SIMULATING CLASSROOM LESSONS: AN AGENT-BASED ATTEMPT

Fred J. Ingram
Department of Management Science
Lancaster University Management School
Lancaster, LA1 4YX, U.K.
fred.ingram@lancaster.ac.uk

Dr Roger J. Brooks
Department of Management Science
Lancaster University Management School
Lancaster, LA1 4YX, U.K.
roger.brooks@lancaster.ac.uk

ABSTRACT
This is an interim report on a project to construct an agent-based simulation that reproduces some of the interactions between students and their teacher in classroom lessons. In a pilot study, the activities of 67 students and 7 teachers during 40 lessons were recorded using a data collection instrument that currently captures 17 student states and 15 teacher states. These data enabled various conceptual models to be explored, providing empirical values and distributions for the model parameters. Using these data, a lesson can be ‘played back’ using a visualization program implemented in NetLogo. A visualization and simulation can be viewed side-by-side and their outputs compared in various ways, e.g. overall class state-transition matrices or individual student state trajectories. The main challenges are the formulation of descriptive rules, establishing what metrics to use to compare lessons, and determining how to validate a simulation.

Keywords: Agent-based Modelling, Agent-based Simulation, Social Simulation, NetLogo

1 INTRODUCTION AND OVERVIEW
The goal of this research is to enhance understanding of the dynamic interactions between students and their teacher in classroom lessons, using an agent-based model (ABM). One aspect of this is to use the model to try to predict the consequences of some standard classroom interventions, for instance:
- What would be the effect of adding a teaching assistant (TA) to a class?
- What happens when a friendship group becomes disengaged and disturbs a lesson?
- What effect would a different table/seating arrangement have?
- Which of two lesson plans would be more effective?

It is often quite difficult, and possibly unethical, to experiment with alternative options for classes. A simulation could be a pragmatic decision-support tool for use by school or college teachers or senior management. Of most interest to them is likely to be the end results after a lesson or series of lessons, such as the degree of productivity (‘time on task’) or the frequency of disruption, or the amount of student participation. There appears to be no existing model that can accomplish these goals, so the challenge is to create one.

Classroom simulations have been developed to help train teachers. Trainees are presented with repeatable scenarios populated with some virtual students and can explore alternative interventions. But these systems rely on the trainee choosing a course of action. They are not full simulations of an entire lesson and there is no simulation of the teacher. There are also training systems in which the students are ‘avatars’, operated by teaching professionals (Ferry, Kervin and Carrington, 2011; Deale and Pastore, 2014; Gregory, 2014).

Agent-based modelling is a computational approach to simulating systems. Agents (actors) are identified along with their states and essential attributes. The principal interactions between agents and their environment are formulated as rules, which can be stochastic in nature. Agent-based modelling and simulation have been used in many different application areas, including in the field of education.

Salgado et al (2014) developed an ABM to investigate ‘differential school effectiveness’. They proposed incorporating factors such as friendships between students and the expectation bias of the teachers, and found that their ABM provided improved causal explanations of differences between schools. Manzo (2013) described the development of an ABM incorporating a network of social influences and noted that with an ABM it is possible to identify causes. Mital et al (2014) used an
ABM and social network analysis to simulate the outcomes of a school project. Student agents had attributes such as aspirations, cognitive ability and perseverance and the relationships between agents could change dynamically. Abrahamson et al (Abrahamson, Blikstein and Wilensky, 2007) developed an ABM to investigate the influence of individual and social factors on collaborative problem-solving activities. They believed their study might be useful in teacher training and educational policy making.

Classroom models have also been constructed. Gamboa-Brooks-Gray (2015) described a classroom ABM in which random class layouts with students were created, but without a teacher. She sought to establish a correlation between the student agents’ ‘classroom environment variables’ (primarily how they feel about their tuition) and their attainment.

It seems there is no published research taking the approach used in this project, namely modelling the spatial arrangement and temporal ‘trajectories’ of students and teacher through various activities or states during lessons, with constant interactions between students and the teacher (and teaching assistant, if present).

2 CONCEPTS FOR THE MODELS

An agent-based model postulates that the agents have specific attributes that contribute to their behaviour and that they engage in specific activities, represented as the states of the agents. A model also hypothesises relationships between agents and their environment and suggests rules of interaction. There are many influences that affect the events in classrooms (see Figure 1 below), some of which are discussed in this section.

![Figure 1 Some of the many influences on students in a classroom lesson](image)

Many factors are known to have an impact on students’ experiences of education, their achievements and their behaviour, for example: socio-economic and family background, parent’s education, ethnicity, gender, special educational needs, individual learning styles, the community and culture in which the school exists, the target academic level set by teachers and/or the school, the range of abilities (heterogeneity) within a class, class sizes, teaching styles (Hirschy and Wilson, 2002; Wigfield and Cambria, 2010; Sun and Shek, 2012; Swinson, 2012; Dillenbourg, 2013; Easby, 2015; Kalambouka et al., 2016; Pampaka and Williams, 2016) Of particular importance are friendship networks and the general influence of peers and friends, especially the academic achievement of peers (Hirschy and Wilson, 2002; Hanushek et al., 2003; Halliday and Kwak, 2012; Blansky et al., 2013). There is also an effect on time spent learning due to different seating arrangements (Schwieso, 1995; Bicard et al., 2012).

Student personality and individuality are also important. These influences could be broken down into factors representing Openness to experience, Conscientiousness, Extraversion, Agreeableness and Neuroticism (the OCEAN model) or Belief, Desire and Intention (the BDI model) or to a number of other alternative concepts (Gibson, 2009, 2011; Stavroula et al., 2016; Jager, 2017).
Some examples of student attributes are:
▪ inclination to follow the teacher’s instructions - related to attitude to teacher and subject
▪ distractibility, e.g. by noises, conversations
▪ tendency to ask for help or to ask questions.

But these student-specific attributes may not be the best predictors of the next state a student agent moves into. What is happening at each moment in a lesson matters. For example, from a student’s view:
▪ am I finding the current activity boring or too easy or too difficult?
▪ am I so tired I can’t concentrate?
▪ are my friends participating – whatever they are doing I might be inclined to copy?

As part of the agent-based modelling process, the most relevant factors are selected to become the agents’ attributes. In addition, out of all the possible actions of the agents, some are selected for modelling. These become the agent states. In this project, the following student agent states were selected (the colour-coding showing which states are considered productive, disruptive, or other (e.g. resting)). There is a similar list for teachers.

<table>
<thead>
<tr>
<th>State</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Being disciplined by teacher or TA</td>
</tr>
<tr>
<td>2</td>
<td>Unproductive, away from own desk and distracting others</td>
</tr>
<tr>
<td>3</td>
<td>In own seat chatting, distracting, socialising, turning around etc.</td>
</tr>
<tr>
<td>4</td>
<td>Intentionally unproductive, not participating, but not distracting others</td>
</tr>
<tr>
<td>5</td>
<td>Class being told off</td>
</tr>
<tr>
<td>6</td>
<td>Unintentionally unproductive: not learning, but not disturbing others</td>
</tr>
<tr>
<td>7</td>
<td>Not sure if productive: just sitting, not disturbing others</td>
</tr>
<tr>
<td>8</td>
<td>Working alone</td>
</tr>
<tr>
<td>9</td>
<td>Working with others</td>
</tr>
<tr>
<td>10</td>
<td>Expressing knowledge to the class</td>
</tr>
<tr>
<td>11</td>
<td>Being appreciated or praised by the teacher</td>
</tr>
<tr>
<td>12</td>
<td>Listening to / interacting with teaching</td>
</tr>
<tr>
<td>13</td>
<td>Receiving individual instruction or support from the teacher</td>
</tr>
<tr>
<td>14</td>
<td>In a group receiving instruction from the teacher</td>
</tr>
<tr>
<td>15</td>
<td>Receiving instruction from the TA in the classroom</td>
</tr>
<tr>
<td>16</td>
<td>Gone out of classroom with the TA support group</td>
</tr>
<tr>
<td>17</td>
<td>None of these (e.g. left room)</td>
</tr>
</tbody>
</table>

Table 1: Brief descriptions of student states
A classroom is definitely a complex system (Abrahamson, Blikstein and Wilensky, 2007; Blikstein, Abrahamson and Wilensky, 2008; Keshavarz et al., 2010; Burns and Knox, 2011) and student behaviour will be difficult to predict. The next state of the teacher, however, is less capricious. In general, the teacher is either following the lesson plan or temporarily side-tracked by having to manage student behaviour or provide unanticipated additional support. Given a lesson plan, it is anticipated that it will be possible to produce a realistic simulation of a teacher, with variable attributes (such as the tendency to offer individual help or the level of strictness). Hence an outline lesson plan is an essential input to a simulation.

3 DATA COLLECTION
A pilot study was conducted at a small UK secondary school and involved 67 students and 7 teachers. Besides using questionnaires to obtain data for some of the agent attributes, a significant amount of effort was invested in collecting data about agent states (activities), with 40 lessons observed and over 20,000 events (state changes) recorded. The intention is to derive some generic student profiles based on an analysis of the combination of questionnaire data and event data, using Principal Components Analysis to reduce the number of factors and Cluster Analysis to identify student groupings.

These data provided useful estimates for agent states, data that do not seem to be available in published research. For example:
▪ The length of time students work before chatting or taking a break.
▪ The length of time students receive individual help from a teacher or teaching assistant.

Although much classroom data collection is still conducted using paper forms, there is increasing use of software tools to ‘mark-up’ event data directly (see for example Vosaic Connect
Ingram and Brooks (https://vosaic.com accessed 6 January 2018)). The lesson events data collection instrument developed for this project is unique and efficient, enabling an observer to record multiple events within seconds of each other. The Microsoft® Excel (2016) based program (screenshot in Figure 2 below) captures the student and teacher states; it also records teaching assistant participation. It creates a log of events, a series of time-stamped agent states. With some additional processing these log files provide student and lesson statistics, as well as being the input to the visualization program, described next.

The lesson events data collection instrument developed for this project is unique and efficient, enabling an observer to record multiple events within seconds of each other. The Microsoft® Excel (2016) based program (screenshot in Figure 2 below) captures the student and teacher states; it also records teaching assistant participation. It creates a log of events, a series of time-stamped agent states. With some additional processing these log files provide student and lesson statistics, as well as being the input to the visualization program, described next.

![Image of Figure 2](image-url)

**Figure 2** The event data collection user interface showing students in groups, in states and participating.

As with all classroom data collection, the observer has to make subjective decisions about how to interpret and then classify an observed event. It is critical that the observer is completely familiar with the identified states of the teacher and students.

Some results from the analysis of these data are described in section 5. A second phase of data collection is planned. This will seek to obtain more accurate values for those agent attributes and states that have been identified as a result of deeper conceptual modelling.

### 4 DATA VISUALISATION AND SIMULATION

The screenshot in Figure 3 below shows a version of the NetLogo-based visualization and simulation program (VizSim) during a run of a lesson visualization. (NetLogo is a popular, open-source agent-based simulation package available at https://ccl.northwestern.edu/netlogo/ accessed on 27th September 2017). The log file produced by the event data collection program during one lesson was the input to the program.

![Image of Figure 3](image-url)

**Figure 3** The NetLogo visualization and simulation program during a run of a lesson visualization.

As with all classroom data collection, the observer has to make subjective decisions about how to interpret and then classify an observed event. It is critical that the observer is completely familiar with the identified states of the teacher and students.

Some results from the analysis of these data are described in section 5. A second phase of data collection is planned. This will seek to obtain more accurate values for those agent attributes and states that have been identified as a result of deeper conceptual modelling.

Although the text is unreadable until one zooms in, one can see that most of the screen contains various attempts at displaying what is happening to individuals and to the class as a whole. The middle section includes a representation of a classroom with students at their desks, the teacher (wolf) is currently helping one student at his/her desk, as is the teaching assistant (cat).

What can’t be seen are the dynamic links that form and disappear during student-student, student-teacher, and student-TA interactions. The colours have significations, for example green represents a ‘productive’ state, blue a ‘resting/other’ state and red a ‘disruptive/distracting’ state. The background of each student becomes pinker as their level of disruptiveness persists.

The right-hand side covers setting up a classroom and/or lesson and includes controls for the visualization and simulation, primarily for choosing a model and setting its parameters. This side also includes the controls for making changes to the classroom or agents as part of a simulation, e.g. moving a student, or changing teacher attributes or the lesson plan.

During the visualisation, the software accumulates detailed and summary data which can be used to compare actual lessons with simulations, or to compare agents, lessons or classes.
5 PRELIMINARY DATA ANALYSIS

To date, data from 51 students and 21 lessons have provided values for a range of agent attributes and model parameters, quantifying many lesson activities. It is hoped that analyses of the questionnaire data in conjunction with the event data will enable construction of the all-important student profiles and teacher profiles – thus making it possible to allocate a profile to an agent for a simulation. A few results concerning student states are shown below. The states were described in Table 1 in section 2.

The histogram in Figure 4 below shows the relative frequencies of the time students spent working alone (state 8). The long ‘tail’ is caused by a few students who can work for extended periods alone, and were allowed to. These data are not showing how long students can study alone before needing a break: the data show how long they did study alone before either taking a break or being interrupted (by the teacher, for example). From these data, one could model an imaginary ‘average’ student as spending a mean of 120 seconds working alone, and, in simulations, use a suitable probability distribution to generate random durations.

Figure 3 The Simulation-Visualization program (VizSim)

Figure 4 Length of time students worked alone (state 8)
The student states are classified as either productive, disruptive or ‘other’ (see Table 1 in section 2). Table 2 below shows the proportions of these states for three the students, together with the overall averages. The variations illustrate that a model should not use average values to represent an imaginary ‘average’ student, but rather that different profiles should be used to model different types of students.

Table 2 How much of a student’s lesson time is productive, disruptive or ‘other’?

<table>
<thead>
<tr>
<th>Student REF</th>
<th>Productive</th>
<th>Disruptive</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>709</td>
<td>99%</td>
<td>0%</td>
<td>1%</td>
</tr>
<tr>
<td>812</td>
<td>84%</td>
<td>11%</td>
<td>4%</td>
</tr>
<tr>
<td>1001</td>
<td>86%</td>
<td>1%</td>
<td>13%</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>90%</strong></td>
<td><strong>4%</strong></td>
<td><strong>5%</strong></td>
</tr>
</tbody>
</table>

One of the results from the event data is a set of state transition matrices (STM) which provide the relative probability that a student will change from one state to another. The overall STM produced for all students in all lessons could represent the behaviour of an imaginary ‘average’ student, although an STM for each profile type would be preferred. These STMs can also provide parameter values for more specific rules of agent interaction. Together, a student STM and a teacher STM enable a purely empirically-driven stochastic simulation.

With the values provided by these data it is possible to provide models with empirically-based parameter values.

6 SOME INITIAL MODELS

Since constructing a simple production-rule model (3 student states, no teacher) to assess project feasibility, and then the subsequent data collection, a few other conceptual models have been developed:

1. **Random state transitions** – everyone changes states randomly; while obviously unrealistic, this does provide a performance baseline.
2. **Empirical State Transition Matrix** – the next state is the one most likely, based on historical data; this model also is inherently unrealistic, mainly because agents are not interacting. However, it also provides a sort of baseline: if the proposed model performs similarly to the empirical data, then the model is promising.
3. **The ‘Max{P,D,R}’ model** – discussed below.
4. **Plausible Rules** – discussed below.

6.1 The Max{P,D,R} model

In this simple model, student agents can be in one of three states, **Productive**, **Disruptive**, or **Resting**, and the teacher has one state, **teaching**. At each tick of the simulation clock, each student has to decide whether to stay in the current state or switch to one of the other two states. The idea is to sum all the factors that tend to increase or decrease productivity, disruption and resting, and so calculate scores that reflect the student’s inclination to be in those states, and then choose the state with the maximum score (or at random if equal maxima). The three scores are:

<table>
<thead>
<tr>
<th>Concept</th>
<th>Construction</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Productive</strong> Score</td>
<td>Add response to teacher, add influence of others being productive, add factor for proximity of teacher.</td>
<td>P</td>
</tr>
<tr>
<td><strong>Disruptive</strong> Score</td>
<td>Add opposite response to teacher, add influence of others being disruptive, subtract factor for proximity of teacher, add factor for need to socialise.</td>
<td>D</td>
</tr>
<tr>
<td><strong>Resting</strong> Score</td>
<td>Add opposite of energy level, add influence of others resting, subtract factor for proximity of teacher.</td>
<td>R</td>
</tr>
</tbody>
</table>
This model assumes that student behaviour is influenced by the following student attributes as well as by two well-known dynamic factors: the proximity of the teacher, and what their friends are currently doing.

**Relevant student attributes**

- Inclination to follow the teacher’s instructions and be productive; $I - t$
- Need for social interaction (assumed to mean disruptive activity, chatting etc.); $s$
- Inclination to copy other students (response to ‘peer pressure’); $c$
- Inclination to rest (basically a measure of energy level); $e$
- Sensitivity to influence of teacher; $\alpha$
- Sensitivity to influence of peers; $\beta$
- Sensitivity to proximity of teacher; $\gamma$
- Influence of energy-level; $r$
- Influence of need to socialize; $d$

<table>
<thead>
<tr>
<th>Attitude to the other student</th>
<th>Friend</th>
<th>Indifferent (or not mentioned)</th>
<th>Enemy*</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of students being productive</td>
<td>$F_b$</td>
<td>$I_p$</td>
<td>$E_p$</td>
</tr>
<tr>
<td>No. of students being disruptive/distracting</td>
<td>$F_d$</td>
<td>$I_d$</td>
<td>$E_d$</td>
</tr>
<tr>
<td>No. of students ‘resting’</td>
<td>$F_r$</td>
<td>$I_r$</td>
<td>$E_r$</td>
</tr>
</tbody>
</table>

These totals are combined and used as shown below. In addition, the user is given some control over the relative weighting of each of these influences:

\[
T_{prox} = \frac{1}{1 + (0.1 \times (distance - 1)^2)}
\]

**Teacher proximity**

The influence of the distance between student and teacher: being closer to teacher increases productive score and reduces other scores. It is based upon the NetLogo ‘patch distance’ between a student agent and the teacher, which is at least 1 unit:

\[
T_{prox} = \frac{1}{1 + (0.1 \times (distance - 1)^2)}
\]

**Influence of peers**

One of the assumptions is that a student may be inclined (according to their attribute values) to copy friends and do the opposite of ‘enemies’. The ‘Friendship’ data were obtained from questionnaires completed by students. At each tick of the simulation clock, for each student, a count is taken of the state of friend, ‘enemy’ and ‘indifferent’ colleagues, leading to 9 totals:

\[
\begin{align*}
P &= \frac{1}{3} \left[ at + 10 \beta c \frac{F_p - E_p}{N} + \gamma T_{prox} \right] \\
D &= \frac{1}{4} \left[ \alpha (1 - t) + 10 \beta c \frac{F_d - E_d}{N} - \gamma T_{prox} + ds \right] \\
R &= \frac{1}{3} \left[ r(1 - e) + 10 \beta c \frac{F_r - E_r}{N} - \gamma T_{prox} \right]
\end{align*}
\]

The model proceeds by calculating the three scores then choosing the state with the maximum score (at random if equal maxima). The controls for $\alpha, \beta, \gamma, r$ and $d$ can be seen in the bottom-right of the simulation program screenshot in Figure 3 in section 4. Figure 5 in section 7 is a screenshot of the Max{P,D,R} model in use.

The development, tuning, testing and assessment of this model is still to be completed. However, the model has numerous drawbacks, so it is unclear how much benefit will arise from pursuing it.
6.2 Plausible rules models

It is here that the complexity of the classroom dynamics is being addressed. The interactions between the agents are being specified using a production-rule-based method (with stochasticity). The approach is to take an observation about what happens in lessons and abstract a rule or principal that might be at work. For example, teachers of all kinds will have experienced waiting for students to respond to an instruction, such as ‘Let’s get our books out’ or ‘Now try the next question’. Students take different lengths of time to respond, and go about the task differently, perhaps directly or perhaps first looking through their bag, etc. Some show their high motivation, some show their rebelliousness. A simplified student rule to convey this behaviour might be:

If the teacher is not teaching and I am expected to be working alone and I’m not, and I have delayed an amount of time that reflects my commitment to this lesson, then I’ll start working.

In the current NetLogo simulation this rule would be formalised separately from the simulation engine using the format 

```
#student,70
#{( state of the-Teacher != 12 ) and ( Expected-Student-State = 8 ) and
( NotInState [8] ) and WillingnessExpressed
#8
```

In this example, the student’s ‘willingness’ is needed to determine a ‘response delay’ for a student. Values for this attribute, called subject-willingness, were derived from the questionnaire data, coded as a score on a 5-point Likert scale. A variable was defined to quantify how much delay each of the 5 levels of subject-willingness would cause: if student’s willingness is 1 then wait 2 seconds, if 2 then wait 4 seconds etc. The function WillingnessExpressed could then be defined:

```
set WILLINGNESS-WAIT [ 2 4 8 16 32 ]
to-report WillingnessExpressed
  report (state-tick-count > (item (subject-willingness - 1) WILLINGNESS-WAIT))
end
```

This example highlights the interplay between conceptual modelling and data collection. The model requires data for response delays. Although the values 2, 4, 8, 16 and 32 were not established from empirical data, these values will eventually be determined from the already collected event data. However, if this had not been possible, now that a need has been identified, a further round of data collection could supply the missing values.

This approach results in many rules and the creation of numerous new variables. The creation of a comprehensive set of rules that recreate lessons acceptably accurately is seen as the core of this research.

7 FUTURE ACTIVITIES

The screenshot in Figure 5 below shows a lesson visualization and simulation running in parallel:

Figure 5 Running a visualization and simulation side-by-side
Just a glance at the plots shows that the simulation (right) has not matched the actual events (left) (but the simulation was running the simplistic Max{P,D,R} model). There is still much to do before the project has a validated ABS capable of investigating the scenarios proposed as research questions, thus making a contribution to teaching analytics and decision-support for schools. In particular:

- the formulation of metrics that characterise lessons, classes and agents, and which can be used to compare them, ideally visually
- the exploration of alternative models, tuning and validating them using the empirical data
- comparison of simulation outputs, using established methods and some novel approaches.

The intention is to pursue the development of ‘production-rules’ to describe agent decision-making, incorporating any other methods that prove useful (Balke and Gilbert, 2014). The aim is to produce a model that adequately represents the dynamics of classroom lessons and has good face validity. Such a model might be quite complex and challenging to analyse. This approach is in keeping with the ‘keep it descriptive stupid’ (KIDS) approach described by Edmonds and Moss (Edmonds and Moss, 2005). They pointed out that it might be more effective to start with a relatively complex model that more accurately describes the real system, rather than the ‘keep it simple stupid’ (KISS) approach which advocates starting as simply as possible. In both approaches, improving the match between the model and reality is typically associated with increasing complexity, but both approaches can lead to simplifications as relationships are discovered and the most important factors identified.

REFERENCES


Easby, J. (2015) GCSE and equivalent attainment by pupil characteristics, 2013 to 2014 (Revised), DfE.


**AUTHOR BIOGRAPHIES**

**FRED INGRAM** has a BSc (Hons) in Computer Science from the University of the Witwatersrand, an MSc in Information Systems from the University of South Africa, an MA in Education from Lancaster University and a BSc (Hons) in Mathematics from the Open University. Much of his career has been in IT, but he was also a teacher for 16 years. He is now pursuing a PhD at Lancaster University.

**ROGER J. BROOKS** is a lecturer in the Management Science department at Lancaster University. He has a B.A. (Hons) degree in Mathematics from Oxford University and an MSc (Eng) and PhD in Operational Research from Birmingham University. He is co-author with Stewart Robinson of a textbook on simulation. His research areas include conceptual modelling, agent-based simulation – health applications, and sports statistics.