven Bramhall, Hayley Horn, Michael Tieu, and Nibhrat Lohia. 2020. SMU Data Science Review QLIME-A Quadratic Local Interpretable Model-Agnostic Explanation Approach QLIME : A Quadratic Local Interpretable Model-Agnostic Explanation Approach. 3, 1 (2020).

[25] Brandi Simonsen, Sarah Fairbanks, Amy Briesch, Diane Myers, and George Sugai. 2008. Evidence-based Practices in Classroom Management: Considerations for Research to Practice. *Education and Treatment of Children* 31, 1 (2008), 351–380. https://doi.org/10.1353/etc.0.0007

[26] Summer S. Braun, Yvonne H.M. van den Berg, and Antonius H.N. Cillessen. 2020. Effects of a seating chart intervention for target and nontarget students. *Journal of Experimental Child Psychology* 191, (2020), 104742. https://doi.org/10.1016/j.jecp.2019.104742

[27] Leo Breiman. 2001. Random forests. *Machine learning* 45, (2001), 5–32.

[28] Joshua Breslau, Elizabeth Miller, Naomi Breslau, Kipling Bohnert, Victoria Lucia, and Julie Schweitzer. 2009. The Impact of Early Behavior Disturbances on Academic Achievement in High School. *Pediatrics* 123, 6 (June 2009), 1472–1476. https://doi.org/10.1542/peds.2008-1406

[29] Edvin Bru. 2006. Factors associated with disruptive behaviour in the classroom. *Scandinavian Journal of Educational Research* 50, 1 (2006), 23–43. https://doi.org/10.1080/00313830500372000

[30] Ann Buchanan and Charlotte Ritchie. 2004. *What works for troubled children?* Barnado's.

[31] Loraine Busetto, Wolfgang Wick, and Christoph Gumbinger. 2020. How to use and assess qualitative research methods. *Neurological Research and Practice* 2, 1 (December 2020), 14. https://doi.org/10.1186/s42466-020-00059-z
[32] Simoneta Caño de las Heras, Carina L. Gargalo, Charlotte Lærke Weitze, Seyed Soheil Mansouri, Krist V. Gernaey, and Ulrich Krühne. 2021. A framework for the development of Pedagogical Process Simulators (P2Si) using explanatory models and gamification. *Comput Chem Eng* 151, (2021), 107350. https://doi.org/10.1016/j.compchemeng.2021.107350
[33] Ana Paula Cardoso, Manuela Ferreira, José Luís Abrantes, Cláudia Seabra, and Cesário Costa. 2011. Personal and pedagogical interaction factors as determinants of academic achievement. *Procedia - Social and Behavioral Sciences* 29, (2011), 1596–1605. https://doi.org/10.1016/j.sbspro.2011.11.402
[34] Martha Carr, Hillary Hettinger Steiner, Brandon Kyser, and Barry Biddlecomb. 2008. A comparison of predictors of early emerging gender differences in mathematics competency. *Learning and Individual Differences* 18, 1 (2008), 61–75. https://doi.org/10.1016/j.lindif.2007.04.005
[35] Scott E. Carrell, Mark Hoekstra, and Elira Kuka. 2018. The Long-Run Effects of Disruptive Peers. *American Economic Review* 108, 11 (November 2018), 3377–3415. https://doi.org/10.1257/aer.20160763
[36] Jeimmy Carrera and Darinka Ramírez-Hernández. 2018. Innovative education in MOOC for sustainability: Learnings and motivations. *Sustainability (Switzerland)* 10, 9 (2018). https://doi.org/10.3390/su10092990
[37] Jen-Wei Chang and Hung-Yu Wei. 2016. Exploring Engaging Gamification Mechanics in Massive Online Open Courses. *Source: Journal of Educational Technology & Society* 19, 2 (2016), 177–203. https://doi.org/10.2307/jeductechsoci.19.2.177
[38] Joseph A. Cobb. 1972. Relationship of discrete classroom behaviors to fourth-grade academic achievement. *Journal of Educational Psychology* 63, 1 (1972), 74–80. https://doi.org/10.1037/h0032247
[39] Jacob Cohen. 2013. *Statistical power analysis for the behavioral sciences*. Academic press.
[40] Carla Colomer, Carmen Berenguer, Belén Roselló, Inmaculada Baixauli, and Ana Miranda. 2017. The impact of inattention, hyperactivity/impulsivity symptoms, and executive functions on learning behaviors of children with ADHD. *Frontiers in Psychology* 8, APR (2017), 1–10. https://doi.org/10.3389/fpsyg.2017.00540
[41] Carla Colomer, Carmen Berenguer, Belén Roselló, Inmaculada Baixauli, and Ana Miranda. 2017. The Impact of Inattention, Hyperactivity/Impulsivity Symptoms, and Executive Functions on Learning Behaviors of Children with ADHD. *Frontiers in Psychology* 08, (April 2017). https://doi.org/10.3389/fpsyg.2017.00540

[42] Alexandra I. Cristea, Ahmed Alamri, Mohammed Alshehri, Filipe Dwan Pereira, Armando M. Toda, Elaine Harada T. de Oliveira, and Craig Stewart. 2024. The engage taxonomy: SDT-based measurable engagement indicators for MOOCs and their evaluation. *User Modeling and User-Adapted Interaction* 34, 2 (2024), 323–374. https://doi.org/10.1007/s11257-023-09374-x

[43] Jessica L. Darby, Brian S. Fugate, and Jeff B. Murray. 2019. Interpretive research: A complementary approach to seeking knowledge in supply chain management. *International Journal of Logistics Management* 30, 2 (2019), 395–413. https://doi.org/10.1108/IJLM-07-2018-0187

[44] Ali Daud, Miltiadis D. Lytras, Naif Radi Aljohani, Farhat Abbas, Rabeeh Ayaz Abbasi, and Jalal S. Alowibdi. 2017. Predicting student performance using advanced learning analytics. *26th International World Wide Web Conference 2017, WWW 2017 Companion* (2017), 415–421. https://doi.org/10.1145/3041021.3054164

[45] Camelia Delcea, Liviu Adrian Cotfas, Liliana Craciun, and Anca Gabriela Molanescu. 2020. An agent-based modeling approach to collaborative classrooms evacuation process. *Safety Science* 121, August 2019 (2020), 414–429. https://doi.org/10.1016/j.ssci.2019.09.026

[46] Department for Education. 2011. Class Size and education in England evidence report. *Research Report DFE-RR169* (2011), 34–35.

[47] Department for Education and Skills. 2002. Mathematical activities for the Foundation Stage Early years. (2002).

[48] Diego Dermeval, Josmário Albuquerque, Ig Ibert Bittencourt, Seiji Isotani, Alan Pedro Silva, and Julita Vassileva. 2019. GaTO: An Ontological Model to Apply Gamification in Intelligent Tutoring Systems. *Front Artif Intell* 2, July (2019), 1–15. https://doi.org/10.3389/frai.2019.00013

[49] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. 2011. From game design elements to gamefulness: Defining "gamification." *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, MindTrek 2011* September (2011), 9–15. https://doi.org/10.1145/2181037.2181040

[50] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. 2011. From game design elements to gamefulness: Defining "gamification." *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, MindTrek 2011* September (2011), 9–15. https://doi.org/10.1145/2181037.2181040

[51] Ankit Dhamija and Deepika Dhamija. 2020. Impact of innovative and interactive instructional strategies on student classroom participation. In *Handbook of Research on Digital Learning*. IGI Global, 20–37.

[52] Guozhu Dong and Huan Liu. 2018. *Feature engineering for machine learning and data analytics*. CRC press.

[53] George J. DuPaul. 2018. Promoting Success Across School Years for Children With Attention-Deficit/Hyperactivity Disorder: Collaborative School-Home Intervention. *Journal of the American Academy of Child & Adolescent Psychiatry* 57, 4 (April 2018), 231–232. https://doi.org/10.1016/j.jaac.2018.02.001

[54] Arief Eko and Priyo Atmojo. 2020. Junior high school students' disruptive behaviour and their expectations on EFL class. *International Online Journal of Education and Teaching (IOJET)* 7, 2 (2020), 509–521.

[55] Edmund T. Emmer and Laura M. Stough. 2001. Classroom management: A critical part of educational psychology, with implications for teacher education. *Educational Psychologist* 36, 2 (2001), 103–112. https://doi.org/10.1207/S15326985EP3602_5

[56] Joshua M. Epstein. 2008. Why model? *Jasss 11*. https://doi.org/10.1016/s0168-9525(01)02382-4

[57] Eli Vibeke Eriksen, Lene Vestad, Edvin Bru, and Simona C.S. Caravita. 2023. Social competencies, classroom relationships, and academic engagement: A latent change score modeling approach among lower secondary school students. *Social Psychology of Education* 27, 2 (2023), 435–459. https://doi.org/10.1007/s11218-023-09841-8

[58] M. Estrella Esturgó-Deu and Josefina Sala-Roca. 2010. Disruptive behaviour of students in primary education and emotional intelligence. *Teaching and Teacher Education* 26, 4 (2010), 830–837. https://doi.org/10.1016/j.tate.2009.10.020

[59] Antonio Jesus Fernandez-Garcia, Juan Carlos Preciado, Fran Melchor, Roberto Rodriguez-Echeverria, Jose Maria Conejero, and Fernando Sanchez-Figueroa. 2021. A real-life machine learning experience for predicting university dropout at different stages using academic data. *IEEE Access* 9, (2021), 133076–133090. https://doi.org/10.1109/ACCESS.2021.3115851

[60] N. D. Finkelstein, W. K. Adams, C. J. Keller, P. B. Kohl, K. K. Perkins, N. S. Podolefsky, S. Reid, and R. Lemaster. 2005. When learning about the real world is better done virtually: A study of substituting computer simulations for laboratory equipment. *Physical Review Special Topics - Physics Education Research* 1, 1 (2005). https://doi.org/10.1103/PhysRevSTPER.1.010103

[61] Jeremy D Finn, Gina M Pannozzo, and Kristin E Voelkl. 1995. Disruptive and Inattentive-Withdrawn Behavior and Achievement among Fourth Graders. *The Elementary School Journal* 95, 5 (1995), 421–434. https://doi.org/10.1086/461853

[62] Christian Fischer, Zachary A. Pardos, Ryan Shaun Baker, Joseph Jay Williams, Padhraic Smyth, Renzhe Yu, Stefan Slater, Rachel Baker, and Mark Warschauer. 2020. Mining Big Data in Education: Affordances

and Challenges. *Review of Research in Education* 44, 1 (2020), 130–160. https://doi.org/10.3102/0091732X20903304

[63] Panagiotis Fotaris and David Wells. 2017. Game-Based Learning in Schools: Trainee Teacher Perceptions in Implementing Gamified Approaches Item Response Theory for student assessment View project Gamification for learning View project ECGBL 2017 Paper-Mini Track on The Teacher's Role, Identity and. October (2017).

[64] Yoav Freund and Robert E. Schapire. 1996. Experiments with a New Boosting Algorithm. *Proceedings of the 13th International Conference on Machine Learning* (1996), 148–156. https://doi.org/10.1.1.133.1040

[65] Wojciech Galuba and Karl Aberer. 2013. Outtweeting the Twitterers - Predicting Information Cascades in Microblogs. *International Immunology* 25, 11 (2013), 613–614. https://doi.org/10.1093/intimm/dxt055

[66] Nigel Gilbert. 2008. Agent-Based Models. https://doi.org/10.4135/9781412983259

[67] Robyn M Gillies. 2004. The effects of cooperative learning on junior high school students during small group learning. *Learning and Instruction* 14, 2 (April 2004), 197–213. https://doi.org/10.1016/S0959-4752(03)00068-9

[68] Carina Girvan and Timothy Savage. 2019. Virtual Worlds: A New Environment for Constructionist Learning. *Computers in Human Behavior* 99, (2019). https://doi.org/10.1016/j.chb.2019.03.017

[69] Ilie Gligorea, Marius Cioca, Romana Oancea, Andra-Teodora Gorski, Hortensia Gorski, and Paul Tudorache. 2023. Adaptive Learning Using Artificial Intelligence in e-Learning: A Literature Review. *Education Sciences* 13, 12 (December 2023), 1216. https://doi.org/10.3390/educsci13121216

[70] Janice D. Gobert and C. Buckley. 2000. Introduction to model-based teaching and learning in science education. *International Journal of Science Education* 22, 9 (2000), 891–894. https://doi.org/10.1080/095006900416839

[71] Luis Gomes and Zita Vale. 2018. Energy Analyzer Emulation for Energy Management Simulators Multi-Agent System for Microgrid Implementation. 620, October (2018), 215–222. https://doi.org/10.1007/978-3-319-62410-5

[72] Alberto Gómez Mármol, Bernardino Javier Sánchez-Alcaraz Martínez, Alfonso Valero Valenzuela, and Ernesto De La Cruz Sánchez. 2018. Perceived violence, sociomoral attitudes and behaviours in school contexts. *Journal of Human Sport and Exercise* 13, 1 (2018), 138–148. https://doi.org/10.14198/jhse.2018.131.14

[73] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. MIT Press.

[74] Margaret E Gredler. 2004. Games and Simulations and Their Relationships to Learning. In *Handbook of research on educational*

communications and technology, 2nd ed. Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US, 571–581.

[75] Mariola C. Gremmen, Yvonne H.M. van den Berg, Eliane Segers, and Antonius H.N. Cillessen. 2016. Considerations for classroom seating arrangements and the role of teacher characteristics and beliefs. *Social Psychology of Education* 19, 4 (2016), 749–774. https://doi.org/10.1007/s11218-016-9353-y

[76] Léo Grinsztajn, Edouard Oyallon, and Gaël Varoquaux. 2022. Why do tree-based models still outperform deep learning on typical tabular data? *Advances in Neural Information Processing Systems* 35, (2022).

[77] David Gunning and David W. Aha. 2019. DARPA's Explainable Artificial Intelligence Program. *AI Magazine* 40, 2 (June 2019), 44–58. https://doi.org/10.1609/aimag.v40i2.2850

[78] J. Mark Halstead and Jiamei Xiao. 2009. Maintaining the balance: Teacher control and pupil disruptions in the classroom. *Cypriot Journal of Educational Sciences* 4, 3 (2009), 142–156.

[79] Richard Harker and Peter Tymms. 2004. The effects of student composition on school outcomes. *School Effectiveness and School Improvement* 15, 2 (2004), 177–199. https://doi.org/10.1076/sesi.15.2.177.30432

[80] Ramzi Haroun and Christine O'Hanlon. 1997. Teachers' perceptions of discipline problems in a jordanian secondary school. *Pastoral Care in Education* 15, 2 (1997), 29–36. https://doi.org/10.1111/1468-0122.00053

[81] John Hattie. 2008. *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement* (1st ed.). Routledge, London. https://doi.org/10.4324/9780203887332

[82] Carlos J. Hellín, Francisco Calles-Esteban, Adrián Valledor, Josefa Gómez, Salvador Otón-Tortosa, and Abdelhamid Tayebi. 2023. Enhancing Student Motivation and Engagement through a Gamified Learning Environment. *Sustainability (Switzerland)* 15, 19 (2023). https://doi.org/10.3390/su151914119

[83] Christothea Herodotou, Claire Maguire, Martin Hlosta, and Paul Mulholland. 2023. Predictive Learning Analytics and University Teachers: Usage and perceptions three years post implementation. In *LAK23: 13th International Learning Analytics and Knowledge Conference*, March 13, 2023. ACM, New York, NY, USA, 68–78. https://doi.org/10.1145/3576050.3576061

[84] Yujia Hong, Nadira Saab, and Wilfried Admiraal. 2024. Approaches and game elements used to tailor digital gamification for learning: A systematic literature review. *Computers and Education* 212, January (2024), 105000. https://doi.org/10.1016/j.compedu.2024.105000

[85] Stephen Houghton, Kevin Wheldall, and Frank Merrett. 1988. Classroom Behaviour Problems which Secondary School Teachers say they find most Troublesome. *British Educational Research Journal* 14, 3 (1988), 297–312. https://doi.org/10.1080/0141192880140306

[86] Ling Hsiao, Irene Lee, and Eric Klopfer. 2019. Making sense of models: How teachers use agent-based modeling to advance mechanistic reasoning. *British Journal of Educational Technology* 50, 5 (2019), 2203–2216. https://doi.org/10.1111/bjet.12844
[87] Ya-Han Hu, Chia-Lun Lo, and Sheng-Pao Shih. 2014. Developing early warning systems to predict students' online learning performance. *Computers in Human Behavior* (2014). https://doi.org/10.1016/j.chb.2014.04.002
[88] Fred J. Ingram and Roger J. Brooks. 2017. Simulating classroom lessons: An agent-based attempt. *Proceedings of the Operational Research Society Simulation Workshop 2018, SW 2018* 2018, 2014 (2017), 230–240.
[89] Yeonju Jang, Seongyune Choi, Heeseok Jung, and Hyeoncheol Kim. 2022. *Practical early prediction of students' performance using machine learning and eXplainable AI*. https://doi.org/10.1007/s10639-022-11120-6
[90] Marta Jarnac de Freitas and Miguel Mira da Silva. 2023. Systematic literature review about gamification in MOOCs. *Open Learning* 38, 1 (2023), 73–95. https://doi.org/10.1080/02680513.2020.1798221
[91] Farid Jauhari and Ahmad Afif Supianto. 2019. Building student's performance decision tree classifier using boosting algorithm. *Indonesian Journal of Electrical Engineering and Computer Science* 14, 3 (2019), 1298–1304. https://doi.org/10.11591/ijeecs.v14.i3.pp1298-1304
[92] George F Jenks. 1963. Generalization in Statistical Mapping. *Annals of the Association of American* 53, 1 (1963), 15–26.
[93] M I Jordan and T M Mitchell. 2015. Machine learning: Trends, perspectives, and prospects. *Science* 349, 6245 (July 2015), 255–260. https://doi.org/10.1126/science.aaa8415
[94] Dorina Kabakchieva. 2013. Predicting student performance by using data mining methods for classification. *Cybernetics and Information Technologies* 13, 1 (2013), 61–72. https://doi.org/10.2478/cait-2013-0006
[95] Esra Mahsereci Karabulut, Selma Ayşe Özel, and Turgay İbrikçi. 2012. A comparative study on the effect of feature selection on classification accuracy. *Procedia Technology* 1, (2012), 323–327. https://doi.org/10.1016/j.protcy.2012.02.068
[96] Karl M. Kapp. 2012. *The Gamification of Learning and Instruction: Game-based Methods and Strategies for Training and Education*. Pfeiffer.
[97] David Kaufman and Alice Ireland. 2019. Simulation as a Strategy in Teacher Education. *Oxford Research Encyclopedia of Education* June 2019 (2019), 1–19. https://doi.org/10.1093/acrefore/9780190264093.013.478
[98] Hassan Khosravi, Simon Buckingham Shum, Guanliang Chen, Cristina Conati, Yi Shan Tsai, Judy Kay, Simon Knight, Roberto Martinez-

Maldonado, Shazia Sadiq, and Dragan Gašević. 2022. Explainable Artificial Intelligence in education. *Computers and Education: Artificial Intelligence* 3, May (2022). https://doi.org/10.1016/j.caeai.2022.100074
[99] Won Pyo Kim, Hyun Jin Kim, Seung Pil Pack, Jae Hyun Lim, Chul Hyun Cho, and Heon Jeong Lee. 2023. Machine Learning-Based Prediction of Attention-Deficit/Hyperactivity Disorder and Sleep Problems with Wearable Data in Children. *JAMA Network Open* 6, 3 (2023), E233502. https://doi.org/10.1001/jamanetworkopen.2023.3502
[100] Thomas A. Kindermann. 2007. Effects of naturally existing peer groups on changes in academic engagement in a cohort of sixth graders. *Child Development* 78, 4 (2007), 1186–1203. https://doi.org/10.1111/j.1467-8624.2007.01060.x
[101] Thomas A. Kindermann. 2016. Peer Group Influences on Students' Academic Motivation. *Handbook of Social Influences in School Contexts: Social-Emotional, Motivation, and Cognitive Outcomes* January (2016), 31–47.
[102] Roger Kneebone. 2003. Simulation in surgical training: educational issues and practical implications. *Medical education* 37, 3 (March 2003), 267–277. https://doi.org/10.1046/j.1365-2923.2003.01440.x
[103] Sotiris Kotsiantis, Christos Pierrakeas, and P Pintelas. 2004. PREDICTING STUDENTS'PERFORMANCE IN DISTANCE LEARNING USING MACHINE LEARNING TECHNIQUES. *Applied Artificial Intelligence* 18, (2004), 411–426.
[104] Alex Krizhevsky, I Sutskever, and G.E Hinton. 2010. ImageNet Classification with Deep Convolutional Neural Networks. *Encyclopedia of Movement Disorders, Three-Volume Set* (2010), V2-257-V2-259. https://doi.org/10.1016/B978-0-12-374105-9.00493-7
[105] Richard N. Landers, Michael B. Armstrong, and Andrew B. Collmus. 2017. How to Use Game Elements to Enhance Learning: Applications of the Theory of Gamified Learning. In *Serious Games and Edutainment Applications*. Springer International Publishing, Cham, 457–483. https://doi.org/10.1007/978-3-319-51645-5_21
[106] Masoud Gholamali Lavasani and Farah Khandan. 2011. Maintaining the balance: teacher control and pupil disruption in the classroom. *Cypriot Journal of Education* 2, (2011), 61–74.
[107] Victor Lavy, M. Daniele Paserman, and Analia Schlosser. 2012. Inside the Black Box of Ability Peer Effects: Evidence from Variation in the Proportion of Low Achievers in the Classroom. *Economic Journal* 122, 559 (2012), 208–237. https://doi.org/10.1111/j.1468-0297.2011.02463.x
[108] Yann Lecun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *Nature* 521, 7553 (2015), 436–444. https://doi.org/10.1038/nature14539
[109] Geertje Leflot, Pol A. C. van Lier, Patrick Onghena, and Hilde Colpin. 2010. The Role of Teacher Behavior Management in the Development

of Disruptive Behaviors: An Intervention Study with the Good Behavior Game. *Journal of Abnormal Child Psychology* 38, 6 (August 2010), 869–882. https://doi.org/10.1007/s10802-010-9411-4

[110] Jaime Leon, Elena Medina-Garrido, and Juan L. Núñez. 2017. Teaching quality in math class: The development of a scale and the analysis of its relationship with engagement and achievement. *Frontiers in Psychology* 8, JUN (2017), 1–14. https://doi.org/10.3389/fpsyg.2017.00895

[111] Christophe Leys, Christophe Ley, Olivier Klein, Philippe Bernard, and Laurent Licata. 2013. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology* 49, 4 (2013), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

[112] A Liaw and M Wiener. 2002. Classification and Regression by randomForest. *R News* 2, 3 (2002), 18–22.

[113] Brian Liu and Madeleine Udell. 2020. Impact of Accuracy on Model Interpretations. (2020).

[114] Chao Yu Liu, Wei Lieh Huang, Wei Chih Kao, and Susan Shur Fen Gau. 2017. Influence of Disruptive Behavior Disorders on Academic Performance and School Functions of Youths with Attention-Deficit/Hyperactivity Disorder. *Child Psychiatry and Human Development* 48, 6 (2017), 870–880. https://doi.org/10.1007/s10578-017-0710-7

[115] Yiyi Liu, Zijie Huang, and Gong Wang. 2023. Student learning performance prediction based on online behavior: an empirical study during the COVID-19 pandemic. *PeerJ Computer Science* 9, (2023), 1–25. https://doi.org/10.7717/peerj-cs.1699

[116] R. Long. 2019. The School Day and Year (England). *House of Commons Library* 07148 (2019).

[117] Scott M. Lundberg and Su In Lee. 2017. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems* 2017-Decem, December (2017), 4766–4775.

[118] Scott M. Lundberg and Su In Lee. 2017. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems* 2017-Decem, Section 2 (2017), 4766–4775.

[119] John W. Maag. 2018. *Behavior Management: From Theoretical Implications to Practical Applications* (3rd ed.). Cengage Learning.

[120] C. M. MacAl and M. J. North. 2010. Tutorial on agent-based modelling and simulation. *Journal of Simulation* 4, 3 (2010), 151–162. https://doi.org/10.1057/jos.2010.3

[121] Charles M. Macal and Michael J. North. 2008. Agent-based modeling and simulation: ABMS examples. *Proceedings - Winter Simulation Conference* December (2008), 101–112. https://doi.org/10.1109/WSC.2008.4736060

[122] Charles Macal and Michael North. 2014. Introductory tutorial: Agent-based modeling and simulation. In *Proceedings of the Winter*

Simulation Conference 2014, December 2014. IEEE, 6–20. https://doi.org/10.1109/WSC.2014.7019874

[123] Garima Malhotra. 2017. Strategies in Research (Inductive, Deductive, Retroductive and Abductive strategies). *International Journal of Advance Research and Development* 2, 5 (2017), 172–180.

[124] Stanley N. N. Mandah and Comfort Enyi Osuagwu. 2020. Characteristics Behaviours and Factors Responsible for Attention Deficit Hyperactivity Disorder (Adhd) Among Senior Secondary School Students in Rivers State , Nigeria. *European Journal of Special Education Research* 6, 1 (2020), 142–164. https://doi.org/10.46827/ejse.v6i1.3171

[125] Md Maniruzzaman, Jungpil Shin, and Md Al Mehedi Hasan. 2022. Predicting Children with ADHD Using Behavioral Activity: A Machine Learning Analysis. *Applied Sciences (Switzerland)* 12, 5 (2022). https://doi.org/10.3390/app12052737

[126] Glyza M Marie Peras, Janine F Corraine Castro, and JR A Mantog. 2023. Mitigating Students' Disruptive Behavior Through Operant Conditioning. 9, 3 (2023), 2023.

[127] R.J Marzano. 2008. The art and science of teaching: a comprehensive framework for effective instruction. *Choice Reviews Online* 45, 11 (July 2008), 45-6316-45–6316. https://doi.org/10.5860/CHOICE.45-6316

[128] Robert J Marzano. 2007. The Art and Science of Teaching: A Comprehensive Framework for Effective Instruction. 2007. ASCD.

[129] David Masad and Jacqueline Kazil. 2015. Mesa: An Agent-Based Modeling Framework. *Proceedings of the 14th Python in Science Conference* April (2015), 51–58. https://doi.org/10.25080/majora-7b98e3ed-009

[130] Salgado Mauricio, Marchione Elio, and Gilbert Nigel. 2014. Analysing Differential School Effectiveness Through Multilevel and Agent-Based Modelling Multilevel Modelling and School Effectiveness Research. 17, 2014 (2014), 1–13.

[131] Quamrul H. Mazumder, Sunzia Sultana, and Fardeen Mazumder. 2020. Correlation between classroom engagement and academic performance of engineering students. *International Journal of Higher Education* 9, 3 (2020), 240–247. https://doi.org/10.5430/IJHE.V9N3P240

[132] Patrick E McKight and Julius Najab. 2010. Kruskal-wallis test. *The corsini encyclopedia of psychology* (2010), 1.

[133] Jennifer C Mcvay and Michael J Kane. 2011. Does Mind Wandering Reflect Executive Function or Executive Failure? *Psychological Bulletin* 136, 2 (2011), 1–16. https://doi.org/10.1037/a0018298.Does

[134] Christine Merrell, Kapil Sayal, Peter Tymms, and Adetayo Kasim. 2017. A longitudinal study of the association between inattention, hyperactivity and impulsivity and children's academic attainment at

age 11. *Learning and Individual Differences* 53, (January 2017), 156–161. https://doi.org/10.1016/j.lindif.2016.04.003

[135] Christine Merrell and Peter Tymms. 2005. Rasch analysis of inattentive, hyperactive and impulsive behaviour in young children and the link with academic achievement. *Journal of Applied Measurement* 6, 1 (2005), 1–18.

[136] Christine Merrell and Peter B Tymms. 2001. Inattention, hyperactivity and impulsiveness: Their impact on academic achievement and progress. *British Journal of Educational Psychology* 71, 1 (2001), 43–56.

[137] Kenneth W. Merrell and Gretchen Gimpel. 2014. *Social Skills of Children and Adolescents*. Psychology Press. https://doi.org/10.4324/9781315805757

[138] Caitlin Mills, Sidney D'Mello, Nigel Bosch, and Andrew M. Olney. 2015. Mind wandering during learning with an intelligent tutoring system. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 9112, (2015), 267–276. https://doi.org/10.1007/978-3-319-19773-9_27

[139] Haradhan Mohajan. 2020. *Munich Personal RePEc Archive Quantitative Research: A Successful Investigation in Natural and Social Sciences*.

[140] Alana Morais, Joseana Araujo, and Evandro Costa. 2015. Monitoring student performance using data clustering and predictive modelling. *Proceedings - Frontiers in Education Conference, FIE* 2015, (February 2015). https://doi.org/10.1109/FIE.2014.7044401

[141] Hussan Munir, Bahtijar Vogel, and Andreas Jacobsson. 2022. Artificial Intelligence and Machine Learning Approaches in Digital Education: A Systematic Revision. *Information* 13, 4 (2022), 203.

[142] Paul Mupa and Tendeukai Isaac Chinooneka. 2015. Factors Contributing to Ineffective Teaching and Learning in Primary Schools: Why are Schools in Decadence? *Journal of Education and Practice* 6, 19 (2015), 125–133.

[143] Marcell Nagy and Roland Molontay. 2023. Interpretable Dropout Prediction: Towards XAI-Based Personalized Intervention. *International Journal of Artificial Intelligence in Education* 0123456789 (2023). https://doi.org/10.1007/s40593-023-00331-8

[144] Funda Nayir. 2017. The Relationship between Student Motivation and Class Engagement Levels. *Egitim Arastirmalari - Eurasian Journal of Educational Research* 2017, 71 (2017), 59–77. https://doi.org/10.14689/ejer.2017.71.4

[145] Regina M. Oliver, Joseph H. Wehby, and Daniel J. Reschly. 2011. Teacher classroom management practices: effects on disruptive or aggressive student behavior. *Campbell Systematic Reviews* 7, 1 (2011), 1–55. https://doi.org/10.4073/csr.2011.4

[146] E. Olivier and I. Archambault. 2017. Hyperactivity, inattention, and student engagement: The protective role of relationships with teachers and peers. *Learning and Individual Differences* 59, September (2017), 86–95. https://doi.org/10.1016/j.lindif.2017.09.007
[147] Lily Orland-Barak and Ditza Maskit. 2017. *Methodologies of Mediation in Professional Learning*. https://doi.org/10.1007/978-3-319-49906-2
[148] Jayanti Owens and Heide Jackson. 2017. Attention-deficit/hyperactivity disorder severity, diagnosis, & later academic achievement in a national sample. *Social Science Research* 61, (2017), 251–265. https://doi.org/10.1016/j.ssresearch.2016.06.018
[149] Bridgitte Owusu-Boadu, Isaac Kofi Nti, Owusu Nyarko-Boateng, Justice Aning, and Victoria Boafo. 2021. Academic Performance Modelling with Machine Learning Based on Cognitive and Non-Cognitive Features. *Applied Computer Systems* 26, 2 (2021), 122–131. https://doi.org/10.2478/acss-2021-0015
[150] P Pandey and M M Pandey. 2015. *Research Methodology: Tools & Techniques*. Bridge Center.
[151] Yancy Vance Paredes, David Azcona, I-han Hsiao, and Alan F Smeaton. 2018. Predictive Modelling of Student Reviewing Behaviors in an Introductory Programming Course. (2018).
[152] Barry Parsonson. 2012. Evidence-Based Classroom Behaviour Management Strategies. *Kairaranga* 13, 1 (2012), 16–23.
[153] Robyn Paul, Sarah Elaine Eaton, Gayle Laird, Nancy Nelson, and Robert Brennan. 2020. Using Agent-Based Modelling for Eer Experimental Design: Preliminary Validation Based on Student Cheating Behaviours. *Canadian Engineering Education Association (CEEA)* (2020), 1–8.
[154] Fabian Pedregosa, Gael Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12, (2011), 2825–2830.
[155] Alejandro Platas-López, Alejandro Guerra-Hernández, Marcela Quiroz-Castellanos, and Nicandro Cruz-Ramirez. 2023. A survey on agent-based modelling assisted by machine learning. *Expert Systems* (May 2023). https://doi.org/10.1111/exsy.13325
[156] Vagelis Plevris, German Solorzano, Nikolaos P. Bakas, and Mohamed El Amine Ben Seghier. 2022. Investigation of Performance Metrics in Regression Analysis and Machine Learning-Based Prediction Models. *World Congress in Computational Mechanics and ECCOMAS Congress* (2022), 0–25. https://doi.org/10.23967/eccomas.2022.155
[157] Michela Ponticorvo, Raffaele Di Fuccio, Andrea Di Ferdinando, and Orazio Miglino. 2017. An agent-based modelling approach to build up

educational digital games for kindergarten and primary schools. *Expert Systems* 34, 4 (2017), 1–9. https://doi.org/10.1111/exsy.12196

[158] Michela Ponticorvo, Raffaele Di Fuccio, Andrea Di Ferdinando, and Orazio Miglino. 2017. An agent-based modelling approach to build up educational digital games for kindergarten and primary schools. *Expert Systems* 34, 4 (2017), 1–9. https://doi.org/10.1111/exsy.12196

[159] Maryam Poorghorban, Susan Jabbari, and Fateme Chamandar. 2018. Mathematics Performance of the Primary School Students: Attention and Shifting. *Journal of Education and Learning* 7, 3 (2018), 117. https://doi.org/10.5539/jel.v7n3p117

[160] Steven F. Railsback and Volker Grimm. 2011. *Agent-Based and Individual-Based Modeling: A Practical Introduction*. Princeton University Press.

[161] Steven Railsback and Volker Grimm. 2012. *Agent-Based and Individual-Based Modeling: A Practical Introduction*.

[162] Elias Ratinho and Cátia Martins. 2023. The role of gamified learning strategies in student's motivation in high school and higher education: A systematic review. *Heliyon* 9, 8 (August 2023), e19033. https://doi.org/10.1016/j.heliyon.2023.e19033

[163] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. "Why should i trust you?" Explaining the predictions of any classifier. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 13-17-Augu, (2016), 1135–1144. https://doi.org/10.1145/2939672.2939778

[164] Kaili Rimfeld, Margherita Malanchini, Laurie J. Hannigan, Philip S. Dale, Rebecca Allen, Sara A. Hart, and Robert Plomin. 2019. Teacher assessments during compulsory education are as reliable, stable and heritable as standardized test scores. *Journal of Child Psychology and Psychiatry and Allied Disciplines* 60, 12 (2019), 1278–1288. https://doi.org/10.1111/jcpp.13070

[165] Raquel Rosés, Cristina Kadar, and Nick Malleson. 2021. A data-driven agent-based simulation to predict crime patterns in an urban environment. *Computers, Environment and Urban Systems* 89, (September 2021), 101660. https://doi.org/10.1016/j.compenvurbsys.2021.101660

[166] Allison M. Ryan. 2001. The Peer Group as a Context for the Development of Young Adolescent Motivation and Achievement. *Child Development* 72, 4 (August 2001), 1135–1150. https://doi.org/10.1111/1467-8624.00338

[167] Salisu Muhammad Sani, Abdullahi Baffa Bichi, and Shehu Ayuba. 2016. Artificial Intelligence Approaches in Student Modeling: Half Decade Review (2010-2015). *International Journal of Computer Science and Network - IJCSN* 5, 5 (2016), 2277–5420.

[168] Karthik Sankaranarayanan, Ralph Laite, and Nataliya Portman. 2017. Neural network analysis of behavioral agent-based service channel data. *IEEE International Conference on Industrial Engineering and*

Engineering Management 2017-Decem, (2017), 309–313. https://doi.org/10.1109/IEEM.2017.8289902
[169] Nicolae-Bogdan Sara, Rasmus Halland, Christian Igel, and Stephen Alstrup. 2015. High-School Dropout Prediction Using Machine Learning: A Danish Large-scale Study. In *ESANN*, 2015. 23rd.
[170] Katie Seaborn and Deborah I. Fels. 2015. Gamification in theory and action: A survey. *International Journal of Human Computer Studies* 74, December (2015), 14–31. https://doi.org/10.1016/j.ijhcs.2014.09.006
[171] Nabila Sghir, Amina Adadi, and Mohammed Lahmer. 2023. Recent advances in Predictive Learning Analytics: A decade systematic review (2012–2022). *Education and Information Technologies* 28, 7 (July 2023), 8299–8333. https://doi.org/10.1007/s10639-022-11536-0
[172] Juliet Popper Shaffer. 1995. Multipl e hypothesis testing. (1995).
[173] Kyoung Min Shin, Sun Mi Cho, Yun Mi Shin, and Kyung Soon Park. 2016. Effects of early childhood peer relationships on adolescent mental health: A 6- to 8-year follow-up study in South Korea. *Psychiatry Investigation* 13, 4 (2016), 383–388. https://doi.org/10.4306/pi.2016.13.4.383
[174] Avanti Shrikumar, Peyton Greenside, and Anshul Kundaje. 2017. Learning important features through propagating activation differences. *34th International Conference on Machine Learning, ICML 2017* 7, (2017), 4844–4866.
[175] Nikita Sivakumar, Cameron Mura, and Shayn M. Peirce. 2022. Innovations in integrating machine learning and agent-based modeling of biomedical systems. *Frontiers in Systems Biology* 2, November (2022), 1–27. https://doi.org/10.3389/fsysb.2022.959665
[176] Jonathan Smallwood, Merrill McSpadden, and Jonathan W. Schooler. 2007. The lights are on but no one's home: Meta-awareness and the decoupling of attention when the mind wanders. *Psychonomic Bulletin and Review* 14, 3 (2007), 527–533. https://doi.org/10.3758/BF03194102
[177] Jonathan Smallwood and Jonathan W. Schooler. 2006. The restless mind. *Psychological Bulletin* 132, 6 (2006), 946–958. https://doi.org/10.1037/0033-2909.132.6.946
[178] Rodrigo Smiderle, Sandro José Rigo, Leonardo B. Marques, Jorge Arthur Peçanha de Miranda Coelho, and Patricia A. Jaques. 2020. The impact of gamification on students' learning, engagement and behavior based on their personality traits. *Smart Learning Environments* 7, 1 (December 2020), 3. https://doi.org/10.1186/s40561-019-0098-x
[179] David P. Smith, Angela Hoare, and Melissa M. Lacey. 2018. Who goes where? The importance of peer groups on attainment and the student use of the lecture theatre teaching space. *FEBS Open Bio* 8, 9 (2018), 1368–1378. https://doi.org/10.1002/2211-5463.12494

[180] Marina Sokolova and Guy Lapalme. 2009. A systematic analysis of performance measures for classification tasks. *Information Processing and Management* 45, 4 (2009), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

[181] Emily Solari. 2014. Longitudinal prediction of 1st and 2nd grade English oral reading fluency in ELL. *Journal of adolescence* 74, 4 (2014), 274–283. https://doi.org/10.1002/pits

[182] Marta Soriano-Pascual, Alberto Ferriz-Valero, Salvador García-Martínez, and Salvador Baena-Morales. 2022. Gamification as a Pedagogical Model to Increase Motivation and Decrease Disruptive Behaviour in Physical Education. *Children* 9, 12 (2022), 1–12. https://doi.org/10.3390/children9121931

[183] Kurt D Squire and Mingfong Jan. 2007. Mad City Mystery: Developing Scientific Argumentation Skills with a Place-based Augmented Reality Game on Handheld Computers. *Journal of Science Education and Technology* 16, 1 (2007), 5–29. https://doi.org/10.1007/s10956-006-9037-z

[184] Latha Subramainan, Moamin A. Mahmoud, Mohd Sharifuddin Ahmad, and Mohd Zaliman Mohd Yusoff. 2016. An emotion-based model for improving students' engagement using agent-based social simulator. *International Journal on Advanced Science, Engineering and Information Technology* 6, 6 (2016), 952–958. https://doi.org/10.18517/ijaseit.6.6.1300

[185] Latha Subramainan, Moamin A Mahmoud, Mohd Sharifuddin Ahmad, and Mohd Zaliman Mohd Yusoff. 2018. A Simulator's Specifications for Studying Students' Engagement in a Classroom. . 206–214. https://doi.org/10.1007/978-3-319-62410-5_25

[186] George Sugai and Robert R. Horner. 2006. A Promising Approach for Expanding and Sustaining School-Wide Positive Behavior Support. *School Psychology Review* 35, 2 (June 2006), 245–259. https://doi.org/10.1080/02796015.2006.12087989

[187] Anna M. Sullivan, Bruce Johnson, Larry Owens, and Robert Conway. 2014. Punish them or engage them? Teachers' views of unproductive student behaviours in the classroom. *Australian Journal of Teacher Education* 39, 6 (2014), 43–56. https://doi.org/10.14221/ajte.2014v39n6.6

[188] Susan R. Swing and Penelope L. Peterson. 1982. The Relationship of Student Ability and Small-Group Interaction to Student Achievement. *American Educational Research Journal* 19, 2 (1982), 259–274. https://doi.org/10.3102/00028312019002259

[189] Lucile Ter-Minassian, Natalia Viani, Alice Wickersham, Lauren Cross, Robert Stewart, Sumithra Velupillai, and Johnny Downs. 2022. Assessing machine learning for fair prediction of ADHD in school pupils using a retrospective cohort study of linked education and healthcare data. *BMJ Open* 12, 12 (2022), 1–9. https://doi.org/10.1136/bmjopen-2021-058058

[190] Robert L. Thorndike. 1953. Who belongs in the family? *Psychometrika* 18, 4 (1953), 267–276. https://doi.org/10.1007/BF02289263

[191] Armando M. Toda, Wilk Oliveira, Ana C. Klock, Paula T. Palomino, Marcelo Pimenta, Isabela Gasparini, Lei Shi, Ig Bittencourt, Seiji Isotani, and Alexandra I. Cristea. 2019. A taxonomy of game elements for gamification in educational contexts: Proposal and evaluation. *Proceedings - IEEE 19th International Conference on Advanced Learning Technologies, ICALT 2019* 2161–377X, (2019), 84–88. https://doi.org/10.1109/ICALT.2019.00028

[192] Nikola Tomasevic, Nikola Gvozdenovic, and Sanja Vranes. 2020. An overview and comparison of supervised data mining techniques for student exam performance prediction. *Computers and Education* 143, February 2019 (2020), 103676. https://doi.org/10.1016/j.compedu.2019.103676

[193] Ivan Tomek. 1976. Tomek Link: Two Modifications of CNN. *IEEE Trans. Systems, Man and Cybernetics* (1976), 769–772.

[194] Maria Tsiakmaki, Georgios Kostopoulos, Sotiris Kotsiantis, and Omiros Ragos. 2020. Transfer learning from deep neural networks for predicting student performance. *Applied Sciences (Switzerland)* 10, 6 (2020). https://doi.org/10.3390/app10062145

[195] Maren Stabel Tvedt and Edvin Bru. 2023. Completion of upper secondary education: Predictions of the psychosocial learning environment and academic achievement. *Learning and Instruction* 88, August (2023), 101824. https://doi.org/10.1016/j.learninstruc.2023.101824

[196] P Tymms. 1999. *Baseline Assessment and Monitoring in Primary Schools* (1st Editio ed.). David Fulton Publisher, London.

[197] P Tymms and S Albone. 2002. Performance indicators in primary schools. *School improvement through performance feedback* (2002), 191–218.

[198] Peter Tymms and Chris Brown. Modelling Educational Systems: An interactive approach to integrating theory and policy.

[199] Anthony Vance, Paul Benjamin Lowry, and Dennis Eggett. 2015. Increasing Accountability Through User-Interface Design Artifacts: A New Approach to Addressing the Problem of Access-Policy Violations. *MIS Quarterly* 39, 2 (February 2015), 345–366. https://doi.org/10.25300/MISQ/2015/39.2.04

[200] Boglarka Vekety, H. N.Alexander Logemann, and Zsofia K. Takacs. 2021. The effect of mindfulness-based interventions on inattentive and hyperactive–impulsive behavior in childhood: A meta-analysis. *International Journal of Behavioral Development* 45, 2 (2021), 133–145. https://doi.org/10.1177/0165025420958192

[201] Stella Vosniadou. 1994. Capturing and modeling the process of conceptual change. *Learning and Instruction* 4, 1 (1994), 45–69. https://doi.org/10.1016/0959-4752(94)90018-3

[202] Christopher W. Weimer, J. O. Miller, and Raymond R. Hill. 2016. Agent-based modeling: An introduction and primer. In *2016 Winter Simulation Conference (WSC)*, December 2016. IEEE, 65–79. https://doi.org/10.1109/WSC.2016.7822080

[203] Philip H Winne and Allyson Fiona Hadwin. 1998. Studying as Self-Regulated Learning. *Metacognition in Educational Theory and Practice* January 1998 (1998), 291–318. https://doi.org/10.4324/9781410602350-19

[204] World Health Organization. 1993. The ICD-10 Classification of Mental and Behavioural Disorders. *World Health Organization* 55, 1993 (1993), 135–139. https://doi.org/10.4103/0019

[205] Janel Wu. 2020. Machine Learning in Education. *Proceedings - 2020 International Conference on Modern Education and Information Management, ICMEIM 2020* (2020), 56–63. https://doi.org/10.1109/ICMEIM51375.2020.00020

[206] Jie Xu, Kyeong Ho Moon, and Mihaela Van Der Schaar. 2017. A Machine Learning Approach for Tracking and Predicting Student Performance in Degree Programs. *IEEE Journal on Selected Topics in Signal Processing* 11, 5 (2017), 742–753. https://doi.org/10.1109/JSTSP.2017.2692560

[207] Gokul Yenduri, Rajesh Kaluri, Dharmendra Singh Rajput, Kuruva Lakshmanna, Thippa Reddy Gadekallu, Mufti Mahmud, and David J. Brown. 2023. From Assistive Technologies to Metaverse - Technologies in Inclusive Higher Education for Students with Specific Learning Difficulties: A Review. *IEEE Access* 11, (2023), 64907–64927. https://doi.org/10.1109/ACCESS.2023.3289496

[208] Wei Zhang, Andrea Valencia, and Ni Bin Chang. 2023. Synergistic Integration Between Machine Learning and Agent-Based Modeling: A Multidisciplinary Review. *IEEE Transactions on Neural Networks and Learning Systems* 34, 5 (2023), 2170–2190. https://doi.org/10.1109/TNNLS.2021.3106777

[209] Liqiu Zhao and Zhong Zhao. 2021. Disruptive Peers in the Classroom and Students' Academic Outcomes: Evidence and Mechanisms. *Labour Economics* 68, June 2020 (2021). https://doi.org/10.1016/j.labeco.2020.101954