Real-Time Prediction of Students' Activity Progress and Completion Rates

Louis Faucon[1], Jennifer K. Olsen[2], Stian Haklev[3], Pierre Dillenbourg[4]

Abstract

In classrooms, some transitions between activities impose (quasi-)synchronicity, meaning there is a need for learners to move between activities at the same time. To make real-time decisions about when to move to the next activity, teachers need to be able to balance the progress of their students as they work at different paces. In this paper, we present a set of estimators that can be used in real time to predict the progress and completion rates of students working on computer-supported activities that can be divided into sequential subtasks. With our estimators, we investigate what effect the average progress rate of the class, a given number of previous steps, or weighting the proportion of progress assigned to each subtask has on predictions of students' progress. We find that accounting for the average class progress rate near the beginning of the activity can improve predictions over baseline. Additionally, weighted subtasks decrease prediction accuracy for activities where the behaviour of faster students diverges from the average behaviour of the class. This paper contributes to our ability to provide accurate student progress predictions and to understand the behaviour of students as they progress through the activity. These real-time predictions can enable teachers to optimize learning time in their classrooms.

Notes for Practice

- This paper advances our knowledge of the challenges of time management and the characteristics of students' progress within single activities in classrooms.

- Several estimators of students' progress rates and the proportion of time required by each subtask of an activity are proposed and tested on student data from 12 university courses ranging from 64 to 241 students.

- The estimators have an average error of less than 2% when predicting the average progress of the class, which can help teachers optimize their use of class time.

Corresponding author [1] *Email: louis.faucon@epfl.ch Address: Computer Human Interfaces for Learning and Instruction, EPFL, CH-1015, Lausanne, Switzerland*

[2] *Email: jennifer.olsen@epfl.ch Address: Computer Human Interfaces for Learning and Instruction, EPFL, CH-1015, Lausanne, Switzerland*

[3] *Email: stian.haklev@epfl.ch Address: Computer Human Interfaces for Learning and Instruction, EPFL, CH-1015, Lausanne, Switzerland*

[4] *Email: pierre.dillenbourg@epfl.ch Address: Computer Human Interfaces for Learning and Instruction, EPFL, CH-1015, Lausanne, Switzerland*

1. Introduction

Let us consider a simple sequence of two learning activities. First, learners individually solve a problem. Second, the teacher forms pairs of learners who produced different solutions in the first activity and asks them to solve a new problem. This scenario is a typical collaborative script. Now, what happens if the teacher gives the students 10 minutes to complete the individual activity, but after 10 minutes, only 90% of the students have produced a solution? This is a frequent problem. The teacher is then confronted with a difficult decision: extend the time allocated for the first activity or force the transition to the next activity with some students being paired randomly. This decision constitutes an ethical dilemma: have 90% of the learners wait (knowing idle learners will often engage in distracting off-task activities) or sacrifice the 10% of learners who do not have a solution. The teacher could make a more informed decision if a system could provide information such as "the completion rate

of the first activity will be around 96% in two more minutes and 100% in nine more minutes." This example illustrates the goal of this contribution: *predicting completion time to optimize the timing of transitions between activities.*

Our purpose is not to point out that individuals have different learning paces; this is one fundamental issue addressed by adaptive instruction. Neither is it our point to remind the reader that the total time budget is restricted and quite often the same for all learners, which creates a tension with the previous point. What our example emphasizes is another timing difficulty: some transitions between activities impose (quasi-)synchronicity (i.e., the need for (almost) all learners to move at (almost) the same time to the next activity). Such synchronicity constraints come from the dependency between the two activities. In the example, the fact that solutions collected in the first activity are used to form pairs in the second one constitutes an orchestration dependency. Instructional design, especially mastery learning theory, has traditionally focused on prerequisite dependencies. Instead, our example illustrates organizational dependencies.

This non-pedagogical dependency reveals the very pragmatic flavour of research on classroom orchestration. In our example, some learners may finish later for reasons endogenous to learning (e.g., a slower learning pace may be due to weak prerequisites or increased cognitive load) but are often due to extraneous pitfalls (e.g., some students join late, some need to pause, some have to reinstall Flash in the middle of the activity). Both the instructional time (i.e., the time devoted to learning activities) and the non-instructional time (i.e., the time spent on non-pedagogical activities such as distributing material, moving chairs for teamwork, and extraneous events) can change the amount of time needed for an activity. Thus, orchestrating a sequence of classroom activities requires dynamic timing, where the time allocated for each activity must be balanced between the whole lesson and the time needed for the individual activities.

To summarize this introduction, let us define the key concepts that guide our contribution:

- Classroom orchestration is the management and adaptation of learning activities in real time to account for events as they occur in the classroom (Dillenbourg, 2013; Kollar & Fischer, 2013; Prieto, Dlab, Gutiérrez, Abdulwahed, & Balid, 2011).

- Transitions are the (supposedly short) periods when learners move from one activity to the next.

- An orchestration dependency exists between two activities, A and B, if A sets up logistical conditions for B (e.g., A produces data used by B). If A is a prerequisite to B, there is an instructional dependency between A and B.

- Dynamic timing is the flexibility to adapt the timing of activities in real time to align with the current needs of the classroom.

In this paper, we aim to contribute to classroom orchestration by investigating how we can predict student progress in real time during an activity to support teachers' decision-making in their classroom. Specifically, we are interested in predicting the progress that students will make over a period of time and the time of completion of the activity based solely on the actions that have been taken in the activity, without any previous information. In this way, our model does not have to be calibrated on a new activity or group of students. Our models were tested on the data gathered across 12 activities run in 4 large lectures. By using 12 datasets, we can test the algorithms across a range of groups and activities to make stronger claims on the generalizability of our results. Moreover, because we are using data collected during class activities as opposed to in the lab, our data will follow more authentic completion patterns. The range and authenticity of the data allow us to investigate the impact that different classroom features have on the progress and completion predictions. The remaining sections of this paper are organized as follows. In Section 2, we discuss related work supporting and modelling classroom timing. In Section 3, we present the formulation of the problem with the datasets and context presented in Section 4. In Sections 4 and 5, we present our proposed estimators and results, respectively. Finally, in Sections 6 and 7, we discuss the results and contributions.

2. Related Work

2.1 Timing Instruction

As our simple example in the introduction illustrated, within formal education settings, the amount of time for instruction on a certain topic is limited (Dillenbourg, 2013; Karweit & Slavin, 1981), so teachers must continuously consider how to balance the time spent on any activity so it is most productive for the class as a whole. However, in the classroom, it can be difficult for teachers to monitor and track the state of all students. It is important to balance giving students enough time to work on an activity so that it is useful for their learning with other activities that need to be accomplished. Normal classroom cues, such as the noise level, may not be appropriate indicators of classroom completion because some students may be discussing the task while others are engaging in off-task behaviour. These factors make it difficult for teachers to estimate the appropriate time to switch to the next activity. Learning analytics could be used to help teachers make these timing decisions.

Within classrooms, there are different classifications of academic learning time at which different interventions can have different effects (Gettinger & Seibert, 2002; Karweit & Slavin, 1981). At the highest level is the *allocated* (planned for) *time* that the teacher has provided for the activity. In the classroom, external constraints can lead to not all of the allocated time being used for instruction (Dillenbourg, 2013). The actual time used for an activity in the classroom is the *instructional time*. Different individual factors of the students, such as their motivation (Malone, 1981) or prior knowledge, can change how students interact with the activity, as well as the use of instructional time, which is measured through the *engaged time*. Finally, although a student may still be engaged with a task, given the nature of the task, the student may no longer be productive, leading to a reduction in the *successful and productive learning time*. When considering activity transitions in the classroom, the teacher is working to maximize the engaged and productive learning times of the students.

Across these different levels of classroom timing, it has been found that an increase in instructional time has a positive effect on learning (Fraser, Walberg, Welch, & Hattie, 1987). When the instructional time is limited, students have a reduced ability to develop knowledge integration of complex topics (Clark & Linn, 2003). Because teachers do not necessarily have control over the amount of time they have with students during the day, it is important to optimize the instructional time that they do have. This can be done by decreasing non-instructional time, such as transitions between classes or activities, which, through support, has been shown to increase time on task (Campbell & Skinner, 2004; Hine, Ardoin, & Foster, 2015; Yarbrough, Skinner, Lee, & Lemmons, 2004). Learning in the classroom can also be supported by influencing the engaged time and successful and productive learning time. For example, when the time limits for an activity are explicitly announced, students increase the number of tasks completed without reducing their accuracy (Rhymer et al., 2002; Van Houten, Hill, & Parsons, 1975; Van Houten & Thompson, 1976). However, students still do not all work at the same pace and may become disengaged when they complete a task.

Within the literature, the primary way that student pacing is addressed is through individual personalized learning (Song, Wong, & Looi, 2012; Xu, Wang, & Wang, 2005). In personalized learning, students work at their own pace and on topics that benefit their learning. Although time is still limited in the classroom, each student can potentially use this time optimally. However, within personalized learning, students must often work individually, because once dependencies are introduced with other students in the class, timing matters again. For example, when students are working collaboratively or as a whole class, they depend on group members or classmates to be ready for the activity. Even on an individual level, such as with peer grading, there may be dependencies where students rely on other students to be ready before they can continue to the next activity. These dependencies can lead to off-task time, where some students finish early and have to wait for other students to be ready. Even giving students a productive task to work on during this time may still have a negative impact (Dillenbourg, Li, & Kidzinski, 2016). To reduce this wait time, teachers need to be able to make informed decisions about when to move to the next task that are optimized for the entire class.

2.2 Progress Awareness Tools

To support teachers in optimally deciding when to move to the next task, there needs to be support for classroom orchestration. The definition of classroom orchestration includes the management and adaptation of learning activities in real time to account for events as they occur in the classroom (Kollar & Fischer, 2013; Prieto et al., 2011). Although learning activities are often planned in advance, variability occurs in the classroom that must be accounted for in real time. Two examples of this variability are groups needing to be changed because a student is absent and students taking more or less time on an activity than planned. In these cases, the teacher plays a central role in assessing the current state of the classroom and adapting the learning activity to fit the current needs of the students, given the external constraints of the learning environment (e.g., limited time, physical classroom environment) (Dillenbourg, 2013; Dillenbourg et al., 2011).

Although much of the current research on classroom orchestration focuses on supporting teachers in recognizing and identifying student work, such as through the identification of misconceptions, wheel spinning, or division of labour (Holstein, Hong, Tegene, McLaren, & Aleven, 2018; van Leeuwen, van Wermeskerken, Erkens, & Rummel, 2017), real-time classroom orchestration also entails supporting the transitions that occur between activities (Dillenbourg, Prieto, & Olsen, 2018). When orchestrating these transitions, the teacher has to consider what transitions are occurring, how the students are going to transition, and when the transition should occur. In the classroom, the when is often the most variable as it can depend on the instruction time and must be balanced between the other activities (Dillenbourg, 2015). An unexpected event in the classroom can force the teacher to re-balance the time for all activities. Even when students are on task, an activity may take more or less time than originally planned and students will complete the activity at different times. All of these factors make timing an important aspect of orchestration that teachers may have difficulty monitoring on their own.

When helping teachers make decisions about when to have an activity transition, current orchestration systems and teacher dashboards account for activity progress in one of two ways: tracking student progress and tracking the passage of time. One

metric that teachers can use to understand when students should switch to the next activity is students' progress in the current activity. Only when a student has completed the current task can they move on to the next. The progress that students are making is not always visible to teachers. Visualizations can be used to increase teachers' awareness of student progress (Slotta, Tissenbaum, & Lui, 2013; Wang, Tchounikine, & Quignard, 2017). One method of visualizing progress is to use mastery grids or heat maps (Loboda, Guerra, Hosseini, & Brusilovsky, 2014). Mastery grids and heat maps display a grid of students and different activities. Each square represents the progress that a specific student has made on a specific activity. This visualization allows the teacher to quickly assess how far students or a class have progressed through a set of activities and where they may be struggling. These visualizations are often used after the fact for reflection (Charleer, Santos, Klerkx, & Duval, 2014; Verbert et al., 2014) but have also been used in real time for teachers to track student progress to better understand which students may need more teacher support (Molenaar & Knoop-van Campen, 2017).

On the other hand, progress through an activity has also been tracked using the passage of time. These types of visualizations are more common as part of systems specifically designed to support classroom orchestration. Orchestration systems can passively indicate the time lapse that has occurred, such as by an increase in a progress bar or a marker on a timeline (Håklev, Faucon, Hadzilacos, & Dillenbourg, 2017; Martinez-Maldonado, Clayphan, & Kay, 2015; Phiri, Meinel, & Suleman, 2016). Because the time that will be spent on a task is often decided during planning, an orchestration system can also assess the time that has passed. The MTDashboard project (Martinez-Maldonado et al., 2015), in addition to showing students' progress through the entire lesson, has an overtime alarm that alerts the teacher to how much over the allotted time they have gone. Similarly, in the Lantern project (Alavi & Dillenbourg, 2012), the colour of a group's light indicates the progress they have made through the problem set, but the speed of the flashing indicates how long they have been waiting for help, which indicates the passage of time to the teacher.

Although both student progress and the passage of time provide different types of information to the teacher, they have not often been combined into one visualization. In some tools, the teacher can track student progress and time through separate visualizations (Martinez-Maldonado et al., 2015). However, in these systems, the teacher is still responsible for interpreting the visualizations separately and then combining the information. Additionally, these visualizations do not provide a prediction, meaning that the teacher can only make decisions based on the current state of students, which is often too late. In this paper, we aim to develop models that combine the aspects of time, progress, and prediction. In other words, through the information provided by the model, the teacher could determine the expected progress that the class will have made in three minutes (or any given amount of time) or how long until the class has reached 95% progress (or any given progress), which is not possible with current methods.

2.3 Modelling Student Behaviours

In terms of analyzing time, the focus has primarily been on finding patterns in student habits or in predicting completion. When students are working in a self-paced environment, such as when learning in a MOOC, they can determine their own schedules and when to work. What time during the day, during the week, and across the whole course students choose to work has been found to be related to learning outcomes and performance (Boroujeni, Sharma, Kidziński, Lucignano, & Dillenbourg, 2016; Fiel, Lawless, & Brown, 2018). However, these models primarily attempt to find patterns in how students use their time rather than predicting timing. Other research has focused on the overall goal of whether students will complete a course (Crossley, Paquette, Dascalu, McNamara, & Baker, 2016; Dekker, Pechenizkiy, & Vleeshouwers, 2009). Using features of student activity while learning, such as clickstream data or prior test scores, we can predict student completion of a class or program. Although this research relates to completion, it relates more to the concept of *whether* a student will reach completion rather than *when* they will reach completion.

2.3.1 Real-Time Predictions

Prediction algorithms are one of the most popular methods in learning analytics and educational data mining. Prediction methods allow us to infer a predicted variable based on other variables in the data. They have been used to predict future events, such as student completion/dropout (Crossley et al., 2016; Dekker et al., 2009) or knowledge tracing and formative assessment (Feng, Heffernan, & Koedinger, 2009; Ming & Ming, 2012; Pavlik Jr., Cen, & Koedinger, 2009), or to predict variables that may not be feasible to collect in an educational setting, such as engagement (Beck, 2005) or affective states (D'mello, Craig, Witherspoon, Mcdaniel, & Graesser, 2008).

Using prediction methods in real time, interventions can be put in place before it is too late to have an impact on the student's learning. A well-known application of predictions in real time is Bayesian knowledge tracing (Corbett & Anderson, 1994) to track student knowledge while using an intelligent tutoring system. Based on this knowledge tracing, new problems for the student can be chosen to match their skill level (Vanlehn, 2006). Making predictions in real time has unique challenges in that the data used for the prediction has to be easily accessible. In many systems, log data is used even when predicting

more complex phenomena. For example, response times can be used to predict engagement (Beck, 2005), student actions can be used to detect student misuse of software (Baker, Corbett, & Koedinger, 2004), and affective state can be inferred from conversational cues (D'mello et al., 2008). Using log data allows models to be developed in a non-intrusive way while still being able to be used in real time.

Real-time predictions require appropriate data processing and restrictions to support the algorithms and students' models. For example, predicting next week's dropout in MOOCs can only be done with the data collected for the current week and the previous weeks, but never with the data from future weeks (Taylor, Veeramachaneni, & O'Reilly, 2014). A solution to this problem has been to use transfer learning (Boyer & Veeramachaneni, 2015; Whitehill, Mohan, Seaton, Rosen, & Tingley, 2017). Transfer learning allows us to train models using data from the final weeks of previously finished MOOCs and use this acquired knowledge to correctly predict students' behaviour for the final weeks of the currently running MOOC. Another proposed solution uses proxy labels in real time from already collected data to train algorithms to predict future data (Whitehill et al., 2017). Our approach in this paper faces the same difficulty in its real-time predictions. The estimators we propose in this paper cannot compute optimal parameters due to unknown future data. We apply methods similar to proxy labelling by training our algorithms solely on data collected during the current activity in the classroom (see Section 3.3 for a more detailed description of our method of handling the real-time aspect of our data collection and prediction mechanism).

2.3.2 Aggregating Classroom Information

Before a visualization can be developed to be displayed to a teacher, we first need to be able to accurately model classroom behaviours. In the areas of research on educational data mining and learning analytics, models of students have a predominant place. Many of the methods focus on finding different patterns in the data that can be used to provide teachers with more information and to inform their decisions in their classroom (Baker & Siemens, 2014). When used to make visualizations, the outputs can be interpreted by humans to make decisions.

The majority of models that are developed target modelling individuals, which can inhibit the teacher in making classroom-level decisions. When information is aggregated, it may be to show the collective student knowledge to get a more complete picture of what the class knows as a whole (Slotta et al., 2013). Additionally, aggregated data can be used to see overall class patterns, such as average activity and actions when video watching (Wachtler, Khalil, Taraghi, & Ebner, 2016) or the behaviours that students engage in across a course (Jovanović, Gašević, Dawson, Pardo, & Mirriahi, 2017). Although these methods can provide useful information about student behaviour and actions over time, they cannot make predictions based on students' previous and current actions.

3. Formulation of the Problem

The problem addressed in this paper consists of designing estimators to predict the future progress of students during activities in the classroom. Specifically, we focus on tasks that can be divided into discrete steps, such as quizzes, as opposed to open-ended tasks, such as essays or code writing. The estimators described in Section 4.2 below can be used in any learning environment that automatically extracts estimates of students' progress in real time. In technology-enhanced classrooms, this extraction can easily be integrated into most learning software.

Before describing the estimators of students' progress in Section 4, we first discuss the key aspects of the problem. In Section 3.1, we define the concepts of progress, progress rate, and completion and explain the context in which our estimators will be implemented and used. In Section 3.2, we point to the main factors and challenges influencing the design, in Section 3.3 we discuss the specificity of real-time predictions, and in Section 3.4 we explain how data from the whole class can improve estimations of individual students' progress.

3.1 Definition of Progress and Completion

We define the notion of *progress* using a value between zero and one that estimates how far through the task students have advanced. A value of zero corresponds to the student not having started the activity, while a value of one corresponds to the student having completed the activity. Reaching *completion* in an activity means that the student has completed all required subtasks and, thus, will wait until the teacher transitions to the next activity. In this sense, we can think of the progress as the percentage of the activity that has been completed if completion is binary, being either done or not. Progress as a percentage can be interpreted in two different ways. It can be evaluated according to the percentage of the subtasks completed or the percentage of the time already passed out of the total time required. Counting the number of subtasks completed is easily implemented within a learning environment. On the other hand, we must predict the amount of time students will need so we can measure their progress in terms of time. Progress in terms of time needed can be a more interesting measure for the teacher but unfortunately cannot be directly known. For this reason, we prefer a definition in terms of percentage of subtasks completed.

In line with this definition of progress, we define *progress rate* as the speed at which a student's progress is increasing over time. We assume that a student cannot make backward progress; that is, once a student has completed a subtask, it remains completed.

We also assume that student progress cannot be actively and precisely measured at all times. Instead, progress is reported based on milestones reached. When students are working on an activity, progress may be made both inside and outside the orchestration system. For example, on a multiple-choice quiz for mathematics, students may use paper to work out the problem before selecting the solution in the system. As they work on the problem on paper, they are making progress on the problem that cannot be precisely measured by the system. With our estimators we restrict ourselves to working with discrete measurements of student progress that the system is able to provide.

Finally, we restrict ourselves to learning activities that can easily be divided into subtasks. Whenever a student completes a subtask, we collect their student ID, time, and progress. For every student with ID s, we then have a set of tuples $(t_{s,0}, p_{s,0}), (t_{s,1}, p_{s,1}), ..., (t_{s,n}, p_{s,n})$. We define the progress at any time t as the linear interpolation from equation (1), where $t_{s,0}$ is the time when the first log is received, which corresponds in our case to a progress $p_{s,0}$ of 0; t is any time during the activity; and $t_{s,k}$ and $t_{s,k+1}$ are two consecutive times at which logs are received with the respective progress of $p_{s,k}$ and $p_{s,k+1}$:

$$p_s(t) = \begin{cases} 0 & \text{if} \quad t_0 \geq t, \\ p_{s,k} + (p_{s,k+1} - p_{s,k}) * (t - t_{s,k})/(t_{s,k+1} - t_{s,k}) & \text{if} \quad t_{k+1} \geq t > t_k, \\ p_n & \text{if} \quad t > t_n. \end{cases} \tag{1}$$

3.2 Primary Challenges

In an ideal world, all students would complete the activity at the same time and work at a steady pace. In classrooms, this does not happen because both endogenous and exogenous pitfalls influence when individual students complete an activity. Challenges in predicting student progress arise due to these individual differences between students. In this section, we present some of the primary challenges to predicting progress, including different start times and varied progress rates between subtasks.

3.2.1 Variability in Students' Start Times

A first challenge for predicting student progress is that students do not all start the activity at the same time for a number of reasons, such as being distracted or taking longer to start their computer or sign in. For example, Figure 1 shows two distributions of the delays of students for starting an activity. The right distribution shows a majority of students starting the activity within 20 seconds of the teacher making the activity available. On the other hand, the left distribution shows a significant proportion of the students starting the activity one or even two minutes after the activity is made available. These examples demonstrate the range of starting situations that can occur depending on different parameters in the classroom. Variability in start times is more of a challenge for teachers, leading to more variability in completion times and a larger variance in student progress.

3.2.2 Variability in Student Progress Rates

A second challenge is accounting for student progress rates, which are often a source of individual differences. For a wide range of possible reasons, different students will require different amounts of time to complete a given activity. Knowing this time distribution of the classroom is a challenge for teachers. A good way to quickly evaluate the spread of the distribution of individual progress rates is to measure how much more time the slower students need as a percentage of how much time the faster students use. For this we can choose, for example, to use the 25% and 75% quartiles of the time distribution to correspond respectively to our estimate of the time needed by faster and slower students. Using this measurement, our data from twelve courses (described in more detail in Section 4.1.2) has three courses where the slower students needed more than 50% more time than the faster students, four courses where the slower students needed between 30% and 50% more time than the faster students, and five courses where the slower students needed between 20% and 30% more time than the faster students. The two histograms in Figure 2 show two examples of activities that have different spreads. The distribution on the left corresponds to one of the courses with a spread of more than 50% according to our proposed measure, and the activity on the right has a spread of less than 25%. These individual differences in progress rate again pose a challenge for the teacher to decide how much time to allocate to an activity when students do not all need the same amount of time.

3.2.3 Imprecision in Expectations of Time Distribution Between Subtasks

A third challenge comes from the lack of knowledge about the proportion of time that each subtask will consume out of the total time. By default, each subtask has an equal proportion of the progress if there is no prior knowledge about the tasks, but,
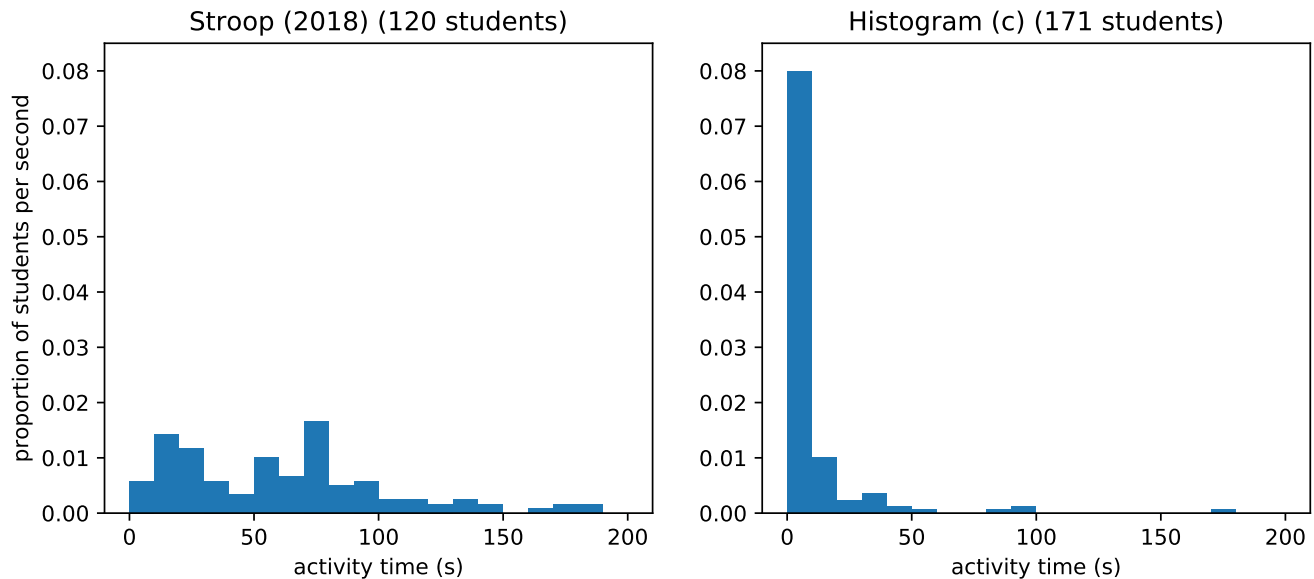
Figure 1. Examples of distribution of students' starting delays for two activities in different classrooms.

in most cases, each subtask does not take the same amount of time to complete. For example, Figure 3 shows the distribution of the proportion of time each student used for each question of a six-question quiz. We can see on this graph that the average time spent on each question varies significantly between the students and between the questions. This quiz is an example of a default progress assignment, with each question receiving an equal amount of progress, that fails to correctly capture the time students really spend on each subtask. In this example, the naive assignment would assume approximately 17% of progress for each question, while we can see from the figure that question 1 had an average time proportion above 25% and the sixth question of less than 10%.

From our experience in classrooms, we identified several factors that can make students' progress rate vary during the activity. One factor could be the skills and preferences of students, which will modify their progress rates accordingly. Additionally, it is not uncommon to design a quiz where questions become harder over time and, thus, require more time to be solved. Another factor is that students are getting used to the task they are doing, which will increase their progress rate (in Figure 3, we see a general downward trend in the proportion of time required by students for each subtask). Finally, in class the teacher will also interact with students either to explain a difficulty or to make them aware of time passing. These interventions are not often recorded but can affect students' progress rates. In general, it is not surprising if not all the subtasks of an activity require the same amount of time to complete. This poses a challenge for estimating and predicting student progress because it makes it difficult to extrapolate the time required by future subtasks from the time used on the previous subtasks.

3.3 Real-Time Predictions of Future Progress

To predict students' future progress, estimators make predictions in real time, implying that at any given time, the estimators must be able to make predictions based solely on the data collected until this particular time. Thus, the estimators have to work with a varying amount of data throughout an activity. At an early stage of an activity, students will have produced only a limited number of logs, but at a later stage, a number of students will have already finished and others will have produced a larger number of log messages. The problem of the prediction for an individual student is illustrated in Figure 4. The figure shows five black dots representing five pairs of time and progress $(t_{s,i}, p_{s,i})$, as defined previously. This example shows an activity with 10 subtasks, so the expected log messages received from the students will be $0, 0.1, 0.2, \ldots, 0.9,$ and 1. In this example, so far, the student has only completed 40% of the activity so has generated logs only up to $p_{s,n} = 0.4$. Another input used by the predictions is the current time, represented on the figure as a vertical black line. We call this the time of prediction and use the notation t_{pred} in the rest of the paper. The estimators must predict students' progress from the latest received log message to the current time t_{pred} and from the current time to completion, which may be several minutes. Figure 4 (left) shows as dashed lines possible predictions using interpolations of the student's progress curve.

The predictions of current and future progress cannot be based solely on the student's logs as input. Figure 4 (right) illustrates this point. The blue dashed line shows a situation in which interpolation from only the sequence of logs completely
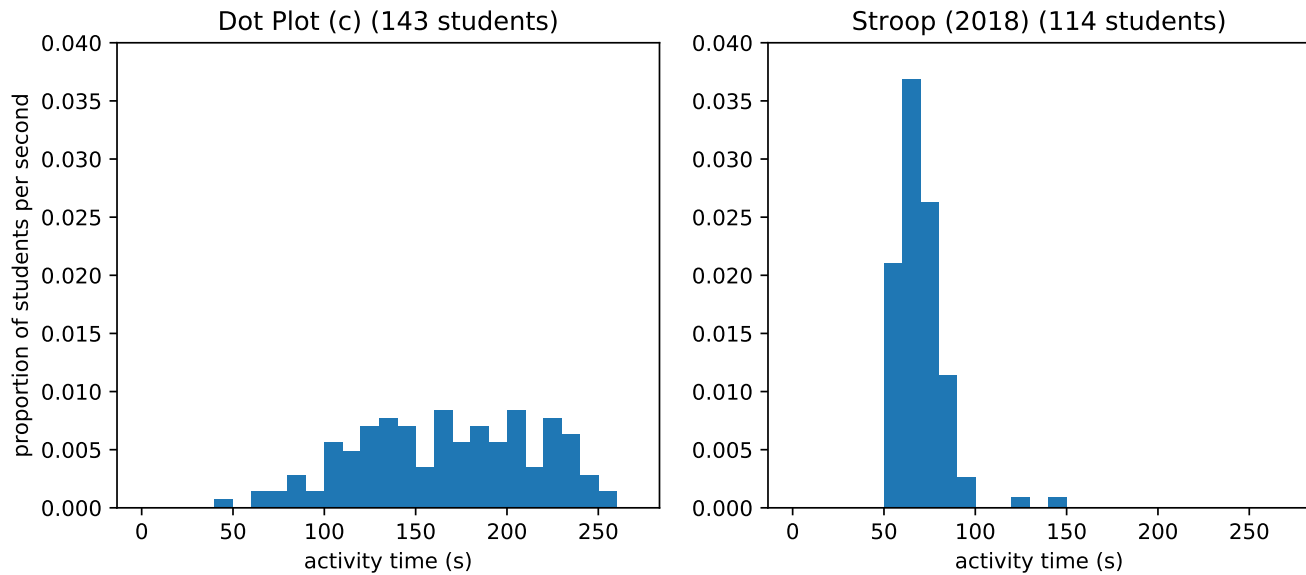
Figure 2. Examples of distribution of the time students take to complete two activities in different classrooms.

fails to realistically predict the progress made between the time of the last received log, $t_{s,n}$, and the current time of prediction, t_{pred}, thus also rendering the prediction of future progress incorrect. Indeed, both the current time of prediction t_{pred} and the progress steps expected from the student matter to the prediction of future progress. The green dashed line shows a prediction that also uses the current time and the expected progress step (0.1 for the example). The information from the five log messages collected is not enough to correctly estimate the student's progress. An equally important piece of information is that no log message has been received between the time of the latest log, $t_{s,n}$, and the time of prediction, t_{pred}.

As part of the real-time aspect of the problem, we also introduce the notion of range of prediction, which is the amount of time in the future for which the estimators compute predictions. In practice, the range of prediction can be arbitrarily long, but predictions should be expected to be less accurate further in the future. In Section 5, we use fixed prediction ranges because they are necessary for evaluating our proposed estimators. Furthermore, the limited prediction range is also forced for the evaluation of the estimators because no data was collected beyond the moment when the teacher decided in class to interrupt the activities.

3.3.1 Interpolation Method

The small number of messages received for each individual student can be interpolated using several models. Figure 4 (left) shows, for example, a possible interpolation using a linear (constant) progress rate (green) or using a logistic curve fitted to the student's progress logs (blue). From the data collected on 12 courses described in Section 4.1.2, we could observe that the curves of individual students' progress are well approximated by linear curves. As we have seen from the example in Figure 3, the subtasks do not usually take exactly the same amount of time. However, no pattern, such as the increasing speed followed by decreasing speed of a logistic curve, could be found. Thus, it was natural to make no prior assumptions on students' progress rates and distribution of time among subtasks. This is why we decided to use linear models. Furthermore, we also investigate in this paper other estimators that will learn the evolution of students' progress rate and the distribution of time among subtasks from the classroom data collected in real time without prior assumptions.

3.4 Using Whole-Class Data

Given the variance in students' start times and progress rates, it is likely that they will progress through a task in a very asynchronous manner. However, information from students who are further ahead in the activity can be used for students that have not yet reached the same point. In addition to a student's past performance being used to predict their progress, the progress from the rest of the class can be used to improve the prediction in some circumstances. For example, for a student who just started, no informed prediction can be made because we do not have any information on their progress rate. However, using the progress that other students have made through the activity, we can make an informed prediction for this student if we assume some sort of similarity with earlier students.
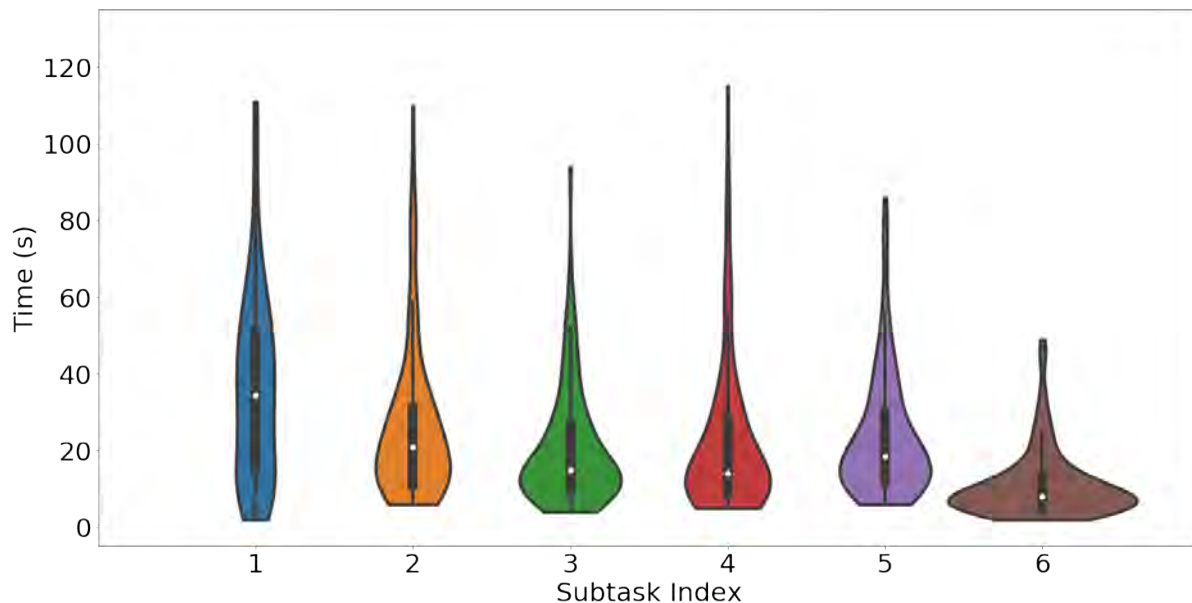
Figure 3. Distribution of proportion of the time used for a six-question quiz for an activity that lasted approximately 260 seconds.

## 4. Methods

4.1 Datasets and Educational Context

4.1.1 Learning Environment

The data we used in our analyses was from active-learning activities conducted using the FROG platform (Håklev et al., 2017). FROG is a web-based platform for designing and executing orchestration graphs, which range from simple sequences of activities to rich integrative collaboration scripts in classrooms. With FROG, students can engage in learning activities across whole-class, collaborative, and individual activities with the ability to share data between activities and use the data to adapt learning activities and make collaborative groupings. With FROG, the teacher can orchestrate the learning activities in real time using teacher dashboards for the different activities and orchestration controls to advance students through the lesson. On the student side, each planned activity is displayed on students' personal devices only when that activity is active. All interactions that students or teachers have within FROG are logged and can be used for the real-time dashboards. We focused on activities conducted using the FROG platform because it collected data at the interaction level with recorded timestamps and progress through an activity (discussed in more detail below), giving us the information needed to track progress.

Because FROG supports a wide range of learning activities, not all activities produce the same types of data, and the information that is logged for a student depends on the activity in which they are engaged. To log progress, we restricted the activities to types that had natural divisions into subtasks, as mentioned above. For example, a quiz can be divided into multiple questions, or a problem set can be divided into the individual problems. Each time a student completed a subtask within the activity, we logged this progress in addition to when the student started the activity (i.e., progress value of zero). The prediction methods presented in this paper are not restricted to the FROG platform but do require data of this nature to be collected, where the progress of each student is logged with a timestamp throughout the learning activity. Using this individual data for each student, we can then aggregate individual progress and completion rates (i.e., progress value of one) to provide an overall class progress and completion rate along with the predicted future trajectory of both measures (see Figure 5 for an example of how this could be displayed).

4.1.2 Datasets

All of our datasets come from two different courses in 2018 and 2019 in which teachers used FROG to support learning activities. The first course was a statistics course for bachelor's students. The course had about 600 students divided into three sections. We used progress logs for two activities, each given once to each section. The activities consisted of two quizzes concerning different data representations (dot plot, boxplot, and histogram). While answering the quiz, the students were also interacting with a data visualization tool that aimed to help them explore and understand the different representations (see
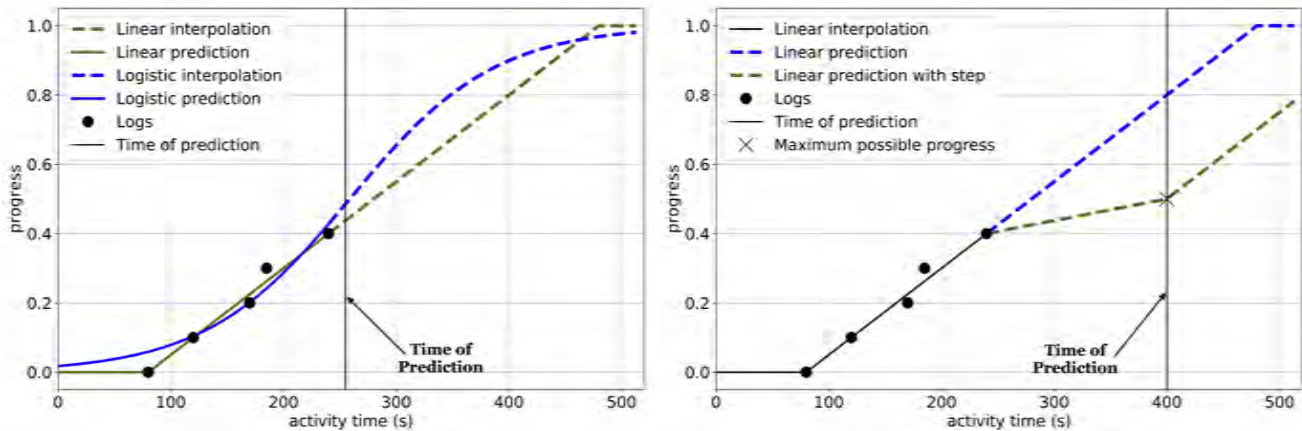
Figure 4. Examples of prediction of a student's future progress using linear interpolation (green, left) and logistic curve fitting (blue, left) and prediction of a student's future progress after a delay without observing any progress from the student (right).

Figure 6). The two quizzes were taken by each section during their two-hour lecture. These activities are referred to as A1 to A6 in Table 1, and the three groups are referred to as (a), (b), and (c).

The second course was given in both 2018 and 2019 for bachelor's students in computer science. The classes had about 120 students, but attendance was often low. Both years, during three different lectures, the students did three experiential activities that aimed to teach concepts of cognition and user interfaces. The three activities were Genealogy Puzzle, Stroop Effect, and Train Ticket Simulator. Genealogy Puzzle is an activity for experiencing the effects of cognitive load. Students had to solve logic problems based on family trees of increasing difficulty, necessitating more effort and working memory. In the Stroop Effect activity, students had to answer a series of very quick questions identifying a word (of a colour) as the text changed colour. Finally, in the activity Train Ticket Simulator, students used four different user interfaces to order three train tickets. These six activities are referred to as A7 to A12 in Table 1.

The datasets from these classes were selected because they provide a range of activities and include students from different institutions and disciplines. For each activity, we had a dataset from more than one class, allowing us to consider variability between students. Moreover, for each class, we had data from more than one activity so we could account for the variability between activities. Because the datasets came from a class activity and not a designed experiment, no personal information was gathered from the students. All of the students were informed before participating that their data may be used for research purposes and were asked for consent. If consent was not provided, the data was deleted after that activity for that participant.

Table 1 shows the number of students who participated, the number of progress log messages received, the number of subtasks for each learning activity, and the length of each activity. Note that three questions were removed from the activity Dot Plot following the lecture given to the first group. This does not affect our analysis because the 12 activities are considered separately and no analysis is made between the repeated activities. Furthermore, for some activities, the number of log messages exceeds the product of the number of students and the number of subtasks. This is because extra log messages are triggered when the activity is activated by the teacher and changing an answer to a previously completed subtask triggers a repeated log message. This does not interfere with the mechanics of the predictions that we describe in Section 4.2.

4.2 Estimators

In this section, we describe several different estimators we have implemented to predict student progress in real time. The different estimators address the specific challenges of the problem that we presented above. The first models, constant progress rate (CPR) and average progress rate (APR), address the challenge of different start times and progress rates of students, while the models local progress rate (LPR) and weighted progress rate (WPR) aim to solve the weight challenge. Finally, the models APR and WPR also use the whole class's data to inform predictions made for individuals.

4.2.1 CPR Model

Our first model assumes that students are making progress at a constant rate through the subtasks of a particular activity. The rate of progress can, however, be different for each student. This assumption leads us to use a linear interpolation to predict the progress of the students after estimating their progress rate. At the time t_{pred} when we are making a prediction, we estimate both the current progress and the APR for each student to compute the linear interpolation. From a student's
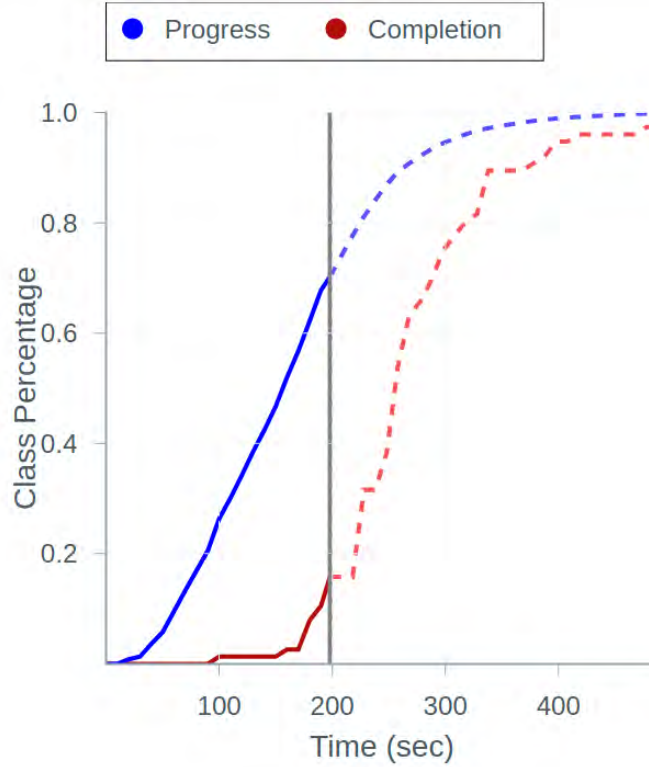
Figure 5. Real-time dashboard of aggregated student progress and completion rates on the FROG platform. The vertical line is the current time; the blue line indicates the average progress of the students and the red line the proportion of the class that has completed the activity; the dashed lines are predictions of future progress and completion rates.

logs, $(t_{s,0}, p_{s,0}), \ldots, (t_{s,k}, p_{s,k})$, with $t_{s,k} < t_{pred}$, we compute the progress rate to be the amount of progress made by the student divided by the time the student has been working. This computation is shown in equation (2), where $\hat{r}_s(t_{pred})$ is the estimated progress rate of student s, $(t_{s,0}, p_{s,0})$ is the first log entry received, and $(t_{s,k}, p_{s,k})$ is the latest log entry received before the current time at which we are making the prediction. Second, we compute the current estimate of the student's progress at the current time t_{pred} following the formula in equation (3). It is important to note that the true value of the student's progress at time t_{pred} is not known because the system only measures the discretized progress from the latest log of the student, $(t_{s,k}, p_{s,k})$. From this, equation (3) estimates the current progress, $\hat{p}_s(t_{pred})$, by adding to the latest progress received, $p_{s,k}$, the minimum of the full quantity of the progress assigned to the next subtask and the quantity of progress we expect the student to have made assuming the estimated progress rate given by $\hat{r}_s$:

$$\hat{r}_s(t_{pred}) = (p_{s,k} - p_{s,0})/(t_{s,k} - t_{s,0}), \tag{2}$$

$$\hat{p}_s(t_{pred}) = min(p_{s,k} + (t_{pred} - t_{s,k}) * \hat{r}_s(t_{pred}), p_{s,k} + step). \tag{3}$$

Using both the prediction of the progress at time t_{pred} and the estimated progress rate, we predict the progress for all time $t > t_{pred}$ with the linear function described in equation (4). This linear function has a slope of $\hat{r}_s$ and a value of $\hat{p}_s(t_{pred})$ at $t = t_{pred}$:

$$\hat{p}_s(t) = \hat{p}_s(t_{pred}) + \hat{r}_s * (t - t_{pred}). \tag{4}$$

The above formulas require collection of at least two logs to be able to make predictions for a student. In the particular case of students who have not yet completed a subtask (only one data point has been collected), the estimated default progress rate is always zero, which leads to a prediction of $\hat{p}_s(t) = 0$ for all $t \geq t_{pred}$.

Table 1. Datasets of Progress Logs Collected

	Activity	Number of students	Number of logs	Number of subtasks	Activity time (s)
A1	Dot Plot (a)	241	2891	8	664
A2	Histogram (a)	227	2538	6	355
A3	Dot Plot (b)	137	1181	5	255
A4	Histogram (b)	149	1485	6	227
A5	Dot Plot (c)	163	1490	5	266
A6	Histogram (c)	171	1756	6	264
A7	Genealogy (2018)	78	704	8	386
A8	Stroop (2018)	120	2459	20	217
A9	Train (2018)	64	782	12	562
A10	Genealogy (2019)	70	667	8	425
A11	Stroop (2019)	85	1853	20	141
A12	Train (2019)	84	1060	12	636
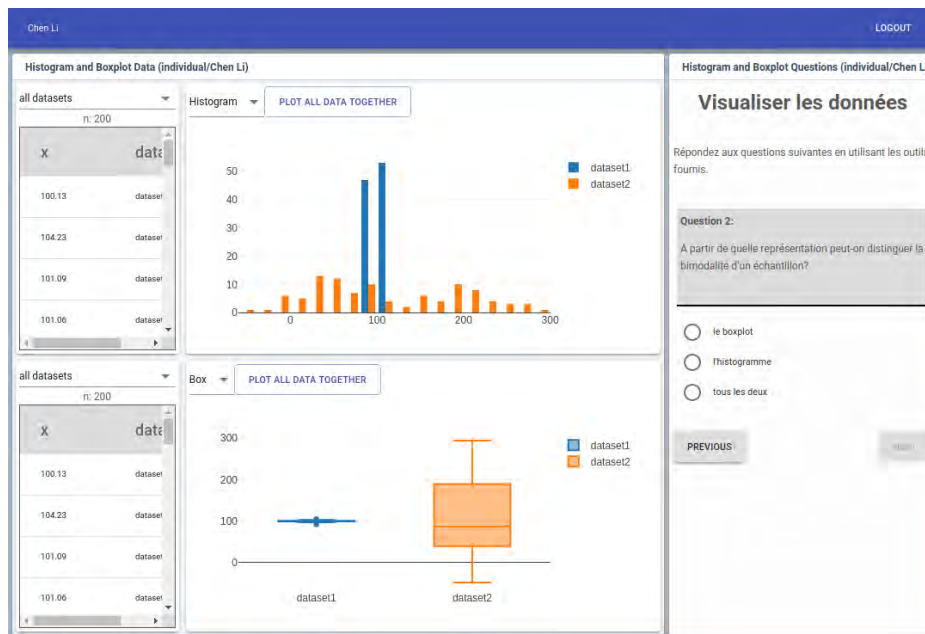
Figure 6. Screenshot of the statistics quiz activities (on the right) with the data visualization tools (on the left).

4.2.2 LPR Model

The assumption that students maintain a constant progress rate is certainly too optimistic, which motivates us to consider this model of non-constant progress rates. The LPR model makes predictions based only on a fixed number κ of the latest progress logs for each student. This method supports a change of progress rate over time, where the most recent progress logs are better indicators of future progress than older progress logs. In the model, κ is a parameter for which we can choose any value $2 \leq \kappa \leq m$, with m being the number of subtasks. Large values correspond to a model very similar to the CPR model, while smaller values only account for the student's current progress rate but are also more subject to noise in the estimation because the model uses fewer data points. In this paper, we compare values of κ from $\kappa = 2$ to $\kappa = 5$.

4.2.3 APR Model

The APR model is a variation of the CPR model. This model uses the APR measured over the data collected from the whole class (see equation (5)) to improve estimates for students' progress rates, especially for students who have only completed a small number of subtasks. Indeed, the CPR model has several limitations, such as assigning a progress rate of zero to students who have not yet completed a subtask or being sensitive to noise in the estimate of the progress rate for students with a small

number of subtasks completed:

$$\bar{r} = \frac{1}{n} * \sum_s (\hat{r}_s).$$ (5)

For this purpose, the APR model computes a new estimate of the progress rate for each student. It does so using the formula described in equation (6), where $\hat{r}'_s$ is the new progress rate of the student. In this formula, the parameter λ is the number of log messages already received from the student. The progress rate of a student with only one log message, for whom we cannot estimate the progress rate, is estimated as $\hat{r}'_s = \bar{r}$. The weight of the APR then decreases for students with more log messages:

$$\hat{r}'_s = (\bar{r} + (\lambda - 1) * \hat{r}_s)/\lambda.$$ (6)

4.2.4 WPR Model

As mentioned in previous sections, one challenge for the predictions is that the estimated progress steps for each subtask of an activity do not correctly estimate the proportion of time students need. Such discrepancies could lead to imprecision in the progress prediction. The WPR model proposes to solve this challenge by using data collected over time from the whole class to measure the proportion of time that students allocated on average to each subtask. This data allows us to transform the progress curves of students to account for the uneven time needed by each subtask. An advantage that the WPR model has over the LPR model is that it can account for more uneven progress changes instead of assuming that a student's progress rate will remain similar to that in their previous subtasks, as is assumed in the LPR model. The WPR model also assumes that the progress rate changes of the students are correlated, which is likely to be the case.

The first step for the WPR predictions is to compute a weight profile that represents the weight of each subtask of an activity. Given the constraints of the problem, the weight profile has to be dynamically estimated from incomplete data during the learning activity in the classroom. In other words, until a student finishes all subtasks, the complete weight profile cannot be known. However, the dynamic estimation cannot be accurate with too little data and, in particular, benefits the most from collecting log data for students who have fully completed the activity. For a student s who has generated the sequence of log messages $(t_{s,0}, p_{s,0}), (t_{s,1}, p_{s,1}), ..., (t_{s,n}, p_{s,n})$, the proportion of time spent on each subtask is computed using equation (7). From the equation, we see that the proportion of time used cannot be computed for subtasks that the student has not yet completed ($k > n$), so these subtasks are assumed to require a proportion $1/m$ of the student's time, where m is the number of subtasks. For the subtasks that have already been completed, the proportion of time needed for each subtask can be directly computed from the student's progress logs:

$$\hat{w}_{s,k} = \begin{cases} \frac{t_{s,k} - t_{s,k-1}}{t_{s,n} - t_{s,0}} \times \frac{n}{m} & \text{if} \quad 1 \leq k \leq n, \\ \frac{1}{m} & \text{if} \quad n < k \leq m. \end{cases}$$ (7)

To compute the weight profile using data from students who have not yet completed the activity, we assume that the proportion of time spent on the subtasks that have not been completed yet will be as assigned by the original weighting (this corresponds to a value of $1/m$ as in equation (7)). Furthermore, because students who have fully completed the activity give us more information about the true weight profile, we give more importance in the computation to the students who have completed more subtasks. This is controlled by a parameter p, which balances the amount of importance that is given. Equation (8) shows how the proportion of time used by each student is aggregated to estimate the true weight profile. In the equation, $\hat{w}_k$ is the aggregated weight given to subtask number k. α_s is the importance in the aggregation given to student s. α_s is computed as an increasing function of the student's progress and controlled by a parameter p. In Section 5, we analyze the quality of the weight profile estimation over time based on the parameter p used to average the data from students:

$$\hat{w}_k = \sum_{s \in S} \alpha_s \hat{w}_{s,k} = \sum_{s \in S} p_{s,n}^p \hat{w}_{s,k}.$$ (8)

Figure 7 is an example of an evolution of the estimated weight profile at different times during an activity (we chose to display every interval of 30 seconds). We see in the figure that the approximated weight profile estimates (blue) change over time to more precisely align with the final whole-class weight profile (orange). The final weight profile is the weight profile
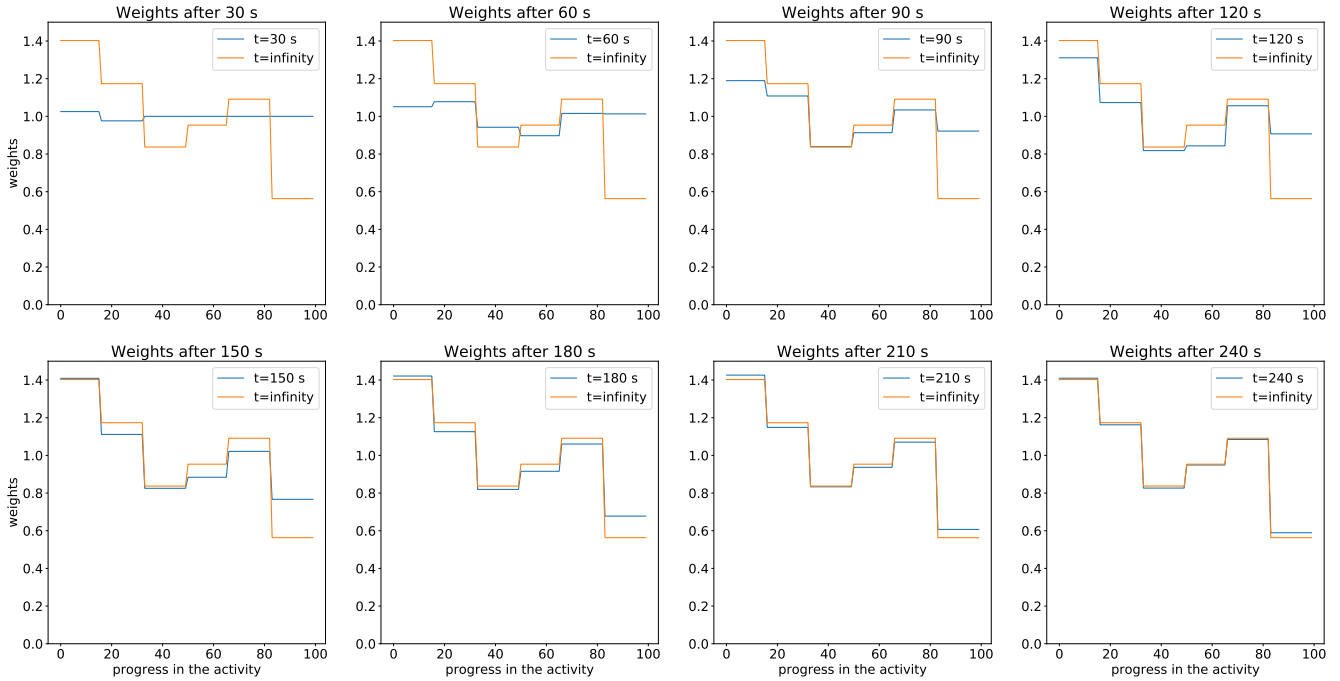
Figure 7. Computation of the weight profile over time. The blue line shows the estimated weight profiles at different times by steps of 30 seconds. The orange line is the actual target.

computed from the whole class after the end of the activity. How quickly the weight profile is correctly approximated is crucial for the WPR estimator because progress predictions will be inaccurate if the estimated weight profile is not similar to actual student behaviour.

The WPR model itself consists of several steps. First, based on the collected data at a given time, an approximation of the weight model is computed as outlined above. This model is used to adjust the proportion of progress assigned to each subtask of the activity from the initial teacher estimation during planning. Then the CPR or APR model is applied to the transformed data. Finally, the progress predictions are turned back to the original scale to be compared to the actual logs collected and to be shown to teachers in the scale that they choose.

4.3 Analysis

To analyze the different estimators, we specify a time of prediction (t_{pred}) and a range of prediction (Δ_t) for each of the 12 courses. The value of t_{pred} corresponds to a time in the class when the estimators have to make a prediction. The value of Δ_t corresponds to how far in the future the estimators are predicting students' progress. In this paper, we evaluate and compare our estimators at different times t_{pred} and ranges of time Δ_t. For each prediction point, the estimators receive only the data collected before t_{pred} and must predict student progress at every future time between t_{pred} and $t_{pred} + \Delta_t$.

For a fixed time t_{pred} and range of prediction Δ_t, we compute predictions with each of the above-mentioned estimators of the interpolated progress curves for every student based only on measurements received before the time t_{pred}. To evaluate the quality of the predictions, we measure the root mean square error (RMSE) for each student by averaging the squared difference between the interpolated progress of the student and the predicted progress at every second. RMSE was chosen because it is a widely adopted error measure for algorithms such as linear regressions (Pelánek, 2015), which are quite similar to the estimators we are presenting in this paper. Equation (9) shows the formula we used to compute the prediction error for each individual student. The sum includes the squared difference between the predicted progress and the real progress estimated for every second of time between $t = t_{pred}$ and $t = t_{pred} + \Delta_t$:

$$RMSE_{t_{pred},\Delta_t}(\hat{p}_s, p_s) = \sqrt{\frac{1}{\Delta_t} \sum_{t=t_{pred}}^{t_{pred}+\Delta_t} (\hat{p}_s(t) - p_s(t))^2}. \tag{9}$$

Because the accuracy of the prediction is different based on the amount of data collected, we evaluate and report on the
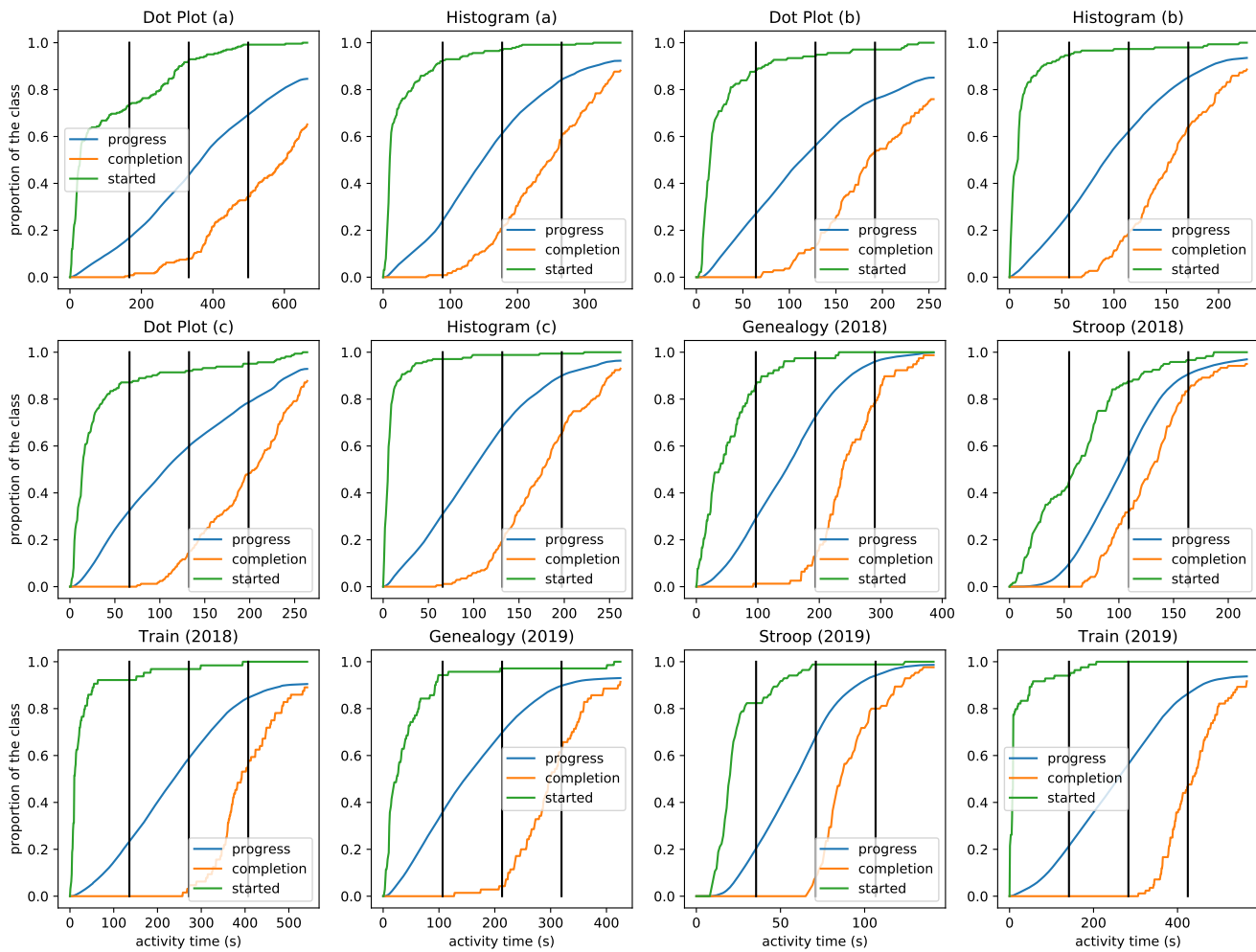
Figure 8. State of each of the 12 classrooms in terms of the delay of students joining the activity (green), the average progress (blue), and the proportion of completion (orange) over time. Each graph has three vertical lines representing the times at 25%, 50%, and 75% of the full time of the activity.

predictions of the different estimators after 25%, 50%, and 75% of the time of the activity has passed. At 25%, a majority of students have not yet completed the activity and only a limited amount of data has been collected. On the other hand, at 75%, a majority of students have already completed the activity and others have made significant progress. Figure 8 shows for each of the 12 activities the aggregated progress and proportion of completion of the class over time, but also the proportion of students who have started the activity. These are important to consider when analyzing the prediction results of our different estimators. For example, at the 25% prediction time for all 12 of the activities, only a very small number of students have completed the activity, so we should not expect the weight profiles estimation to be sufficiently accurate.

For the individual prediction errors measured and reported in Section 5, students who had already completed the activity were not included because they had already reached completion, so prediction of their future progress is trivial. Furthermore, we also did not expect the estimators to make predictions for students who join the activity after the specific time of the prediction. These two factors explain why the number of student varies depending on the time in the activity (25%, 50%, or 75%), as can be seen in Tables 2 and 3 in Section 5.

We compare the four previously described estimators: CPR, APR, LPR, and WPR. For LPR we used several values of the parameter κ. Again, the number κ corresponds to the number of most recent log messages that are considered in evaluating the progress rate. We tested all values of κ from two to four, which was decided based on our average number of subtasks in an activity because values larger than four would become equivalent to the CPR model. The LPR models with the different parameters are referred to as LPR2, LPR3, and LPR4 in Table 2 in Section 5.

For each activity and each time (25%, 50%, and 75%), we performed a repeated measure ANOVA to compare the prediction error of our estimators. When the repeated measure ANOVA was significant, we additionally performed a post hoc analysis using paired t-tests comparing the CPR model to each of the other models in terms of the RMSE error for each student. At each time of prediction we also performed the same statistical tests on the data from the 12 activities concatenated to give an idea of the overall performance of the estimators. However, more variance in the error measure is due to the differences between the activities rather than the differences between the predictions, as can be seen from Table 2 in Section 5. Finally, we also compare the estimates of the weight profile for different values of the estimator parameter.

So far, we have only explained how several estimators can be implemented to predict the progress of individual students. Some of these estimators use data from the whole class. Another interesting and possibly useful task is to predict the average progress of the class. This prediction allows us to build dashboards displaying aggregated information about the classroom, such as the one in Figure 5. This prediction could be made directly, but in our situation, because we developed several estimators to predict individual students' progress, another approach is available. In the following section, we evaluate how averaging individual predictions can be a good estimate of the future average progress of the classroom. These estimates are particularly useful for communicating information to teachers in large classrooms.

5. Results

The results we describe in this section are threefold. First, we analyze the optimal parameter for the weight estimation algorithm to balance the estimate between fast and slow students. Second, we compare the estimators on individual predictions. Finally, we compare the estimators for prediction of the average progress of the whole class. Section 5.1 contains our results on the dynamic estimation of the weight profiles. Section 5.2 compares the four estimators we described in Section 4. We report results of statistical tests for predictions made at different times during the activity, and we compare the different estimators and analyze the prediction errors. The results of all of the comparisons are displayed in Table 2. For the aggregated data, we report on the fit of the data when considering aggregated progress and completion rate over the whole class in contrast to individual students. The average error rate of the estimators is shown in Table 3. Even though our work has been focused on individual students, the error of the aggregated predictions is a good indicator of the usefulness of our estimators in the classroom.

### 5.1 Evaluation of Dynamic Estimations of Weight Profiles

As mentioned in Section 4.2.4, the estimate of the weight profile used by the WPR model can be controlled by a parameter p that balances how much of the data of the first students who finish the activity influences the estimate of the subtask weights. We compared the quality of the estimate of the weight profile over time during an activity for different values of p. The main difficulty in estimating the weight profile comes from differences between the first students completing the activity and the rest of the students because the first estimates of the weight profile will be based on the data of the first students who complete the activity. Figure 9 shows the quality of the estimates over time for all 12 activities, depending on the choice of p. We can see that using too high a value for p leads to over-fitting to the behaviour of the fastest students, which is not completely similar to the rest of the class. This is shown in particular for the activities Histogram (a, b, and c) and Genealogy (2018). Otherwise, we see that higher values of the parameter usually lead to a faster estimation of the weight profile. For example, this can be seen for the activities Dot Plot (a and b) and Stroop (2018). Based on the overall performance for the different parameters, we decided to use $p = 2$ in our predictions.

5.2 Evaluation of Individual Predictions

The results displayed in Table 2 show that for the predictions at 25% of the time, the repeated measure ANOVA shows a significant difference (p-value < 0.05) for 8 of the 12 activities (A1, A2, A4, A6, A7, A8, A11, A12). Additionally, the repeated measure ANOVA shows a significant difference between the estimators when aggregating the error measure from all 12 activities. The post hoc analysis using a paired t-test comparison between each estimator and CPR shows that APR, LPR3, LPR4, and WPR are all significantly different from CPR overall (p-value < 0.01), with lower RMSEs. We note that LPR3, LPR4, and WPR only improved the error rate by a few percent, while APR reduced the error by nearly 30%. Post hoc analysis on each of the eight activities shows that APR was significantly different from CPR, with APR performing better. Also, all of the models were significantly different from CPR for A8, again with lower RMSEs than CPR. Additionally, we also observe that LPR2 was significantly different for four activities from CPR, with a better performance for A8 and A12 and a worse performance for A1 and A6. Finally, the results show that WPR was significantly different from CPR for six activities, four better and two worse.

At the time of 50% of the activities, the repeated measure ANOVA is significant overall and for six of the individual activities (A3, A6, A7, A8, A9, and A11, p-value < 0.05). For the overall measure, the estimators LPR3, LPR4, and WPR
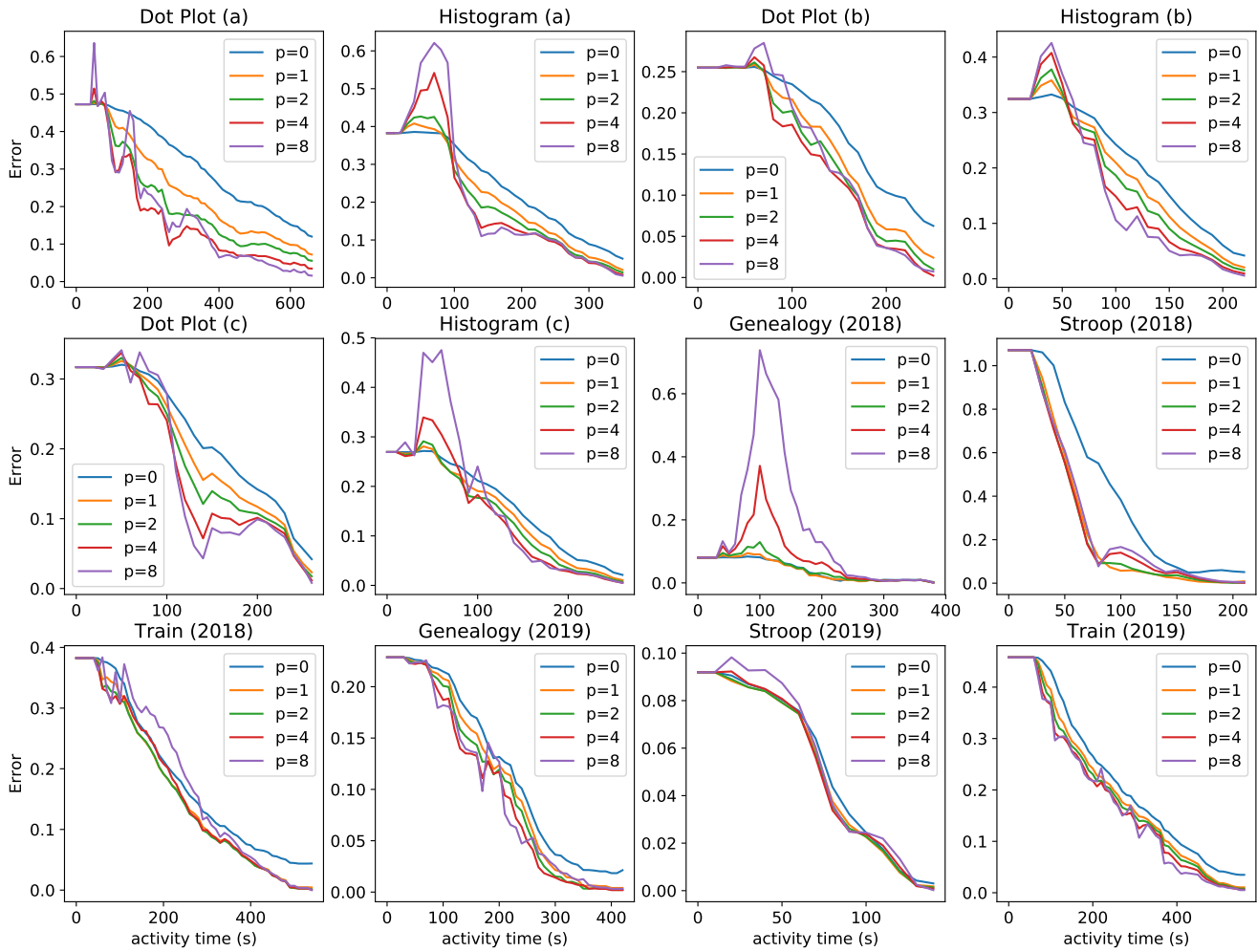
Figure 9. Evolution of the error rate for estimating the weight profile over time for a range of parameters for all 12 activities. The parameter p controls the importance given to faster students in the estimates of the weight profile.

differ significantly from the CPR estimator (using paired t-tests, p-value < 0.01), with the CPR estimator having a lower performance. WPR has the lowest average error of all six estimators at this time for the overall measure. As with the 25% time, we again observe that the estimators APR, LPR2, LPR3, LPR4, and WPR perform significantly differently than CPR for predicting the progress on A8 (p-value < 0.01), again with CPR performing worse. LPR2 also has a significantly different error rate on activities A7, A3, and A6, with a lower error rate for A7 and a higher rate for A3 and A6 than for CPR. WPR has a significantly different error rate than CPR on A11, with WPR having a lower error rate, but APR has a significantly different error rate than CPR on A3, with the error rate being higher for APR.

At the time of 75% of the activities, the repeated measure ANOVA is significant on the set of all the activities and also for activities A1, A2, A3, A9, and A10. The paired t-test over all the activities shows that APR performed significantly differently than CPR on average, with APR performing worse, and APR is the estimator with the highest average error rate. LPR2 and LPR4 also performed significantly differently than CPR, with LPR2 performing worse, while LPR4 performed better. The post hoc analysis shows that the APR estimator has higher prediction errors than the CPR estimator, with this difference being significant (p-value < 0.05) for activities A1, A2, and A3. LPR2 and LPR3 also have significantly different prediction errors than CPR on activity A1, with the CPR errors being lower. The t-tests comparing CPR to the other models for activities A9 and A10 were not significant.

For our analysis, we compared the baseline CPR model to our APR model, which uses the average progress of the whole class to estimate the progress rate for students with few data points. We found that 25% of the way through the activity, the APR model outperforms the CPR model on most of the activities, but the opposite is true after 75% of the time has passed, in

Table 2. Mean Prediction Error for Predictions at 25%, 50%, and 75% of the Time for Each Activity and Averaged over All the Activities

		25% time						
		CPR	APR	LPR2	LPR3	LPR4	WPR	*n*
All*		0.273	**0.197****	0.271	**0.270****	**0.271****	**0.267****	1395
A1*	Dot Plot (a)	0.322	**0.272***	0.331**	0.323	0.322	0.331**	190
A2*	Histogram (a)	0.355	**0.210****	0.357	0.356	0.355	**0.352****	214
A3	Dot Plot (b)	0.281	0.277	0.285	0.281	0.281	0.282	129
A4*	Histogram (b)	0.313	**0.195****	0.317	0.314	0.313	0.310	145
A5	Dot Plot (c)	0.232	0.239	0.233	0.232	0.232	0.231	150
A6*	Histogram (c)	0.270	**0.146****	0.285**	0.273**	0.270	0.268	165
A7*	Genealogy (2018)	0.151	**0.071****	0.155	0.152	0.150	0.169**	66
A8*	Stroop (2018)	0.446	**0.229****	**0.344****	**0.363****	**0.384****	**0.353****	54
A9	Train (2018)	0.196	0.160	0.168	0.187	0.196	0.176	59
A10	Genealogy (2019)	0.142	0.134	0.135	0.136	0.142	0.131	66
A11*	Stroop (2019)	0.105	**0.058****	0.112	0.107	0.104	**0.102***	75
A12*	Train (2019)	0.244	**0.176****	**0.219***	0.247	0.246	**0.204****	82
		50% time						
		CPR	APR	LPR2	LPR3	LPR4	WPR	*n*
All*		0.136	0.131	0.139	**0.132****	**0.131****	**0.128****	1307
A1	Dot Plot (a)	0.208	0.190	0.217	0.213	0.209	0.211	211
A2	Histogram (a)	0.158	0.156	0.179	0.168	0.159	0.156	177
A3*	Dot Plot (b)	0.147	0.189*	0.167**	0.148	0.147	0.147	115
A4	Histogram (b)	0.155	0.136	0.163	0.152	0.156	0.148	118
A5	Dot Plot (c)	0.151	0.159	0.158	0.150	0.151	0.147	131
A6*	Histogram (c)	0.102	0.104	0.118**	0.106	0.102	0.099	136
A7*	Genealogy (2018)	0.031	0.028	**0.022***	0.029	0.029	0.033	66
A8*	Stroop (2018)	0.239	**0.141****	**0.146****	**0.142****	**0.174****	**0.142****	67
A9*	Train (2018)	0.074	0.094	0.065	0.064	0.066	0.063	60
A10	Genealogy (2019)	0.065	0.070	0.053	0.054	0.059	0.056	65
A11*	Stroop (2019)	0.044	0.027	0.040	0.042	0.041	**0.036****	77
A12	Train (2019)	0.083	0.091	0.087	0.081	0.070	0.077	84
		75% time						
		CPR	APR	LPR2	LPR3	LPR4	WPR	*n*
All*		0.103	0.123**	0.110*	0.104	**0.101***	0.103	646
A1*	Dot Plot (a)	0.131	0.158*	0.149**	0.142*	0.134	0.138	157
A2*	Histogram (a)	0.110	0.142*	0.114	0.108	0.106	0.109	92
A3*	Dot Plot (b)	0.137	0.184*	0.150	0.135	0.137	0.141	60
A4	Histogram (b)	0.108	0.123	0.107	0.104	0.103	0.097	51
A5	Dot Plot (c)	0.120	0.122	0.134	0.124	0.121	0.127	79
A6	Histogram (c)	0.082	0.090	0.082	0.084	0.083	0.083	59
A7	Genealogy (2018)	0.018	0.016	0.015	0.017	0.016	0.019	17
A8	Stroop (2018)	0.137	0.122	0.089	0.082	0.100	0.059	16
A9*	Train (2018)	0.042	0.076	0.044	0.038	0.038	0.050	28
A10*	Genealogy (2019)	0.053	0.079	0.047	0.051	0.051	0.048	26
A11	Stroop (2019)	0.028	0.026	0.024	0.018	0.013	0.018	16
A12	Train (2019)	0.046	0.059	0.056	0.056	0.044	0.049	45

In the first column, * shows activities for which significant differences between the models were found using a repeated measure ANOVA (*p*-value < 0.05). In the fourth to eighth columns, * and ** show the significance of a paired t-test between the particular estimator and the CPR estimator (*: *p*-value < 0.05, **: *p*-value < 0.01).

which case the APR model performs worse on three of the activities. This result indicates that at the beginning of the activity, the APR is a good predictor for the progress rate of other students, while it is not a good predictor for students who start later in the activity. Students who start the activity after 50% or 75% of the time has already passed are likely to be very different from students who began the activity right away. Thus, the APR for other students of the class is not a good predictor for their progress rate.

Second, we compared models that addressed the variability in subtask time, the LPR and WPR models, to the baseline CPR model. The LPR model is based on the assumption that the rate at which students make progress on an activity will evolve over time. From this assumption, the model we proposed computes progress rate based only on recent logs. For this model, we compared three different variations that used a progressively larger number of previous log messages for each student. One activity that performed consistently better for the LPR models was Stroop (2018), which performed better for all three types of LPR models for the 25% and 50% time points. The Stroop (2018) activity was uniquely different from the other activities, where the main point of the activity was to complete each task as quickly as possible. However, after the teacher started the activity, all of the students took time to read the instructions, which added time to their first subtask and made their starting slope quite different from that of the rest of the activity. This part of the activity was changed the following year, so we do not observe a similar discrepancy for Stroop (2019) (see Figure 10). For Stroop (2018), where there was a predictable change in time between subtasks, assuming that the progress rate could vary over time was a good choice. This result also indicates that the model performance depends on the type of activity or on the classroom. Our results do not clearly indicate which parameter for the LPR model performed best. It seems to be highly dependent on the activity and the students. The choice of the parameter is a trade-off between the noise of the measurements and the flexibility to adapt quickly to changes in progress rate.
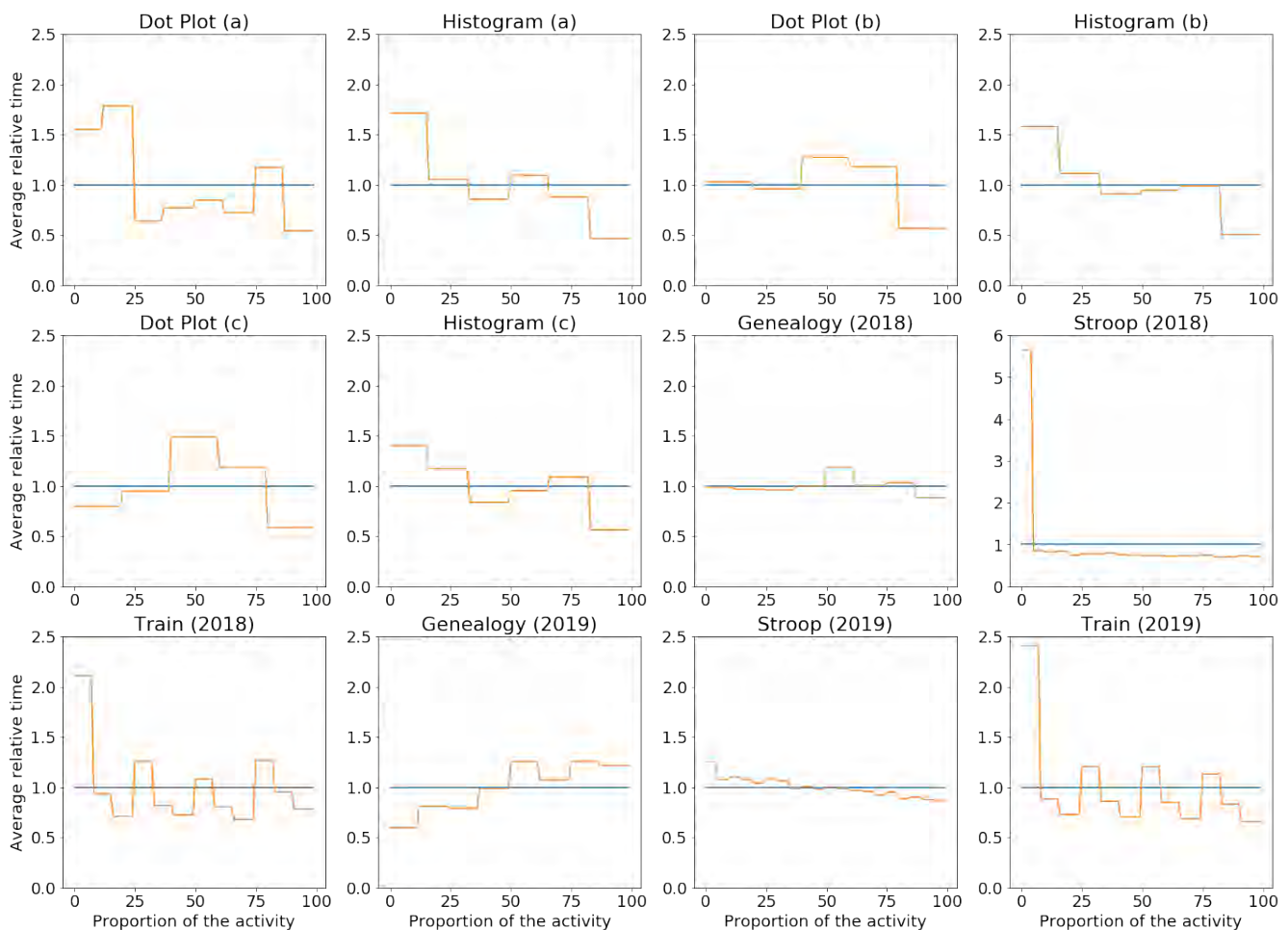
Figure 10. Weight profile of the 12 activities. Because of a large divergence, the vertical scale of the weight profile for Stroop (2018) is different from that for the other activities.

Moreover, near the end of the activity at the 75% time point, there are fewer statistically significant results for the LPR models even though we often see the same trends as at earlier time points. This is most likely due to many students having finished the activity by that time point and thus being excluded from the analysis, lowering our statistical power. For example, Stroop (2018) goes from having $n = 67$ at the 50% time point to $n = 16$ at the 75% time point, which is a 76% reduction in the number of students. It may also mean that any information added from the different model variations does not apply to the last students finishing, who may perform differently than the other students.

Finally, the WPR model showed similar results to the LPR model. Like the LPR model, WPR is an adaptation of the CPR model based on the assumption that the progress proportion varies between subtasks. This model becomes very interesting in situations where the progress proportion of each subtask varies strongly from the chosen assignment. Unlike the LPR models, the WPR model takes into account varied change, where there may not be a pattern in the weights. For example, if a quiz gets steadily harder as it progresses, then there is a uniform change. However, if the difficult questions are randomly placed throughout the quiz, then the change is varied and performance on a previous question cannot be used to predict current performance. Although the weight data is still sparse at the 25% time point because few students, if any, have completed the task, it is still beneficial for predicting 9 of the 12 activities (with statistical significance for 4 activities). The WPR model may not have had a better prediction on all activities for several reasons. The first is that most of the prediction error comes from individual noise, which cannot be more accurately predicted from aggregated information about the whole class. Second, the weight profile for 11 of the activities is not as divergent from the original equal-weight assignment as for the Stroop (2018) activity (see Figure 10), which means that the benefits of the WPR model are not as large. Additionally, the divergence in the weight profile of Stroop (2018) is exactly at the beginning of the activity, so the WPR estimator can quickly estimate it well, but other estimators are actually more affected by this divergence. Finally, for some activities, the faster students are not similar to the rest of the class. This leads to lower prediction performance for the estimators that use aggregated data of the class (APR and WPR) because the APR and the weight model estimation will be biased toward the behaviour of the faster students.

5.3 Evaluation of Aggregated Predictions

In Section 5.2, we compared the quality of our proposed estimators for predicting individual progress. The error rate aggregated over the whole class also matters to justify the usefulness of the method for displaying class-level data. Table 3 shows the same error measurement as reported in the previous sections using equation (9), with the difference that the progress predictions are first averaged over the whole class. This leads, as expected, to lower error rates for two reasons. First, the cases where the future progress of a student is overestimated are now being compensated for by cases where the future progress has been underestimated, thus reducing the error in the aggregated prediction. The prediction error will stay high for an estimator if it mostly overestimates or mostly underestimates the future progress in a systematic way. For example, this is what happens for the APR model, which even though it has lower prediction error for each individual student at the 25% time point, it has a higher aggregated prediction error for the activities Genealogy and Histogram (a), (b), and (c). Second, lower errors occur in the aggregated data because students with progress value one, for whom predictions are trivial, were not included in the predictions in the individual case but had to be included in the aggregated case. In this case, because the students have already finished the activity, they do not have any error.

From Table 3, we see that the aggregated prediction errors range from 0.001 for several activities and models at the 75% time point to 0.404 for the CPR model for the Stroop (2018) activity at the 25% time point. The results are satisfactory to be used by teachers. For example, for the 75% time mark, which is when the teacher is closer to deciding to transition to the next activity, for the baseline estimator the error is the highest for activities A8 and A10 at 1.7% and 1.8%. An error of this scale means that the prediction is off by usually less than the progress of two or three students for a classroom of about 150 students. For these two activities, we see that the WPR model has an even lower prediction error, with a decrease of 39% for A10 and 76% for A8.

6. Discussion

In this paper, we aimed to address the problem of dynamically predicting students' future progress during activities in the classroom. We presented a basic model for making predictions in real time and proposed three variations of this model that accounted for different challenges of predicting the overall progress of a class through an activity. For each of our model variations, we found that we could better predict students' progress compared to the baseline model for a subset of our datasets. By analyzing the differences in performance between these datasets and the time of the predictions, we can form a better understanding of completion patterns for different activity types, as well as of how students engage with these activities.

When we evaluate the patterns of the model successes as a whole, the first one that emerges is that aggregated information about the whole class can be used to improve the performance of progress predictions for individual students. It is important to

Table 3. Mean Prediction Error for Aggregated Predictions at 25%, 50%, and 75% of the Time for Each Activity

		25% time						
		CPR	APR	LPR2	LPR3	LPR4	WPR	*n*
A1	Dot Plot (a)	0.140	0.213	0.146	0.141	0.140	0.126	192
A2	Histogram (a)	0.249	0.143	0.248	0.249	0.249	0.244	216
A3	Dot Plot (b)	0.152	0.242	0.153	0.152	0.152	0.151	129
A4	Histogram (b)	0.198	0.127	0.198	0.198	0.198	0.184	145
A5	Dot Plot (c)	0.096	0.210	0.093	0.096	0.096	0.094	150
A6	Histogram (c)	0.160	0.096	0.158	0.160	0.160	0.139	166
A7	Genealogy (2018)	0.085	0.047	0.086	0.086	0.085	0.086	67
A8	Stroop (2018)	0.404	0.172	0.319	0.343	0.363	0.308	54
A9	Train (2018)	0.125	0.051	0.065	0.097	0.124	0.100	59
A10	Genealogy (2019)	0.058	0.108	0.047	0.052	0.057	0.046	66
A11	Stroop (2019)	0.068	0.014	0.066	0.066	0.065	0.061	75
A12	Train (2019)	0.196	0.111	0.113	0.167	0.193	0.139	82
		50% time						
		CPR	APR	LPR2	LPR3	LPR4	WPR	*n*
A1	Dot Plot (a)	0.061	0.091	0.059	0.058	0.061	0.039	230
A2	Histogram (a)	0.032	0.053	0.037	0.034	0.032	0.026	224
A3	Dot Plot (b)	0.023	0.122	0.028	0.022	0.023	0.027	133
A4	Histogram (b)	0.042	0.024	0.037	0.040	0.042	0.026	146
A5	Dot Plot (c)	0.065	0.108	0.055	0.061	0.065	0.049	154
A6	Histogram (c)	0.008	0.027	0.036	0.021	0.011	0.023	169
A7	Genealogy (2018)	0.008	0.009	0.005	0.005	0.006	0.009	76
A8	Stroop (2018)	0.133	0.065	0.077	0.078	0.101	0.072	106
A9	Train (2018)	0.026	0.013	0.011	0.010	0.011	0.006	62
A10	Genealogy (2019)	0.030	0.047	0.013	0.018	0.022	0.012	68
A11	Stroop (2019)	0.029	0.011	0.025	0.028	0.027	0.023	84
A12	Train (2019)	0.034	0.021	0.012	0.011	0.006	0.011	84
		75% time						
		CPR	APR	LPR2	LPR3	LPR4	WPR	*n*
A1	Dot Plot (a)	0.014	0.051	0.038	0.027	0.020	0.039	240
A2	Histogram (a)	0.002	0.026	0.009	0.007	0.004	0.015	226
A3	Dot Plot (b)	0.007	0.053	0.015	0.006	0.007	0.009	133
A4	Histogram (b)	0.015	0.005	0.005	0.010	0.012	0.004	146
A5	Dot Plot (c)	0.013	0.021	0.021	0.014	0.013	0.023	157
A6	Histogram (c)	0.004	0.011	0.012	0.008	0.005	0.010	170
A7	Genealogy (2018)	0.002	0.002	0.001	0.002	0.002	0.002	78
A8	Stroop (2018)	0.017	0.007	0.010	0.010	0.013	0.004	116
A9	Train (2018)	0.002	0.019	0.007	0.004	0.004	0.004	64
A10	Genealogy (2019)	0.018	0.028	0.015	0.015	0.016	0.011	69
A11	Stroop (2019)	0.004	0.003	0.001	0.001	0.001	0.002	84
A12	Train (2019)	0.005	0.012	0.013	0.013	0.006	0.010	84

note that the models using whole-class data (APR) and automated weighting of progress assignment to subtask (WPR) each increased the prediction performance for over half the activities at the beginning of an activity. Still, there are cases remaining where our results show that students' individual variability is more important than the average behaviour of the class. Thus, the prediction did not improve compared to the CPR model. However, it was uncommon for a model that used aggregated performance to perform significantly worse than the baseline model, indicating that these models could be a good starting point, independent of the activity type. Additionally, our results show us when during the in-class activity the different estimators perform best, which indicates how we could design a combined estimator by appropriately selecting the different modelling approaches that we proposed.

A second pattern that emerges is that there is less of a difference between models as time increases. As mentioned earlier, this difference could be explained by the decrease in the number of students who have not finished the activity, which leads to a decrease in statistical power for our analysis. However, it may also be due to the types of students who complete the activity later. There can be multiple reasons for students to take longer on an activity: they started later, they solve subtasks at a slower pace, or they have off-task behaviour. With the later start and the slower pace, we could expect a relationship with the way the students perform and how earlier students perform. In the case of starting later, it may just be a shift in time, while with the slower pace, it may just be extending the time frame. On the other hand, when the difference is not a simple transformation, as would be the case with off-task behaviour, the enhanced models may not improve the prediction. In future research, it may be beneficial to assess the relationship between earlier and later students' progress to investigate where these differences may arise.

Finally, we see that within any given time point, there is no clear pattern when assessing the results within an activity or class. For example, the Stroop (2018) activity performed better for all three of the LPR models, while the Stroop (2019) model did not. These results indicate that there was not a common pattern of time taken on activities across these two classes. This can be confirmed by looking at the progress curves in Figure 8, in which we can see that although the curves are similar, there is more of a lag at the beginning of the activity in Stroop (2018). This results in differences in the weight profiles of these two activities (see Figure 10). This lag illustrates how external classroom events such as an orchestration action of the teacher can affect the timing of an activity. This lag is not present in any of the other activities for the same students, so it is not a property of the students but may have been due to something that happened in the class that day. These results illustrate the importance of the timing model being able to take into account unexpected events that may happen in the classroom in real time. The proposed models account for these real-time events through different methods. How to combine our different techniques to obtain the optimal predictions will require future research.

To test our models, we used 12 datasets to see how they performed in a range of situations. Across the 12 datasets, there were 5 different activities and 5 different sets of students. Having an overlap of activities and students allowed us to analyze patterns that may have emerged associated with activity or student traits. As mentioned above, we found many discrepancies between the same activities and the same students when applied to different situations. These findings provide more strength to our approach, in which we do not rely on a trained model, which then may overgeneralize, and do not make any assumptions at the beginning of the activity. We are then able to apply our models to any new activity, student group, or situation without affecting the accuracy of our predictions. However, all of our datasets were collected from a higher-education setting. In this case, our results may have limited generalizability to K–12 education if students follow drastically different progression patterns.

In practice, over all the estimators we presented, the error rates for the prediction are quite satisfactory. For example, looking at the 75% time of the activities, the prediction errors in terms of RMSE are nearly never above 0.15 and are on average 0.103 for the CPR model. An error means that the prediction estimator on average wrongly predicts one more or one less subtask completed for an activity with ten subtasks. Given the unpredictability of student behaviour in general, we found this to be a very satisfactory result. Figure 3 illustrates well the very high variance in the time students take to complete subtasks of an activity, and Section 3 details the difficulty of predicting students' progress.

In addition to the error rates for the individual predictions, we also evaluated the error rate for predicting the average progress of the classroom. In this case, the errors are much lower. For example, at the time 75% (which is the most relevant time because it is closer to the moment that the teacher actually decide to transition to the next task), the error rate of the averaged prediction is always below 0.02. This could be the result of the estimator making perfectly accurate predictions for 98 students of a class of 100 students and completely wrong predictions for the remaining 2 students. We again find this range of error rate for a tool aimed at supporting teachers' decisions satisfactory.

6.1 Future Work and Limitations

This paper provides an initial step toward the prediction of activity completion. Several future steps can be taken to build upon this work. First, the current models do not provide insight into the reasoning behind the different completion patterns. For

example, if students increase their speed per task during an activity, it is not clear what mechanism is leading to this increase. The students may be becoming more familiar with the task, the task difficulty could be decreasing, or the students could be no longer engaged with the activity and quickly clicking through, among other reasons. Because these models were not designed to provide insight into the process but rather provide an accurate prediction for future progress and completion, future work is needed. By understanding the underlying mechanisms affecting students' progress, teachers can not only transition activities at optimal times but also adapt the activity to students' needs.

Additionally, the estimators described in this paper are based solely on discrete measurements of student progress during activities that can be naturally divided into a sequence of subtasks. The estimators we presented are limited to this particular case and would not directly generalize to other contexts. This limitation could be extended in three ways. First, progress could be measured continuously and be generalized to activities that are not naturally divisible into subtasks, such as writing essays or brainstorming. A second improvement worth investigating is to extend our use of current whole-class data to improve the prediction to study how data collected previously from the same class or data collected previously from the same activity can be used in the predictions (Boyer & Veeramachaneni, 2015). The estimators described in this paper are limited to using only data from the current activity and do not use other sources of data. This has the advantage of making the estimators easily usable without preparation but misses opportunities to increase the quality of predictions. Learning from other sources of data would also allow the use of more data-intensive algorithms for prediction, such as deep neural networks, because they will use more data than the real-time estimators we describe in this paper. Finally, a third extension is to use other sources of data not directly related to students' progress. For example, evaluating in real time the state of the student (idle/active) during the activity could inform the predictions to know when the student is not likely to make progress. This could be done using eye tracking or video recording (Hutt et al., 2017).

In this work, we focus on making a separate progress prediction for each student, but because the goal of these predictive models is to support teacher orchestration, the individual student data needs to be aggregated to provide the teacher with an overview of the progress and completion for the class as a whole. A simple average of students' progress over time is not sufficient information. For example, when the average progress is at 50%, this could indicate that all students have a progress of 50% or that half of the students have completed the activity while the other half have not started (or anywhere in between). Because an average of the progress of the class can be explained by several states of the class, in current work using these estimators in the classroom, we display both the average progress and the rate of completion. In future research, we will investigate how to display this information efficiently to teachers and the impact that the progress prediction has on teacher orchestration.

Additionally, it may be beneficial to go beyond the prediction of student progress to predict the optimal time to switch and what the trade-offs may be for taking more or less time (Dillenbourg et al., 2016). As part of these trade-offs, it is important to consider not only how much further in the activity a student will progress if given more time but what the impact on their learning may be. To investigate this impact, we need to explore the differences in learning gains that occur between different levels of completion of an activity. Using this relationship, we could build a system that automatically estimates optimal transition times between activities and could provide more meaningful information to teachers about how transitions will affect student learning.

Because our models allow us to predict the progress of students at a given point in time or to predict the time at which students will reach a given point of progress, they can be used in the classroom to support activity transitions. In the classroom, it can be difficult for teachers to monitor all students at the same time, particularly when students are working on individual or small-group tasks. Using teacher dashboards, teachers can easily monitor and process the data and actions of students in a consolidated form (Klerkx, Verbert, & Duval, 2017; Schwendimann et al., 2016). Our models can be used as the back end to the dashboard to provide this information to teachers. How this information is provided is still an open question, with research only beginning to investigate the different levels of orchestration support that can be provided (van Leeuwen, Rummel, & van Gog, 2019). In the case of timing, the information could be shown as a graph, as in Figure 5, or as a set of recommendations based on the prediction. How best to use these models to provide teachers with timely information is a direction for future research.

7. Conclusion

In this paper, we present estimators to predict the future progress of students as they work on individual activities in the classroom. Our techniques principally address the challenges arising from individual differences in students. Furthermore, two of our estimators also use data collected from the whole class to improve individual predictions. By predicting the future progress of students, we address a practical issue in the classroom of teachers not being able to fully monitor their students and not knowing when to transition to the next activity. Specifically, in large lectures, the teacher cannot move around the classroom, and common indicators of completeness, such as the noise level in the classroom, can often be misleading. By

having an accurate prediction of when students may complete an activity, teachers can make informed orchestration moves, such as giving time warnings instead of suddenly switching to a new activity. Teachers being able to make more informed decisions about activity time can prevent students who have already completed the activity from waiting too long as well as preventing those who are still working on a task from getting cut off too early.

Additionally, by analyzing the strengths of different variations of our estimators, we gained further insight into students' behaviour during classroom activities. Our work showed how different strategies for predicting students' progress would perform based on individual differences from and similarity to the average of the whole class. Overall, our work contributes to the field of learning analytics by introducing a novel approach to evaluating students' progress and progress rates during activities in classrooms. Moreover, our work has direct implications for supporting teachers' orchestration by addressing their need for time management support.

Acknowledgments

Declaration of Conflicting Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

References

Alavi, H. S., & Dillenbourg, P. (2012). An ambient awareness tool for supporting supervised collaborative problem solving. *IEEE Transactions on Learning Technologies*, *5*(3), 264–274. https://dx.doi.org/10.1109/TLT.2012.7

Baker, R., Corbett, A. T., & Koedinger, K. R. (2004). Detecting student misuse of intelligent tutoring systems. In *7th International Conference on Intelligent Tutoring Systems* (ITS2004), 30 August–3 September 2004, Maceió, Alagoas, Brazil (pp. 531–540). Springer. https://dx.doi.org/10.1007/978-3-540-30139-4_50

Baker, R., & Siemens, G. (2014). Educational data mining and learning analytics. In R. K. Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences.* Cambridge, UK: Cambridge University Press. https://dx.doi.org/10.1017/CBO9781139519526.016

Beck, J. E. (2005). Engagement tracing: Using response times to model student disengagement. In C. Looi, G. I. McCalla, B. Bredeweg, & J. Breuker (Eds.), *Artificial Intelligence in Education—Supporting Learning through Intelligent and Socially Informed Technology, Proceedings of the 12th International Conference on Artificial Intelligence in Education,* AIED 2005, 18–22 July 2005, Amsterdam, Netherlands (Vol. 125, pp. 88–95). IOS Press. Retrieved from http://www.booksonline.iospress.nl/Content/View.aspx?piid=1296

Boroujeni, M. S., Sharma, K., Kidziński, Ł., Lucignano, L., & Dillenbourg, P. (2016). How to quantify student's regularity? In *Proceedings of the 11th European Conference on Technology Enhanced Learning* (EC-TEL 2016), 13–16 September 2016, Lyon, France (pp. 277–291). Springer. https://dx.doi.org/10.1007/978-3-319-45153-4_21

Boyer, S., & Veeramachaneni, K. (2015). Transfer learning for predictive models in massive open online courses. In *Proceedings of the 17th International Conference on Artificial Intelligence in Education* (AIED 2015), 22–26 June 2015, Madrid, Spain (pp. 54–63). Springer. https://dx.doi.org/10.1007/978-3-319-19773-9_6

Campbell, S., & Skinner, C. H. (2004). Combining explicit timing with an interdependent group contingency program to decrease transition times: An investigation of the timely transitions game. *Journal of Applied School Psychology*, *20*(2), 11–27. https://dx.doi.org/10.1300/J370v20n02_02

Charleer, S., Santos, J. L., Klerkx, J., & Duval, E. (2014). Improving teacher awareness through activity, badge and content visualizations. In *Proceedings of the 13th International Conference on Web-Based Learning* (ICWL 2014), 14–17 August 2014, Tallinn, Estonia (pp. 143–152). Springer. https://dx.doi.org/10.1007/978-3-319-13296-9_16

Clark, D., & Linn, M. C. (2003). Designing for knowledge integration: The impact of instructional time. *The Journal of the Learning Sciences*, *12*(4), 451–493. https://dx.doi.org/10.1207/S15327809JLS1204_1

Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, *4*(4), 253–278. https://dx.doi.org/10.1007/BF01099821

41

Crossley, S., Paquette, L., Dascalu, M., McNamara, D. S., & Baker, R. S. (2016). Combining click-stream data with NLP tools to better understand MOOC completion. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (LAK 2016), 25–29 April 2015, Edinburgh, United Kingdom (pp. 6–14). New York, USA: ACM. https://dx.doi.org/10.1145/2883851.2883931

Dekker, G. W., Pechenizkiy, M., & Vleeshouwers, J. M. (2009). Predicting students drop out: A case study. In T. Barnes, M. Desmarais, C. Romero, & S. Ventura (Eds.), *Proceedings of the Second International Conference on Educational Data Mining* (EDM '09), 1–3 July 2009, Cordoba, Spain. Retrieved from http://www.educationaldatamining.org/EDM2009/uploads/proceedings/dekker.pdf

Dillenbourg, P. (2013). Design for classroom orchestration. *Computers & Education*, *69*, 485–492. https://dx.doi.org/10.1016/j.compedu.2013.04.013

Dillenbourg, P. (2015). *Orchestration Graphs*. Lausanne, Switzerland: EPFL Press. Retrieved from https://www.epflpress.org/produit/735/9782940222841/orchestration-graphs

Dillenbourg, P., Li, N., & Kidzinski, L. (2016). *The Complications of the Orchestration Clock* (Tech. Rep.). Portland Press. Retrieved from https://portlandpress.com/DocumentLibrary/Umbrella/Wenner%20Gren/Vol%2088/PPL$_{Wenner_c}h$02.pdf

Dillenbourg, P., Prieto, L. P., & Olsen, J. K. (2018). Classroom orchestration. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann (Eds.), *International Handbook of the Learning Sciences* (pp. 180–190). Routledge. https://dx.doi.org/10.4324/9781315617572

Dillenbourg, P., Zufferey, G., Alavi, H. S., Jermann, P., Do, L. H. S., Bonnard, Q., . . . Kaplan, F. (2011). Classroom orchestration: The third circle of usability. In *Proceedings of the Ninth Annual Conference on Computer-Supported Collaborative Learning* (CSCL2011), 4–8 July 2011, Hong Kong, China (Vol. 1, pp. 510–517). International Society of the Learning Sciences. Retrieved from https://infoscience.epfl.ch/record/168741?ln=en

D'mello, S. K., Craig, S. D., Witherspoon, A., Mcdaniel, B., & Graesser, A. (2008). Automatic detection of learner's affect from conversational cues. *User Modeling and User-Adapted Interaction*, *18*(1-2), 45–80. https://dx.doi.org/10.1007/s11257-007-9037-6

Feng, M., Heffernan, N., & Koedinger, K. (2009). Addressing the assessment challenge with an online system that tutors as it assesses. *User Modeling and User-Adapted Interaction*, *19*(3), 243–266. https://dx.doi.org/10.1007/s11257-009-9063-7

Fiel, J., Lawless, K. A., & Brown, S. W. (2018). Timing matters: Approaches for measuring and visualizing behaviours of timing and spacing of work in self-paced online teacher professional development courses. *Journal of Learning Analytics*, *5*(1), 25–40. https://dx.doi.org/10.18608/jla.2018.51.3

Fraser, B. J., Walberg, H. J., Welch, W. W., & Hattie, J. A. (1987). Syntheses of educational productivity research. *International Journal of Educational Research*, *11*(2), 147–252. https://dx.doi.org/10.1016/0883-0355(87)90035-8

Gettinger, M., & Seibert, J. K. (2002). Best practices in increasing academic learning time. *Best Practices in School Psychology*, *IV*, 773–787. Retrieved from https://psycnet.apa.org/record/2006-03715-049

Håklev, S., Faucon, L., Hadzilacos, T., & Dillenbourg, P. (2017). Orchestration graphs: Enabling rich social pedagogical scenarios in MOOCs. In *Proceedings of the Fourth ACM Conference on Learning@Scale* (L@S 2017), 20–21 April 2017, Cambridge, Massachusetts, USA (pp. 261–264). New York, USA: ACM. https://dx.doi.org/10.1145/3051457.3054000

Hine, J. F., Ardoin, S. P., & Foster, T. E. (2015). Decreasing transition times in elementary school classrooms: Using computer-assisted instruction to automate intervention components. *Journal of Applied Behavior Analysis*, *48*(3), 495–510. https://dx.doi.org/10.1002/jaba.233

Holstein, K., Hong, G., Tegene, M., McLaren, B. M., & Aleven, V. (2018). The classroom as a dashboard: Co-designing wearable cognitive augmentation for K–12 teachers. In *Proceedings of the Eighth International Conference on Learning Analytics and Knowledge* (LAK 2018), 5–9 March 2018, Sydney, Australia (pp. 79–88). New York, USA: ACM. https://dx.doi.org/10.1145/3170358.3170377

Hutt, S., Mills, C., Bosch, N., Krasich, K., Brockmole, J., & D'Mello, S. (2017). Out of the fr-eye-ing pan: Towards gaze-based models of attention during learning with technology in the classroom. In *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization* (UMAP 2017), 9–12 July 2017, Bratislava, Slovakia (pp. 94–103). New York, USA: ACM. https://dx.doi.org/10.1145/3079628.3079669

Jovanović, J., Gašević, D., Dawson, S., Pardo, A., & Mirriahi, N. (2017). Learning analytics to unveil learning strategies in a flipped classroom. *The Internet and Higher Education*, *33*(4), 74–85. Retrieved from https://psycnet.apa.org/doi/10.1016/j.iheduc.2017.02.001

Karweit, N., & Slavin, R. E. (1981). Measurement and modeling choices in studies of time and learning. *American Educational Research Journal*, *18*(2), 157–171. https://dx.doi.org/10.3102F00028312018002157

Klerkx, J., Verbert, K., & Duval, E. (2017). Learning analytics dashboards. *Handbook of Learning Analytics*, *1*, 143–150. https://dx.doi.org/10.18608/hla17.012

Kollar, I., & Fischer, F. (2013). Orchestration is nothing without conducting—But arranging ties the two together!: A response to Dillenbourg (2011). *Computers & Education*, *69*, 507–509. https://dx.doi.org/10.1016/j.compedu.2013.04.008

Loboda, T. D., Guerra, J., Hosseini, R., & Brusilovsky, P. (2014). Mastery grids: An open source social educational progress visualization. In *Proceedings of the Ninth European Conference on Technology Enhanced Learning* (EC-TEL 2014), 16–19 September 2014, Graz, Austria (pp. 235–248). Springer. https://dx.doi.org/10.1007/978-3-319-11200-8_8

Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science*, *5*(4), 333–369. https://dx.doi.org/10.1207/s15516709cog0504_2

Martinez-Maldonado, R., Clayphan, A., & Kay, J. (2015). Deploying and visualising teacher's scripts of small group activities in a multi-surface classroom ecology: A study in-the-wild. *Computer Supported Cooperative Work (CSCW)*, *24*(2-3), 177–221. https://dx.doi.org/10.1007/s10606-015-9217-6

Ming, N., & Ming, V. (2012). Predicting student outcomes from unstructured data. In *Workshop and Poster Proceedings of the 20th Conference on User Modeling, Adaptation, and Personalization* (UMAP 2012), 16–20 July 2012, Montreal, Canada. Retrieved from http://ceur-ws.org/Vol-872/pale2012_paper_2.pdf

Molenaar, I., & Knoop-van Campen, C. (2017). Teacher dashboards in practice: Usage and impact. In *Proceedings of the 12th European Conference on Technology Enhanced Learning* (EC-TEL 2017), 12–15 September 2017, Tallinn, Estonia (pp. 125–138). Springer. https://dx.doi.org/10.1007/978-3-319-66610-5_10

Pavlik Jr., P. I., Cen, H., & Koedinger, K. R. (2009). Performance factors analysis—A new alternative to knowledge tracing. (online submission) Retrieved from http://pact.cs.cmu.edu/koedinger/pubs/AIED%202009%20final%20Pavlik%20Cen%20Keodinger%20corrected.pdf

Pelánek, R. (2015). Metrics for evaluation of student models. *Journal of Educational Data Mining*, *7*(2), 1–19. Retrieved from https://jedm.educationaldatamining.org/index.php/JEDM/article/download/JEDM087/pdf_12

Phiri, L., Meinel, C., & Suleman, H. (2016). Streamlined orchestration: An orchestration workbench framework for effective teaching. *Computers & Education*, *95*, 231–238. https://dx.doi.org/10.1016/j.compedu.2016.01.011

Prieto, L. P., Dlab, M. H., Gutiérrez, I., Abdulwahed, M., & Balid, W. (2011). Orchestrating technology enhanced learning: A literature review and a conceptual framework. *International Journal of Technology Enhanced Learning*, *3*(6), 583. https://dx.doi.org/10.1504/IJTEL.2011.045449

Rhymer, K. N., Skinner, C. H., Jackson, S., McNeill, S., Smith, T., & Jackson, B. (2002). The 1-minute explicit timing intervention: The influence of mathematics problem difficulty. *Journal of Instructional Psychology*, *29*(4), 305–311. Retrieved from https://psycnet.apa.org/record/2002-11085-011

Schwendimann, B. A., Rodriguez-Triana, M. J., Vozniuk, A., Prieto, L. P., Boroujeni, M. S., Holzer, A., . . . Dillenbourg, P. (2016). Perceiving learning at a glance: A systematic literature review of learning dashboard research. *IEEE Transactions on Learning Technologies*, *10*(1), 30–41. https://dx.doi.org/10.1109/TLT.2016.2599522

Slotta, J. D., Tissenbaum, M., & Lui, M. (2013). Orchestrating of complex inquiry: Three roles for learning analytics in a smart classroom infrastructure. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (LAK 2013), 8–12 April 2013, Leuven, Belgium (pp. 270–274). New York, USA: ACM. https://dx.doi.org/10.1145/2460296.2460352

Song, Y., Wong, L.-H., & Looi, C.-K. (2012). Fostering personalized learning in science inquiry supported by mobile technologies. *Educational Technology Research and Development*, *60*(4), 679–701. https://dx.doi.org/10.1007/s11423-012-9245-6

Taylor, C., Veeramachaneni, K., & O'Reilly, U.-M. (2014). Likely to stop? Predicting stopout in massive open online courses. *arXiv preprint arXiv:1408.3382*.

Van Houten, R., Hill, S., & Parsons, M. (1975). An analysis of a performance feedback system: The effects of timing and feedback, public posting, and praise upon academic performance and peer interaction 1. *Journal of Applied Behavior Analysis*, *8*(4), 449–457. https://dx.doi.org/10.1901/jaba.1975.8-449

Van Houten, R., & Thompson, C. (1976). The effects of explicit timing on math performance. *Journal of Applied Behavior Analysis*, *9*(2), 227. https://dx.doi.org/10.1901%2Fjaba.1976.9-227

van Leeuwen, A., Rummel, N., & van Gog, T. (2019). What information should CSCL teacher dashboards provide to help teachers interpret CSCL situations? *International Journal of Computer-Supported Collaborative Learning*, *14*, 261–289. https://dx.doi.org/10.1007/s11412-019-09299-x

van Leeuwen, A., van Wermeskerken, M., Erkens, G., & Rummel, N. (2017). Measuring teacher sense making strategies of learning analytics: A case study. *Learning: Research and Practice*, *3*(1), 42–58.

https://dx.doi.org/10.1080/23735082.2017.1284252

Vanlehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education*, *16*(3), 227–265. https://dx.doi.org/10.5555/1435351.1435353

Verbert, K., Govaerts, S., Duval, E., Santos, J. L., Assche, F., Parra, G., & Klerkx, J. (2014). Learning dashboards: An overview and future research opportunities. *Personal and Ubiquitous Computing*, *18*(6), 1499–1514. https://dx.doi.org/10.1007/s00779-013-0751-2

Wachtler, J., Khalil, M., Taraghi, B., & Ebner, M. (2016). On using learning analytics to track the activity of interactive MOOC videos. In *Proceedings of the LAK 2016 Workshop on Smart Environments and Analytics in Video-Based Learning,* 26 April 2016, Edinburgh, United Kingdom (pp. 8–17). Retrieved from http://ceur-ws.org/Vol-1579/paper3.pdf

Wang, P., Tchounikine, P., & Quignard, M. (2017). Chao: A framework for the development of orchestration technologies for technology-enhanced learning activities using tablets in classrooms. *International Journal of Technology Enhanced Learning*, *10*(1/2), 1–21. https://dx.doi.org/10.5555/3202163.3202164

Whitehill, J., Mohan, K., Seaton, D., Rosen, Y., & Tingley, D. (2017). Delving deeper into MOOC student dropout prediction. *arXiv preprint arXiv:1702.06404*.

Xu, D., Wang, H., & Wang, M. (2005). A conceptual model of personalized virtual learning environments. *Expert Systems with Applications*, *29*(3), 525–534. https://dx.doi.org/10.1016/j.eswa.2005.04.028

Yarbrough, J. L., Skinner, C. H., Lee, Y. J., & Lemmons, C. (2004). Decreasing transition times in a second grade classroom: Scientific support for the timely transitions game. *Journal of Applied School Psychology*, *20*(2), 85–107. https://dx.doi.org/10.1300/J370v20n02_06