ARTICLE IN PRESS

Learning and Instruction xxx (xxxx) xxxx

Contents lists available at ScienceDirect

Learning and Instruction

journal homepage: www.elsevier.com/locate/learninstruc

What can moment-by-moment learning curves tell about students' self-regulated learning?

Inge Molenaar[a,*], Anne Horvers[a], Ryan S. Baker[b]

[a] Radboud University, Netherlands
[b] University of Pennsylvania, USA

ARTICLE INFO

Keywords:
Primary education
Learning analytics
Adaptive learning technologies
Self-regulated learning

ABSTRACT

Many students in primary education learn arithmetic using adaptive learning technologies (ALTs) on tablets every day. Driven by developments in the emerging field of learning analytics, these technologies adjust problems based on learners' performance. Yet, until now it is largely unclear how students regulate their learning with ALTs. Hence, we explored how learners regulate their effort, accuracy and learning with an ALT using moment-by-moment learning curves. The results indicated that moment-by-moment learning curves did reflect students' accuracy and learning, but no associations with effort were found. Immediate drops were associated with high prior knowledge and suboptimal learning. Immediate peaks were associated with robust learning and pointed to effective student regulation. Close multiple spikes showed moderate learning and lower initial levels of accuracy but, with system support, these students seemed able to regulate their learning. Separated multiple spikes indicated reduced learning and accuracy and potentially signal the inability of students to regulate their learning. In this light, moment-by-moment learning curves seem to be valuable indicators of accuracy regulation during learning with ALTs and could potentially be used in interventions to support SRL with personalized visualizations.

1. Introduction

In the Netherlands alone over 250,000 students in primary education use adaptive learning technologies (ALTs) such as Snappet, Muiswerk, Taalzee/Rekentuin, Got it and PulseOn on a daily basis (Kennisnet, 2014). These technologies, mainly used in Mathematics, Arithmetic, Dutch and English, capture rich data about students' performance during learning (Greller & Drachsler, 2012; Papamitsiou & Economides, 2014). Students' performance data is used to assess their knowledge and skills and to adjust instructional materials to the students' needs (Corbett & Anderson, 1995; Klinkenberg, Straatemeier, & Van Der Maas, 2011). The ALTs currently most frequently used in Dutch schools use an indicator of learners' knowledge during learning to select instructional materials, which is an approach often used to adapt education with technology (Aleven, McLaughlin, Glenn, & Koedinger, 2017). Research has shown that these systems support students' learning effectively in comparison to traditional non-adaptive instruction (Aleven et al., 2017; Faber, Luyten, & Visscher, 2017; Molenaar & Knoop - van Campen, 2019). Although ALTs successfully use student data to adjust instructional materials to learners needs, there has been little focus on how these systems influence students' self-regulated learning (SRL), and self-regulated learning is not a focus of most of the ALTs being used at scale (Winne & Baker, 2013). Instead, up until now research has primary focused on the cognitive effects of learning with ALTs (Aleven et al., 2017; Greene & Azevedo, 2010). The unique contribution of this paper is to increase our understanding of how students regulate their learning with ALTs.

1.1. Self-regulated learning with adaptive learning technologies

SRL theory defines learning as a goal-oriented process, in which students make conscious choices, working toward learning goals (Zimmerman, 2000). Self-regulating learners use cognitive activities (reading, practicing and elaboration) to study a topic, use metacognitive activities (orientation, planning, monitoring and evaluation) to control and monitor their learning and motivate themselves to engage in an appropriate level of learner effort (Greene & Azevedo, 2007). SRL is a dynamic iterative process that evolves over time and several researchers have expressed the need to explore time and order in SRL further (Greene & Azevedo, 2010; I.; Molenaar & Järvelä, 2014; Winne, 2010). Following the COPES model, effective self-regulating learners use learning goals to plan their learning and they obtain these goals by

* Corresponding author.
E-mail address: i.molenaar@pwo.ru.nl (I. Molenaar).

https://doi.org/10.1016/j.learninstruc.2019.05.003
Received 2 July 2018; Received in revised form 10 May 2019; Accepted 13 May 2019

adjusting their effort over time accordingly in the control loop (Hadwin, 2011). In addition, they continuously monitor whether their actions reach the level of accuracy that supports progress towards their learning goal in the monitoring loop (Greene & Azevedo, 2010; Winne, 2010). Hence, learner control and monitoring support the alignment between learners actions, effort, accuracy and learning goals. Accuracy has an important signaling role for learners to become aware of the need to adjust their effort. When accuracy drops due to factors other than lack of knowledge, students should increase effort to uphold accuracy. Hence accuracy can be viewed as a function of knowledge and effort and can be regulated by adjusting one or both of these elements.

In adaptive learning technologies that adjust instructional material to learners' knowledge states, part of the control and monitoring loop is taken over by the technology. These technologies select instructional materials (problems and/or instruction) based on an estimate of the students' current knowledge and/or the probability that a problem will be solved correctly by the student (Corbett & Anderson, 1995; Klinkenberg et al., 2011). Hence, the system executes elements of the control and monitoring loop normally performed by learners. In the control loop, the ALT selects instructional materials aligned with the students' learning goal that are adjusted to their current knowledge. This reduces the need for learners to plan their actions in the control loop. In the monitoring loop, the ALT overtakes the corrective selection of learner actions. When a learner does not make sufficient progress, many ALTs automatically reduce the difficulty of the problems. Without system support corrective actions are dependent on a student's monitoring, which is often imperfect (Roebers, 2017; van Loon, de Bruin, van Gog, & van Merriënboer, 2013). System control can therefore partially overcome students' reduced ability to adequately regulate effort and consequently accuracy in the monitoring loop.

Nevertheless, another important element in the control and monitoring loop, namely the adjustment of student' effort to uphold accuracy (Hadwin, 2011; Winne, 2010), remains the task of the student. Even though many ALTs assume that the right level of student effort is related to a specific probability of success, it is up to the student to apply the level of effort needed to maintain accuracy. Consequently, because regulation of effort and consequently accuracy plays an important role in regulation of learning, students continue to control an important element of self-regulated learning in these environments. Even though positive effects on students' learning have been found when comparing students’ learning with ALTs to traditional learning environments (Aleven et al., 2017; Faber et al., 2017; Molenaar & Knoop-van Campen, 2016), we have little insight into how students actually regulate their effort and accuracy within these systems and how it affects their learning (Bannert et al., 2017).

1.2. Utilization deficiency and personalized visualizations

Two important concepts related to how students regulate their learning are effectiveness and efficiency. The effectiveness relates to the aim of instruction to help learners acquire new knowledge and skills and to apply this in different contexts. When students engage in deep processing and constructive actions, the integration of knowledge into mental models is supported, leading to knowledge gain and transfer of knowledge (Bannert, Hildebrand, & Mengelkamp, 2009). In order to effectively integrate new knowledge and skills into existing knowledge structures, students should invest effort. More effective learning is indicated by a higher learning gain and transfer of knowledge to new contexts (Koedinger, Corbett, & Perfetti, 2012). Both the control and monitoring loop help learners to evaluate the effectiveness of their learning and adjust the required effort accordingly. Here the *efficiency* of learners' regulation, which is indicated by the relation between learners' actions and learning achievements, comes into play. Learners evaluate the cost of their actions in light of their learning effectiveness and amount of effort and time invested. Efficient regulators select those actions that yield the highest learning gain in relation to the least effort and time investment. For example, a learner has the choice to reach a learning goal by making numerous incremental steps that each require effort over a longer time versus taking large steps that require effort over a shorter time. In this example, the second scenario is more efficient (if it results in the same outcome) because the student invests less time. The efficiency of the learners' regulation is closely related to the control loop in which students select learning actions to reach their learning goals. This loop supports *efficient learning* without wasting time on learning actions that do not contribute to progress (Azevedo, 2009) and helps students to regulate the effort needed to successfully engage in the selected actions. In the monitoring loop, learners assess whether their actions and effort produce sufficient accuracy to make progress toward their learning goal. Thus, both the control and monitoring loop support efficient and effective learning.

A complicating factor is that the majority of learners do not regulate their learning sufficiently, leading to less efficient and effective learning (Azevedo, Moos, Greene, Winters, & Cromley, 2008). Students are faced with a so-called *utilization deficiency*, the failure to adequately activate the control and monitor loops during learning (Winne, 2010). Different techniques, such as prompts (Bannert, Sonnenberg, Mengelkamp, & Pieger, 2015), scaffolding (Molenaar, van Boxtel, & Sleegers, 2011) and virtual agents interacting with students (Azevedo et al., 2012), have been used to assist learners in developing self-regulated learning. There are two main disadvantages to such techniques. First, using these techniques students remain the acting agents and these techniques only support students' regulation when students follow the systems' advice. Research indicates that students very often do not use the provided support, either ignoring the support or choosing not to follow the advice given (Bannert et al., 2015; Harley, Taub, Azevedo, & Bouchet, 2018). This contrasts with adaptive learning technologies, where the technology takes over part of the regulation from the student. Hence this situation can be viewed as a hybrid form of human-system regulation. Second, the techniques mentioned above are not helping learners to make explicit inferences about how their learning actions, i.e. effort and accuracy, relate to progress towards their learning goal over time. As Winne states: *Without reliable, revealing and relevant data that* support *learners to make valid inferences about how they control and monitor their learning learners will be handicapped* (Winne & Hadwin, 2013). Therefore, it is important to provide continuous feedback to learners to accurately control and monitor their progress towards learning goals. Personalized visualizations that indicate how users regulate effort and accuracy could be powerful indicators to support both learner efficiency and effectiveness. Therefore a promising solution to overcome learners’ utilization deficiencies and hence support learner regulation of effort and thereby accuracy could be developed based on transforming learner data into personalized visualizations for learners. For example, Arroyo et al. (2007) showed students visualizations of their behavior and progress in order to convince them to stop gaming the system.

1.3. Moment-by-moment learning curves as personalized visualizations

A potential way to form personalized visualizations are moment-by-moment learning curves. Distinct from classic “learning curves” (e.g. Newell & Rosenbloom, 1981), moment-by-moment learning curves visualize the probability that a student learned a specific skill at a particular problem during learning (Baker, Hershkovitz, Rossi, Goldstein, & Gowda, 2013). These curves show how much the learner is likely to have learned at each problem-solving opportunity, which is a representation of progress over time. Moment-by-moment learning curves do not show accuracy directly (unlike classic learning curves) but instead show a distillation of student progress. Moment-by-moment learning curves also do not show effort directly, but a student looking at their own recent progress can connect that progress with their recent effort. In previous research Baker, Corbett, and Aleven (2008) found that the spikiness of the probability that a student just learned at a

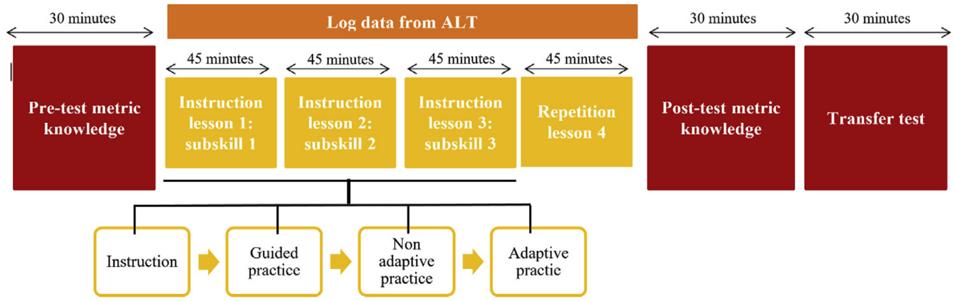


Fig. 1. Study design.

specific moment was associated with learning gain. This has been replicated in several studies (Baker, Goldstein, & Heffernan, 2011). Moreover, the visual patterns of the probability the student just learned (see Fig. 4) proved to be relatively stable and 7 different patterns were found that have been related to different learning outcomes (Baker et al., 2013). For example, immediate peak curves that showed a peak in learning early in the learning episode have been found to be associated with longer-term retention of learning. Immediate drop curves, where the probability of learning was high initially but went down rapidly, have been associated with higher immediate post-test scores.

Hence, evidence that learning outcomes are related to the form of moment-by-moment learning curves has been found. However, it is not currently known how (or whether) the visual patterns reflect students' regulation. At the same time, the curves visualize the probability that students have just learned which might also be informative of students' regulation in ALTs. Therefore, we propose to assess the relation between students' effort, accuracy and learning and different form of moment-by-moment learning curves to further understand how students regulate their learning with ALTs.

2. Purpose of the study

Students' learning with adaptive learning technologies (ALTs) on tablets leaves rich traces of data that capture many details of their learning process (Gašević, Dawson, & Siemens, 2015). Although ALTs successfully use student data to adjust instruction to learners' performance, we know very little about how students regulate their learning with ALTs (Winne & Baker, 2013). SRL theory defines learning as a goal-oriented process in which students make conscious choices, working toward learning goals (Zimmerman, 2000). Effective learners motivate themselves to engage in an appropriate level of effort and monitor accuracy to increases their learning (Boekaerts, 1999). Moment-by-moment learning curves, distilled automatically from students' data traces, visualize student progress in learning over time, which can be used to think about the role that effort plays in learning progress. In this study, we explore the value of moment-by-moment learning curves to further understand how students regulate their learning with ALTs. We study this in the context of real-world, grade five arithmetic learning in the Netherlands.

The following research questions are addressed:

1. What form of Moment-by-Moment learning curves are found?
2. How are Moment-by-Moment learning curves associated with effort and accuracy?
3. How are Moment-by-Moment learning curves associated with learning?
4. How do Moment-by-Moment learning curves interact with problem complexity in the subskills?

3. Method

98 students participated in this study. Students were in grade five of primary school and divided between four schools that participated in this study. The schools were located in the south and west of the Netherlands and had a diverse population. The inclusion criterion was that students had to participate in at least 3 lessons; 3 students were excluded from analysis. 51 boys (53%) and 44 girls (47%) participated in this study. The students were between 10 and 12 years old with a mean of 10.87(sd = 0.45). 5 students missed the pre-test and 4 students did not participate in the post-test and transfer test.

3.1. Design

This study was conducted with a pre-test/post-test design. All students worked on 3 arithmetic subskills in 4 lessons of 45 min each. The lessons were a blended learning scenario with a mix of teacher instruction and individual practice with the ALT. The measurements took place prior to the first lesson (pre-test) and after the completion of all lessons (post-test and transfer test). After the pre-test, students received instruction following the direct instruction model and practiced the three subskills during 3 lessons of 45 min on three consecutive days. In the fourth lesson students were instructed to practice those subskills for which they had not yet received 2 stars. The stars indicated that students had reached their average ability level. Fig. 1 shows the design of this study.

3.2. Materials

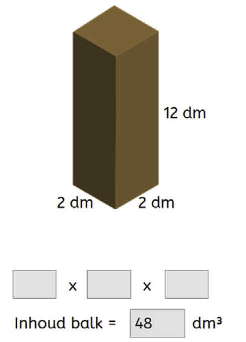
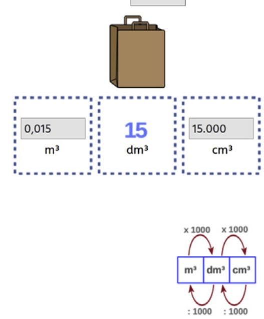
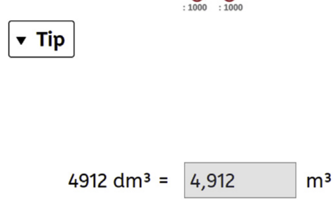
3.2.1. Adaptive learning technology studied

The adaptive learning technology used in this study runs on tablet computers and is widely used for arithmetic and spelling across schools in the Netherlands. In this ALT, students practice subskills after they had received instruction from their teachers and engaged in guided practice under supervision of their teacher. Individual practice takes place in two modes: during non-adaptive practice all students in the class worked on the same set of pre-selected problems, during adaptive practice problems were adjusted to the student' knowledge on the subskill.

The technology used a derivative of the Elo rating as an adaptive algorithm (Elo, 1978; Klinkenberg et al., 2011) that depends on an estimate of the knowledge of a student on a subskill and an estimate of each problem difficulty. The system used an item response model based on the Elo (1978) rating system as a scoring rule, see Klinkenberg for formulas. In the scoring rule problems were selected with a mean success probability of .75 and this parameter was selected based on previous studies (Eggen & Verschoor, 2006; Jansen, Hofman, Savi, Visser, & van der Maas, 2016). The estimate of the students' knowledge on a subskill and the problem difficulty were updated with every answered problem (Klinkenberg et al., 2011). This allows the system to calibrate while students are learning which optimizes the adaptive selection of the problems.

An additional feature in the ALT was that students received direct feedback, indicating a correct or incorrect response, immediately after finishing a problem. Teachers are also given dashboards on which they

Table 1
Examples of problems for each subskill.

	subskill	Example
Subskill 1	Calculate capacity using the formula: 'capacity = length x width x height'	
Subskill 2	Convert from common capacity units to cubic meters (cm^3, dm^3, m^3)	
Subskill 3	Convert cubic meters units (cm^3, dm^3, m^3) to liter units (cl^3, dl^3, l^3)	

can closely follow the performance of individual students. In this particular ALT, there is no error classification by the system – unlike systems such as Cognitive Tutor Algebra (e.g. Aleven et al., 2017), the system does not attempt to identify the misconception associated with a specific error. Instead, student errors can be viewed through the dashboard and teachers can provide elaborated feedback to students based on that information during the individual practice phase. This specific ALT also does not offer scaffolding or prompting functionality. Hence the adaptive selection of problems is currently the only adaptive mechanism at play in this system. Although this is relatively limited adaptivity compared to many intelligent tutoring systems (Aleven et al., 2017), it is in line with much of the adaptive learning software used in the Netherlands, the USA, and elsewhere in the world.

3.2.2. Subskills

The three subskills all included different aspects of measurements of capacity (see Table 1). The Dutch metric system units for measuring capacity were used.

The problems related to the first subskill "Calculate capacity using the formula: 'capacity = length x width x height" were relatively easy because students were given a formula to solve the problem. Also, in this subskill, examples were used to support students' problem solving. The problems related to the second subskill "Convert between common cubic capacity units" were of medium difficulty. Students were asked to convert from common capacity units into cubic meters (cm^3, dm^3, m^3). Finally, problems within the third subskill "Convert cubic meters units to liter units" were hard. Students were asked to convert cubic meters (cm^3, dm^3, m^3) into cubic liter units (cl^3, dl^3, l^3) without a formula, see Fig. 2 for more examples.

3.3. Measurements

3.3.1. Pre- and post-test

The pre- and post-test consisted of 24 items, 8 items per subskill. The items in the pre- and post-test were structurally similar, but different digits were used. The difficulty level of the items, as indicated by the ALT, was used to balance both tests. Fig. 2 provides examples of the items for each subskill. The overall Cronbach's alpha for the whole pre-test was 0.93 with 0.94 for subskill 1, 0.93 for subskill 2 and 0.74 for subskill 3 respectively. The overall Cronbach's alpha for the post-test was 0.91 with 0.74 for subskill 1, 0.92 for subskill 2 and 0.78 for subskill 3 respectively. Learning gain was calculated as the difference between pre- and post-test. The values (given in the results section below) indicated that there was limited evidence for a ceiling effect, requiring a more complex measure of learning gain.

3.3.2. Transfer test

The transfer test consisted of 15 items that tested students' understanding of the relations between meter units and liter units. The Cronbach's alpha for the transfer test was 0.85. Fig. 3 shows an example of an item in the transfer test. The values (given in the results section below) indicated that there was limited evidence for a ceiling effect requiring a more complex measure of learning gain.

3.3.3. Process measures from the ALT log data

The logs of the ALT stored data of students' practice activities, including a date and time stamp, student identifier, problem identifier, learning objective identifier, knowledge score after the problem and correctness of the answer given. Based on this information the following indicators of *effort* and *efficiency* were calculated.

Effort is measured by two indicators per subskill: the number of unique problems a student completed to practice this subskill, and the total number of attempts across all problems. The number of attempts was different from the number of problems as students can attempt to solve a problem multiple times.

Accuracy is calculated by dividing the number of correctly answered problems by the total number of problems completed. This is again calculated both for unique problems and for problem solving attempts at the level of each subskill. Table 2 provides an overview of all measures calculated and their definition.

3.4. Procedure

On the first day students were given 30 min to complete the pre-test after which the first instruction lesson of 45 min was given. The two other instruction lessons and the repetition lesson were given on separate consecutive days following the first lesson. On the fifth day students were given 30 min to complete the post-test and 30 min to complete the transfer test. Each instruction lesson started with 10 min of instruction given by the teacher. The instruction was standardized by using an instruction protocol. After the instruction, the teacher practiced 6 to 8 problems together with the students. Then students continued to work on problems within that particular subskill within the ALT. First students completed a set of 12–15 non-adaptive problems (the same problems for all students in the class) after which they continued to work on adaptively selected problems for the remaining time in the lesson. In the fourth lesson the three subskills of the previous lessons were repeated and practiced. Students were instructed to work on each subskill until the system had assessed that they had reached their average ability level as indicated by the system by two stars.

3.5. Moment-by-moment learning curves

The moment-by-moment learning curves were derived based on an algorithm developed by Baker et al. (2013). The curves indicated a student's learning at each practice opportunity for a specific subskill.

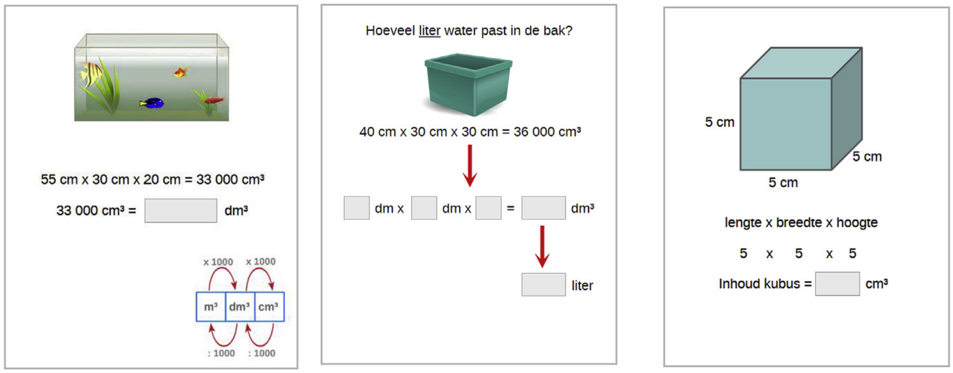


Fig. 2. Examples of items for the three subskills in the pre- and post-test.

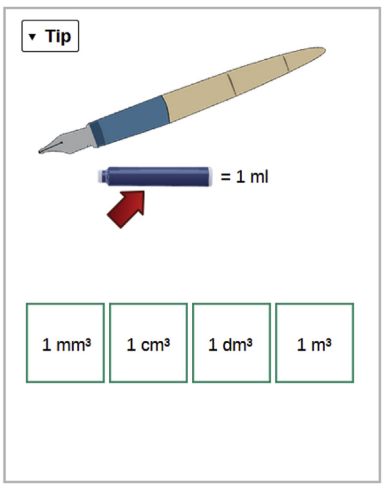


Fig. 3. Example of a problem in the transfer test.

Table 2
Overview of learning, effort and accuracy measures.

Learning measures	Definition
Prior knowledge	Pre-test, one per subskill
Post Knowledge	Post-test, one per subskill
Gain	Post-test/pre-test per subskill
Transfer score	Transfer test score
Effort measures	***log file data***
Unique problems	Number of unique problems completed per subskill
Problem solving attempts	Number of problem solving attempts per subskill
Accuracy measures	***log file data***
Accuracy unique problems	Correct problems/total unique problems completed
Accuracy problem solving attempts	Correct problems/total problem solving attempts

The subskills used in this paper are drawn from design of the ALT used in this paper's analyses, and in turn are included in the Dutch national curriculum guidelines (*SLO, http://tule.slo.nl/*). We use the term “subskill” rather than “knowledge component” (e.g. Baker et al., 2013) to reflect this origin of the skills. As such, we map the moment by moment learning curves directly to the skills measured by the system, the approach also used in (Baker et al., 2011). While more basic mathematical skills are relevant for the current skills, few adaptive learning systems at use at scale attempt to trace back errors to prerequisite skills. The logic behind this is that in a well-designed system and curriculum, the student should already have demonstrated mastery on prerequisite skills before reaching current content, and therefore prerequisite/basic mathematical skills should represent a small proportion of current student errors. A student lacking a prerequisite skill will likely fail to progress in general on the current content, and will appear to have not learned (correctly) within the moment-by-moment curves. As such, this limitation is unlikely to impact the results of our approach in a substantial fashion.

Moment-by-moment learning curves visualized the probability of a student learning a specific subskill at a specific practice opportunity using the moment-by-moment learning model (Baker et al., 2011). This probability was denoted as P(*J*) which refers to the probability that a student just learned. A high value of P(*J*) indicated that the probability that a student learned during that practice opportunity was high, whereas a low value of P(*J*) suggested that the probability that a student learned at that practice opportunity was low. The moment-by-moment learning model used a two-step process that combines two methods to estimate the probability that a student learned at a particular practice opportunity (Baker et al., 2011, 2013).

First, Bayesian Knowledge Tracing (BKT) (Corbett & Anderson, 1995) was applied to the log data to infer the probability of each student knowing each subskill at each time point. Knowledge tracing assumed a two-stage learning model in which each subskill was either in a *learned* or *unlearned* state. A subskill can make a transition from the unlearned to the learned state prior to instruction and practice or at each opportunity to apply the subskill. The goal of knowledge tracing was to produce estimates of $P(L_n)$, the estimated probability that a student had learned the subskill by time n, where n = 0 before the first opportunity to apply the subskill that the student encounters, n = 1 after the first opportunity, n = 2 after the second opportunity, and so on. This estimate was updated after every practice opportunity and the model assumed there was no forgetting. According to this model, knowledge does not map perfectly to performance; a student might be in the unlearned state and guess the answer correctly or in the learned state and make a mistake, giving an incorrect answer.

The probability that student performance does not match student knowledge is calculated by two performance parameters: S, the probability that a student who knows a skill will nonetheless produce an error, and G, the probability that a student who does not know a skill will nonetheless produce a correct response. A slip might occur, for instance, when a student knows how to do multiplication correctly, but makes an error adding up the numbers. Parameters are also calculated for the probability that a student who does not know a skill will learn it at any given practice opportunity (T), and the probability that a student knows the subskill before starting practice (L_0). The T parameter indicates the probability that a subskill will make the transition from the unlearned to the learned state following an opportunity to practice, regardless of the correctness of the answer. The L0 parameter is frequently above 0, as many students will already know a subskill before it is introduced by the teacher.

The common “brute force”/“grid search” method of fitting BKT was used to select the values of the parameters for each subskill. In the brute force/grid search approach to fitting BKT, all possible combinations of the four parameters were tried to obtain the estimate for each subskill for all students. The guessing and slipping probabilities were restricted to be below 0.3 and 0.1 following prior published recommendations (e.g. Baker et al., 2008; Corbett & Anderson, 1995). Next, all potential parameter combinations were tried for each subskill in order to determine which set of parameters fitted best, at a granularity of 0.01. Going forward, the combination of parameter estimates that resulted in the best sum of squared residuals (SSR) was used to calculate student knowledge at the first attempt at each subskill.

The formulas are given below, adapted, with permission, from Corbett & Anderson, 1995.

Calculation of SSR for each combination of L_0, T, Guess, Slip:

Likelihoodcorrect, indicating the probability that a student answers correctly, was calculated for each practice opportunity based on the previous estimated probability that the student knew the current subskill (prevL) and the probabilities for Guess and Slip. The possibility of slip is applied to the current estimated probability the student knew the skill, and the possibility of guess is applied to the current estimated probability the student did not know the skill (1 minus the current estimated probability they knew the skill). After that, Likelihoodcorrect was subtracted from the actual correctness of students' answer (StudentCorrectness) and squared to get the Squared Residual (SR). Then SR was summed to get SSR. The equations for these calculations were:

$$\text{Likelihoodcorrect} = (\text{prevL} * (1.0-\text{Slip})) + ((1.0-\text{prevL}) * \text{Guess})$$

$$\text{SSR} = \sum (\text{StudentCorrectness} - \text{likelihoodcorrect})^2$$

After computing the SSR, we computed the student's new probability of knowledge (the next action's previous probability of knowledge, prevLgivenresult). We do so first by updating our estimate of the student's prior knowledge based on their current accuracy:

$$\text{prevLgivenresult} = \frac{(\text{prevL}*(1.0 - \text{Slip})}{(\text{prevL}*(1-\text{Slip})) + ((1.0-\text{prevL})*(\text{Guess}))} \quad (when\ \ "WRIGHT")$$

$$\text{prevLgivenresult} = \frac{(\text{prevL} * (\text{Slip})}{(\text{prevL}*(\text{Slip})) + ((1.0-\text{prevL})*(1.0-\text{Guess}))} \left(when\ \ "WRONG"\right)$$

Bayesian Knowledge Tracing then updates the student's knowledge state based on the possibility that they learned from the current opportunity to practice the subskill

$$\text{prevL(next action)} = \text{prevLgivenresult} + (1.0-\text{prevLgivenresult})*\text{T}$$

The parameter values that led to the model that best fit the data in this study were:

In the second step, to estimate the probability that a student just learned, P(*J*), the estimates of the BKT parameters (L_0, T, G, S), were used in a second set of formulas (Baker et al., 2008). Table 3 displays the values for this study. In this set of formulas all practice opportunities were labeled as step ***N*** (indicating the ***N***th opportunity for a given student to learn the given subskill) with the probability that the student knew the subskill at that time, according to BKT, as input for the formula. Within “learning at step ***N***” on a subskill, we defined “at step ***N***” as the time when the student enters their first answer for a problem, in accordance with the past usage of the BKT algorithm (e.g. Baker et al., 2008; Baker et al., 2011; Corbett & Anderson, 1995). The probability that the student knew the subskill before answering on step ***N*** was estimated using BKT. Then, information about the performance on the following two steps (***N+1, N+2***) was used to estimate the probability that the student learned the subskill during that particular step ***N***. For example, if the student probably did not know the subskill at step ***N***, but answered ***N+1*** and ***N+2*** correctly, it was relatively likely that the student learned the subskill at step ***N***. Correspondingly, if the answers at steps ***N+1*** and ***N+2*** are incorrect, it is relatively unlikely that the student learned the subskill at step ***N.***

The formulae below are adapted, with permission, from Baker et al., 2013, p. 12.

We can assess the probability that the student learned the subskill at step ***N,*** given information about the actions at steps ***N+1*** and ***N+2*** (which we term $\boldsymbol{A}_{+1+2}$), as:

$$P\,(\boldsymbol{J}) = P(\sim \boldsymbol{L}_n \hat{}\ \boldsymbol{T} \mid \boldsymbol{A}_{+1+2})$$

In other words, the estimate P(J) is computed as the probability that the student did not know the subskill ($\sim L_n$, or 1 minus L_n, the estimated probability the student knew the skill immediately before the practice opportunity) and then learned it at the current practice opportunity (represented with the parameter T, which represents the probability that if the subskill is not yet learned, the subskill is learned at this practice opportunity). Classical Bayesian Knowledge Tracing assumes that P(T) does not fluctuate across situations – that a stable parameter can represent learning across all situations a subskill is used. The Moment-by-Moment Learning Approach (Baker et al., 2011) used here assumes that actual learning is contextual, and calculates it using the formula above. Note that P(J) captures whether the skill was just learned, distinct from P(L_n), which represents whether the skill is currently known (had been learned by time n). This notation, though a bit confusing, is used in order to be consistent with past published articles (e.g. Baker et al., 2011; Corbett & Anderson, 1995).

We estimate P(***J***)'s value using Bayes' Rule:

$$P\left(\sim \boldsymbol{L}_n \hat{}\ \boldsymbol{T} \mid \boldsymbol{A}_{+1+2}\right) = \frac{P\,(\boldsymbol{A}_{+1+2} \mid \sim\ \boldsymbol{L}_n \hat{}\ \boldsymbol{T}\,) * P\,(\sim \boldsymbol{L}_n \hat{}\ \boldsymbol{T})}{P\,(\boldsymbol{A}_{+1+2})}$$

The base probability P($\sim \boldsymbol{L}_n \hat{}\ \boldsymbol{T}$) is computed fairly simply, using the student's current value for P($\sim \boldsymbol{L}_n$) from Bayesian Knowledge Tracing (1 minus P($\boldsymbol{L}_n$), the current probability of knowing the skill), and the Bayesian Knowledge Tracing model's value of P(***T***) for the current subskill:

$$P(\sim \boldsymbol{L}_n \hat{}\ \boldsymbol{T}\,) = P\,(\sim \boldsymbol{L}_n)P(\boldsymbol{T})$$

The probability that the actual responses at time ***N+1*** and ***N+2***, P($\boldsymbol{A}_{+1+2}$), would be seen was computed as a function of the probability of the responses given each possible case: 1) the subskill was already known, P($\boldsymbol{L}_n$), 2) the subskill was unknown but was just learned, P($\sim \boldsymbol{L}_n \hat{}\ \boldsymbol{T}$), or 3) the subskill was unknown and was not learned, P($\sim \boldsymbol{L}_n \hat{}\ \sim \boldsymbol{T}$), combined with the contingent probabilities of each of these cases – an application of the extended form of Bayes' Rule.

$$P(\boldsymbol{A}_{+1+2}) = P(\boldsymbol{A}_{+1+2} \mid \boldsymbol{L}_n)P(\boldsymbol{L}_n) + \boldsymbol{P}(\boldsymbol{A}_{+1+2} \mid \sim\ \boldsymbol{L}_n \hat{}\ \boldsymbol{T})P(\sim \boldsymbol{L}_n \hat{}\ \boldsymbol{T}) + \boldsymbol{P}(\boldsymbol{A}_{+1+2} \mid \sim\ \boldsymbol{L}_n \hat{}\ \sim \boldsymbol{T})\,P(\sim \boldsymbol{L}_n \hat{}\ \boldsymbol{T})$$

The probability of the actions at time ***N+1*** and ***N+2***, in each of these three cases, was a. function of the Bayesian Knowledge Tracing model's parameters for the probability of a student guessing (***G***), slipping (***S***), and learning the subskill (***T***). In order to calculate the probability of each possible case of estimated student knowledge, we must consider all four potential scenarios of performance at actions ***N+1*** and ***N+2***. In the formulae below, correct answers are written as **C** and

Table 3
BKT parameter values derived from data.

	L_0	T	Guess	Slip
Subskill 1	.536	.101	.232	.1
Subskill 2	.027	.059	.250	.1
Subskill 3	.001	.149	.299	.1

incorrect answers are written as ~ **C**. The possible scenarios are: correct/correct (**C**, **C**); correct/incorrect (**C**, ~**C**); incorrect/correct (~**C**, **C**); and incorrect/incorrect (~**C**, ~**C**):

Case where next two actions are right, Case where next action is right and following action is wrong, and student already knew subskill and student already knew subskill

$$P(\boldsymbol{A}_{+1+2} = \boldsymbol{C},\ \boldsymbol{C}|\boldsymbol{L}_n) = P(\sim\boldsymbol{S})^2\ \ P(\boldsymbol{A}_{+1+2} = \boldsymbol{C},\ \sim\boldsymbol{C}|\boldsymbol{L}_n) = P(\sim(\boldsymbol{S})P(\boldsymbol{S})$$

Case where next action is wrong and following action is right, Case where next two actions are wrong,and student already knew subskill and student already knew subskill

$$P(\boldsymbol{A}_{+1+2} = \sim\boldsymbol{C},\ \boldsymbol{C}|\boldsymbol{L}_n) = P(\boldsymbol{S})P(\sim\boldsymbol{S})\ \ P(\boldsymbol{A}_{+1+2} = \sim\boldsymbol{C},\ \sim\boldsymbol{C}|\boldsymbol{L}_n) = P(\boldsymbol{S})^2$$

Case where next two actions are right, Case where next action is right and following action is wrong, and student just learned subskill and student just learned subskill

$$P(\boldsymbol{A}_{+1+2} = \boldsymbol{C},\ \boldsymbol{C}\ |\!\sim \boldsymbol{L}_n\hat{}\ \boldsymbol{T}) = P(\sim\boldsymbol{S})^2\, P(\boldsymbol{A}_{+1+2} = \boldsymbol{C},$$
$$\sim\boldsymbol{C}\ |\!\sim \boldsymbol{L}_n\hat{}\ \boldsymbol{T}) = P(\sim\boldsymbol{S})P(\boldsymbol{S})$$

Case where next action is wrong and following action is right, Case where next two actions are wrong. ,

and student just learned subskill and student just learned subskill

$$P(\boldsymbol{A}_{+1+2} = \sim\boldsymbol{C},\ \boldsymbol{C}\ |\!\sim \boldsymbol{L}_n\hat{}\ \boldsymbol{T}) = P(\boldsymbol{S})P(\sim\boldsymbol{S})\ \ P(\boldsymbol{A}_{+1+2} = \sim\boldsymbol{C},$$
$$\sim\boldsymbol{C}\ |\!\sim \boldsymbol{L}_n\hat{}\ \boldsymbol{T})\ = P(\boldsymbol{S})^2$$

Case where next two actions are right, and student did not know subskill and did not just learn it

$$P(\boldsymbol{A}_{+1+2} = \boldsymbol{C},\ \boldsymbol{C}\ |\!\sim \boldsymbol{L}_n\hat{}\ \sim\boldsymbol{T}) = P(\boldsymbol{G})P(\sim\boldsymbol{T})P(\boldsymbol{G}) + P(\boldsymbol{G})P(\boldsymbol{T})P(\sim\boldsymbol{S})$$

Case where next action is right and following action is wrong, and student did not know subskill and did not just learn it

$$P(\boldsymbol{A}_{+1+2} = \boldsymbol{C},\ \sim\boldsymbol{C}\ |\!\sim \boldsymbol{L}_n\hat{}\ \sim\boldsymbol{T}) = P(\boldsymbol{G})P(\sim\boldsymbol{T})P(\sim\boldsymbol{G}) + P(\boldsymbol{G})P(\boldsymbol{T})P(\boldsymbol{S})$$

Case where next action is wrong and following action is right, and student did not know subskill and did not just learn it

$$P(\boldsymbol{A}_{+1+2} = \sim\boldsymbol{C},\ \boldsymbol{C}\ |\!\sim \boldsymbol{L}_n\hat{}\ \sim\boldsymbol{T}) = P(\sim\boldsymbol{G})P(\sim\boldsymbol{T})P(\boldsymbol{G}) + P(\sim\boldsymbol{G})P(\boldsymbol{T})$$
$$P(\sim\boldsymbol{S})$$

Case where next two actions are wrong,and student did not know subskill and did not just learn it

$$P(\boldsymbol{A}_{+1+2} = \sim\boldsymbol{C},\ \sim\boldsymbol{C}\ |\!\sim \boldsymbol{L}_n\hat{}\ \sim\boldsymbol{T}) = P(\sim\boldsymbol{G})P(\sim\boldsymbol{T})P(\sim\boldsymbol{G}) + P(\sim\boldsymbol{G})$$
$$P(\boldsymbol{T})P(\boldsymbol{S})$$

Through this process, each action was labeled with estimates of the probability P(***J***) that the student learned the subskill at that time.

3.5.1. *Coding of the moment-by-moment learning curves graphs*

The P(***J***) of each unique problem was plotted in the moment-by-moment learning curves graphs per subskill. Only graphs with 15 or more problems were included, resulting in 265 graphs. The form of the graphs were coded following Baker et al. (2013) by 2 trained human coders. We used 4 codes (see Table 4) to code the graphs, as the previously found forms "single peak", "constant" and "plateau" (e.g. Baker et al., 2013) were not present in our data. It is worth noting that we define immediate peak somewhat differently from Baker et al. (2013), defining an immediate peak as occurring in the first 10 practice opportunities rather than the first 2 practice opportunity, in order to separate the learning occurring during the pre-test problems all students received from the adaptive problems received later. All graphs were coded by two coders resulting in an overall kappa of 0.74 indicating substantial agreement.

3.5.1.1. *Analysis*. We used t-tests to assess students' learning on the three subskills. To determine the how the two effort variables (unique problems solved and problem solving attempts), the two accuracy variables (accuracy problems solved and problem solving attempts) and the three learning (pre-test, post-test and gain) were related to different forms of MbMLC, we fitted seven Linear Mixed Models, fit using REML (Restricted Maximum Likelihood), with unique problems solved, problem solving attempts, accuracy problems solved, accuracy problem solving attempts, pre-test, post-test and gain each as dependent variable, a fixed effect for form of MbMLC and random effects for student and subskill, using the lme4 package in R (Bates, Maechler, Bolker, & Walker, 2015). The significance of parameter estimates was determined by t-tests using Satterthwaite's method (Kuznetsova, Brockhoff, & Christensen, 2017). In order to test the differences between the four moment-by-moment learning curves we used post-hoc pairwise comparisons of estimated means with Kenward-Roger adjustment of the degrees of freedom (Lenth, 2007). We applied the Tukey correction for multiple comparisons. Finally we counted the number of different MbMLC that contributed to the transfer score and used a linear regression model to estimate the contribution of each curve to students' transfer score. The interaction between problem difficulty (easy, medium, hard) for different subskills and the form of the moment-by-moment learning curves was assessed using a chi-square analysis.

4. Results

4.1. *Descriptive*

The students undertook a total of 16,751 problems in 19,433 problem solving attempts, see Table 5. The average number of problems completed per student was 176.33 ($sd = 52.84$) and the average number of problem solving attempts was 204.56 ($sd = 63.70$). The average accuracy for problems was .83 ($sd = 0.18$) and 0.67 ($sd = 0.19$) for problem solving attempts. The students' average score on the pre-test was 8.53 ($sd = 6.15$) and 16.66 ($sd = 6.11$) on the post-test, resulting in an average learning gain of 8.85 ($sd = 6.42$). The average transfer score was 10.25 ($sd = 4.11$). A paired samples *t*-test indicated that students learned $t(85) = 12.54$, $p < 0.001$, $r = 0.35$, which is a medium effect, during the four lessons overall and for all the three subskills independently.

Table 4
Coding rules for moment-by-moment learning curves.

Curve form	Rules	Meaning
Immediate drop	The curve starts high, drops within the first few practice opportunity and remains low afterwards.	High prior knowledge, the student may already know the subskill.
Immediate peak	The curve starts low, peaks within the first 10 practice opportunities and remains low afterwards.	Quick initial learning indicating the student quickly understands the subskill.
Close multiple spikes	The curve starts low, shows 2 or more peaks within the first 25 practice opportunities and remains low afterwards.	Consecutive set of learning events during practice towards understanding the subskill.
Separated multiple spikes	The curve starts low, shows 2 or more peaks within the first 25 practice opportunities and continues to show peaks thereafter.	Multi-phase gradual learning during practice towards understanding the subskill.

Table 5
The descriptive.

	Subskill 1		Subskill 2		Subskill 3		All subskills	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Pretest	5.16	3.22	1.61	2.58	1.76	1.94	8.53	6.13
Posttest	6.83	1.61	5.73	2.85	4.02	2.49	16.66	6.08
Gain	1.83	2.96	4.36	3.21	2.51	2.44	8.85	6.15
Problems	52.97	18.88	53.71	19.98	54.41	23.33	176.33	52.66
Attempts	61.48	23.22	69.80	27.80	73.27	30.97	204.56	63.70
Accuracy problems	.91	.13	.79	.21	.79	.17	.83	.18
Accuracy attempts	.79	.14	.61	.19	.59	.16	.67	.19

4.2. Form of moment-by-moment learning curves

We found four visual patterns of learning curves, shown in Fig. 4. *Immediate peak* were most frequent with 118 curves (45%), followed by *immediate drop* (occurred 66 times, 25%). Both *separated multiple spikes* (46 occurrences, 17%). and *close multiple spikes* (35 occurrences, 13%) were quite frequent, whereas *single spike* curves did not occur.

4.3. Effort and form of moment-by-moment learning curves

Differences in students' effort were compared for the form of the moment-by-moment learning curves, see Fig. 5. There was no difference in the number of unique problems solved between different moment-by-moment learning curves, see Table 6, nor for the number of problem solving attempts, see Table 7.

4.4. Accuracy and form of moment-by-moment learning curves

The differences in students' accuracy were compared for the form of the moment-by-moment learning curves, see Fig. 6. Student accuracy on unique problems differed significantly between different moment-by-moment learning curves, see Table 8. *Immediate drop* showed higher accuracy than the other curves (*immediate peak* $t(209.15) = 5.03$, $p < 0.001$, *close multiple spikes* $t(193.66) = 5.73$, $p < 0.001$, *separate multiple spikes* $t(205.05) = 4.94$, $p < 0.001$). and the accuracy of the *immediate peak* was higher than *close multiple spikes* $t(212.02) = 2.53$, $p < 0.01$, *separate multiple spikes* $t(226.69) = 2.11$, $p < 0.03$). The accuracy of problem solving attempts also differed significantly between different moment-by-moment learning curves, Table 9. All curves differed significantly from each other with respect to accuracy on problem solving attempts, with a downward trend in accuracy as curve complexity increased, see Table 10.

4.5. Learning and form of moment-by-moment learning curves

Differences in students' learning were compared for the form of the moment-by-moment learning curves, see Fig. 7. Pre-test scores differed significantly between different moment-by-moment learning curves, see Table 11. *Immediate drop* was associated with higher pre-test scores than all other curves (*immediate peak* $t(236.29) = 13.51$, $p < 0.001$, *close multiple spikes* $t(238.369.78$, $p < 0.001$, *separate multiple spikes* $t(245.22) = 8.33$, $p < 0.001$). Also, post-test scores differed significantly between different moment-by-moment learning curves, see Table 12. *Immediate drop* was associated with higher post-test scores than *close multiple spikes* $t(221.70) = 2.93$, $p < 0.01$. *Immediate peak* was associated with higher post-test scores than *close multiple spikes* $t(205.95) = 2.68$, $p < 0.01$. The gain scores also differed significantly

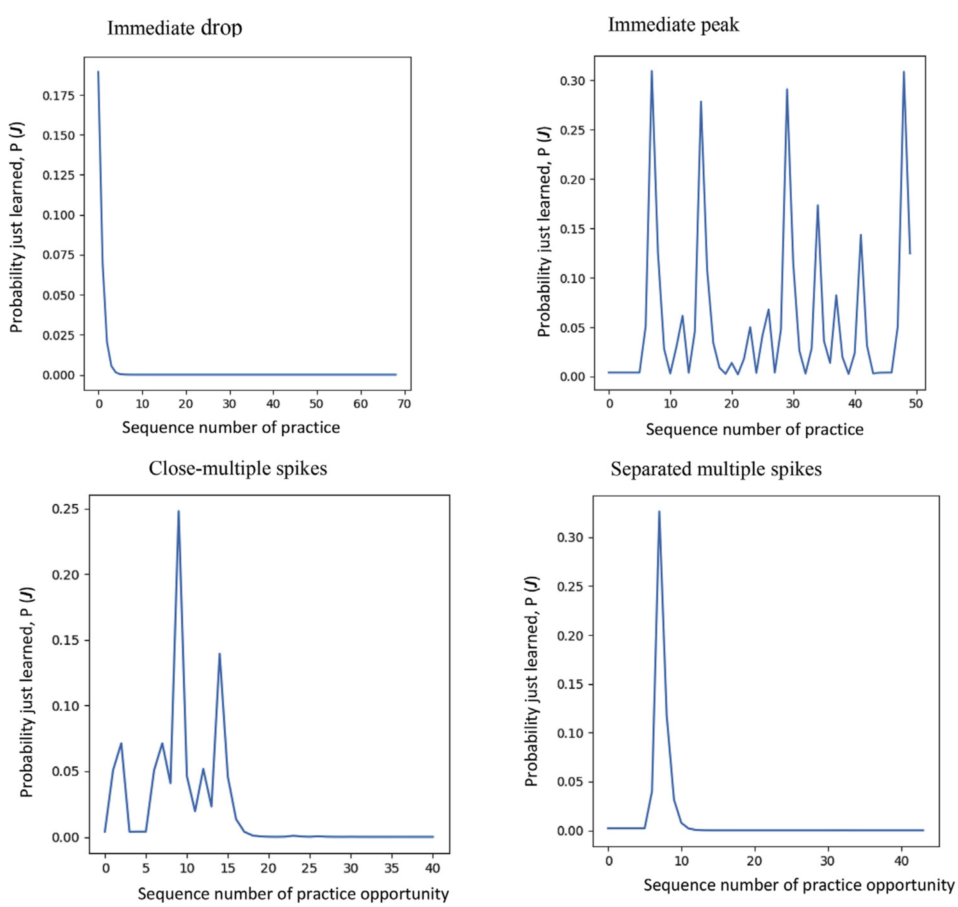


Fig. 4. Form of moment-by-moment learning curves.

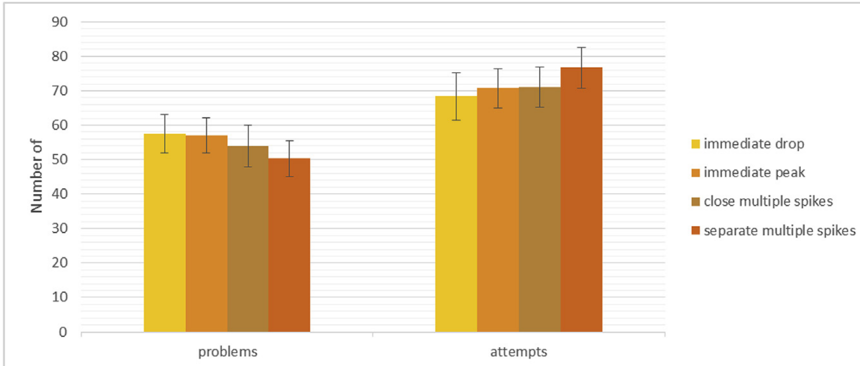


Fig. 5. Effort seen for form of learning curves.

Table 6
Mixed linear model problems solved by form of moment-by-moment learning curve.

	β	SE
Fixed Part		
Intercept immediate drop	54.92***	2.43
Immediate peak	1.43	2.57
Close multiple spikes	1.13	3.22
Separated multiple spikes	−1.73	3.69
Random Part		
Intercept student	171.8	13.11
Intercept subskill	0.00	0.00

$N_{problem} = 265$, $N_{student} = 92$, $***p < 0.001$, $**p < 0.01$.

Table 7
Mixed linear model problems solving attempts by form of moment-by-moment learning curve.

	β	SE
Fixed Part		
Intercept immediate drop	69.74***	4.84
Immediate peak	0.40	3.97
Close multiple spikes	0.43	5.20
Separated multiple spikes	5.70	5.86
Random Part		
Intercept student	211.16	14.87
Intercept subskill	31.91	5.65

Nattempt = 265, Nstudent = 92, ***p < 0.001, **p < 0.01.

Table 8
Mixed linear model percentage correct problems by form of moment-by-moment learning curve.

	β	SE
Fixed Part		
Intercept immediate drop	0.92***	0.02
Immediate peak	−0.09***	0.02
Close multiple spikes	−0.13***	0.02
Separated multiple spikes	−0.13***	0.03
Random Part		
Intercept student	0.01	0.09
Intercept subskill	0.01	0.08

$N_{pc\_correct} = 265$, $N_{student} = 92$, $***p < 0.001$, $**p < 0.01$.

Table 9
Mixed linear model percentage correct attempts by form of moment-by-moment learning curve.

	β	SE
Fixed Part		
Intercept immediate drop	0.76***	0.04
Immediate peak	−0.08***	0.02
Close multiple spikes	−0.13***	0.02
Separated multiple spikes	−0.19***	0.03
Random Part		
Intercept student	0.01	0.10
Intercept subskill	0.01	0.06

$N_{pc\_attempt} = 265$, $N_{student} = 92$, $***p < 0.001$, $**p < 0.01$.

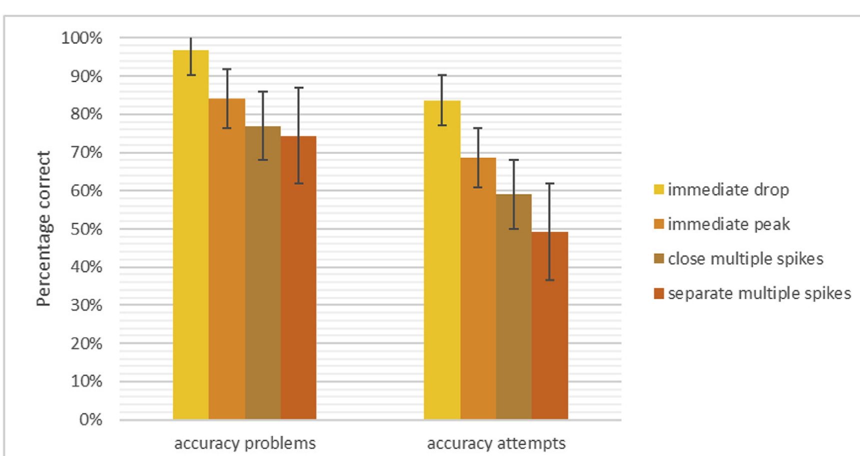


Fig. 6. Differences in accuracy between form of moment-by-moment learning curves.

Table 10
Contrast between forms of MbMCL on accuracy of problem solving attempts.

	df	*t*
Immediate drop versus immediate peak	220.02	4.55***
Immediate drop versus close multiple spike	226.14	5.56***
Immediate drop versus separated multiple spike	239.75	7.06***
Immediate peak versus close multiple spike	207.93	2.74***
Immediate peak versus separated multiple spike	223.59	5.15***
Close multiple spike versus separate multiple spike	206.54	2.63***

between different moment-by-moment learning curves, see Table 13. Because of its high pre-test scores, *immediate drop* was associated with a significantly lower gain than all other curves (*immediate peak* $t(212.20) = -9.70$, $p < 0.001$, *close multiple spikes* $t(209.98) = 2.89$, $p < 0.00$, *separate multiple spikes* $t(223.68) = 2.79$, $p < 0.01$). *Immediate peak* showed significantly higher gain compared to all other curves (*close multiple spikes* $t(204.41) = -5.16$, $p < 0.001$, *separate multiple spikes* $t(224.45) = -4.34$, $p < 0.001$).

Finally, we counted the number of different MbMLC that contributed to the transfer score and used a linear regression model to estimate the contribution of each curve to transfer. Transfer was significantly predicted by the count of all MbMLC, see Table 14. Specifically, *separate multiple spikes* have a negative impact on the transfer score.

4.6. Problem difficulty and form of moment-by-moment learning curves

Next, we examined the interaction between subskills and form of different moment-by-moment learning curves. There was a significant difference in form of moment-by-moment learning curves for the different subskills, chi-square analysis $\chi^2(df = 6,\ N = 265) = 73.91$, $p < 0.001$, see Fig. 8. More *immediate drop* curves are found for the easy problems in subskill 1 than for the medium and hard problems found in subskills 2 and 3. *Immediate peaks* occur frequently for all subskills, but slightly less often for the hard problems in subskill 3. Both *close multiple spikes* and *separated multiple spikes* became more frequent as the problem difficulty increased. This indicated that the form of the curves seems to be affected by the problem difficulty of the subskill.

5. Discussion

This exploratory study investigated the proposition that data generated in ALTs can not only facilitate adaptation of instructional materials to learners' performance, but also has the potential to show how learners regulate their learning over time. We had hypothesized that moment-by-moment learning curves could be indicative of learners'

Table 11
Mixed linear model pre-test by form of moment-by-moment learning curve.

	β	SE
Fixed Part		
Intercept immediate drop	5.99**	0.64
Immediate peak	−4.29***	0.31
Close multiple spikes	−4.07***	0.41
Separated multiple spikes	−3.91***	0.46
Random Part		
Intercept student	1.23	1.10
Intercept subskill	0.98	0.99

$N_{pre\text{-}test} = 250$, $N_{student} = 87$, $***p < 0.001$, $**p < 0.01$.

Table 12
Mixed linear model post-test by form of moment-by-moment learning curve.

	β	SE
Fixed Part		
Intercept immediate drop	6.11**	0.74
Immediate peak	−0.37	0.32
Close multiple spikes	−1.26**	0.42
Separated multiple spikes	−0.89	0.49
Random Part		
Intercept student	2.79	2.79
Intercept subskill	1.67	1.67

$N_{post\text{-}test} = 256$, $N_{student} = 89$, $***p < 0.001$, $**p < 0.01$.

regulation of effort and accuracy. We performed this study in arithmetic lessons within classrooms that worked over several days with this ALT. Students solved easy, medium and hard problems respectively for subskill 1, 2 and 3. The study examined how the form of moment-by-moment learning curves interacted with problem difficulty per subskill and was associated with students' effort, accuracy, learning per subskill. We found clear interactions between different moment-by-moment learning curves, regulation of accuracy and learning. Below we reflect on the meaning of the moment-by-moment learning curves in general and specifically for self-regulated learning.

Immediate drops were proposed to indicate that students already knew the skill (Baker et al., 2013). In this study it was found that *immediate drops* were indeed associated with high prior knowledge (pre-test), post-test knowledge and transfer scores, but were also associated with low learning gains which aligns with earlier findings. At the same time immediate drops were related to high accuracy. This indicated that students already knew the subskill to a large extent and were over-practicing to some degree, a hypothesis proposed in Baker et al. (2013). In the light of self-regulated learning, the data indicated that these students had little trouble in regulating their accuracy and the

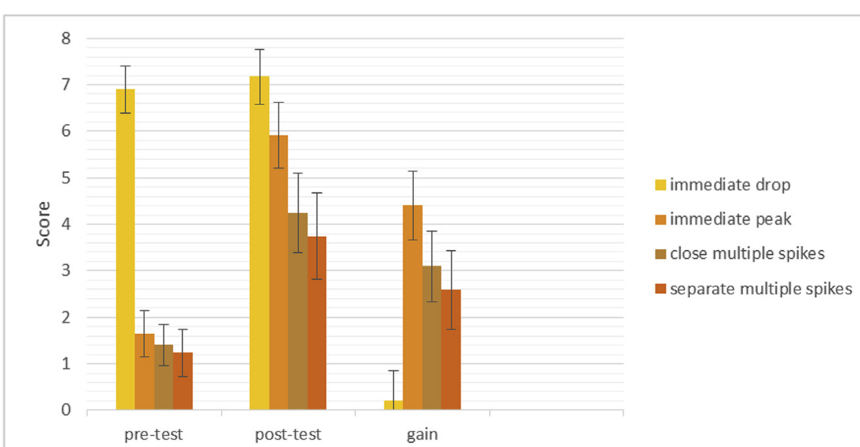


Fig. 7. Differences in learning between form of moment-by-moment learning curves.

Table 13
Mixed linear model gain by form of moment-by-moment learning curve.

	β	SE
Fixed Part		
Intercept immediate drop	0.33	0.61
Immediate peak	3.91***	0.39
Close multiple spikes	2.71***	0.51
Separated multiple spikes	2.60***	0.58
Random Part		
Intercept student	2.61	1.61
Intercept subskill	0.71	0.84

N_{gain} = 241, $N_{student}$ = 84, ***$p < 0.001$, **$p < 0.01$.

Table 14
Linear regression model predicting transfer by form of moment-by-moment learning curve.

	β	SE
Constant	11.31***	2.38
Immediate drop	.11	.94
Immediate peak	.06	.88
Close multiple spikes	-.73	.92
Separated multiple spikes	−2.28*	1.01

Note: $R^{2=}$ *0.13,* Nstudent = 92, ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$.

knowledge did transfer to other contexts. The efficiency of learning under these conditions might be questioned, but students in this study had no means to select different subskills to work on. A possible implication is that ALTs could provide more challenges to learners showing immediate drop curves to enhance learning efficiency.

Immediate peaks were proposed to indicate quick initial learning (Baker et al., 2013). Indeed, low prior knowledge (pre-test scores) indicated that students did not know the subskill in advance and high post-test knowledge showed that students had learned the subskill. Consequently, learning gains as well as transfer scores were high for these curves, indicating effective learning. This curve was associated with relatively high accuracy on unique problems and problem solving attempts. This seems to indicate that students were able to maintain reasonably high levels of accuracy. Based on this, the learning these students can be viewed as efficient and effective. However, it can be questioned whether students showing these curves actually benefit from ALT's adaptive support because these students appear to have been effectively regulating their own learning before the adaptive learning system started to support the students.

Close multiple spikes were expected to indicate a consecutive set of learning events (Baker et al., 2013). We found that they were associated with low prior knowledge, moderate post knowledge, moderate transfer and learning gain. At the same time, the accuracy of unique problems and the accuracy of the problem solving attempts was lower than for *immediate peak* and *drop*. However, the accuracy of problem solving attempts was higher than for *separate multiple spikes*. This indicated that students experienced challenges in sustaining their accuracy. Closer examination of the curves shows that the spikes in close multiple spikes curves tend to occur when students work on non-adaptive problems. When students started to practice with problems that are adjusted to the knowledge level of the student, students did show more gradual learning. This could indicate that students were incapable of effectively regulating their accuracy themselves, but did better when the ALT was taking over part of their regulation. Future research should explore the interaction between students and system regulation in more detail. For now, we view these patterns as indicating students who were unable to fully regulate their own learning during the initial non-adaptive problems but whose regulation challenges were possibly resolved by the adaptive learning system taking over control and monitoring their progress.

Finally, based on previous findings, *separated multiple spikes* were expected to indicate gradual learning (e.g. Baker et al., 2013). In the current study these patterns were associated with low prior knowledge and low to moderate post-knowledge, moderate learning gain and lower transfer compared to the other curves. Students showed reduced accuracy both in terms of the unique problems completed as well as the number of problem solving attempts. This indicates that students hardly learned and these curves also seemed indicative of students experiencing problems regulating their accuracy. Both learning efficiency and effectiveness are relatively low. It is arguable whether, in the context of an ALT, this curve indeed reflected gradual learning. We postulate that the curves might actually signal ineffective regulation, in which students were unable to maintain accuracy. The probability that the student had just learned in this situation showed an apparently random pattern that might be indicative of unsuccessful regulation. Even when the system adjusted the difficulty of problems the students were unable or unwilling to sustain higher levels of accuracy.

These conclusions were further supported by the finding that problem difficulty per subskill interacts with the form of the moment-by-moment learning curves. The easy problems in subskill 1 were primarily associated with *immediate drops* and *peaks*, whereas the medium problems in subskill 2 were associated with a decrease in immediate drops and an increase in *close* and *separate multiple spikes*. This result suggests that students did not have prior knowledge of the subskill 2 and that more students experienced challenges in regulating their effort and

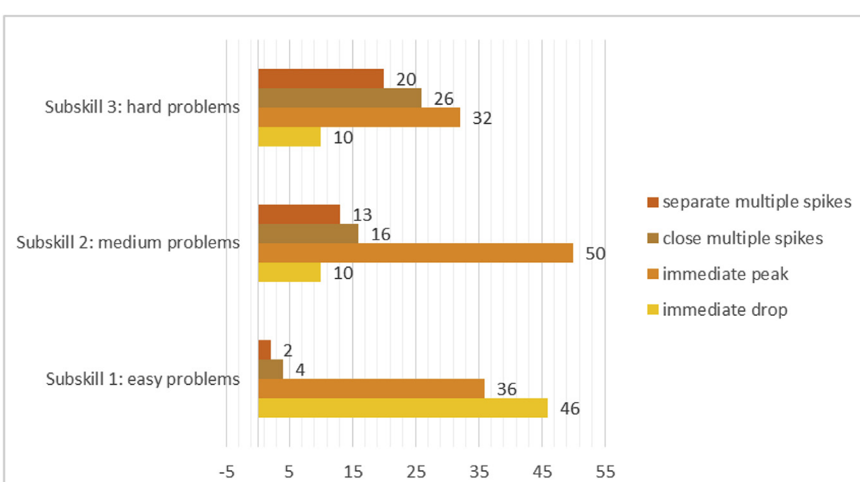


Fig. 8. Frequency of form of moment-by-moment learning curves for subskills with problems of different difficulty.

accuracy on this subskill. The hard problems in subskill 3 were associated with even fewer *immediate peaks* and more *close* and *separate multiple spikes*. Hence, it seemed that fewer students managed to learn this subskill without system support and more students experienced problems, even with system support. Analysis at the individual student level could further provide insights into how human-system regulation supports learners during learning with ALTs and whether there are specific characteristics that identify learners that experience problems.

Hence MbMLC reflects the efficiency of students' regulation of their learning over time. Immediate drops were indicative of inefficient regulation as students continued to practice even thought they had learned the subskill. Immediate peaks were indicative of efficient regulation where students experienced no problems in regulating their accuracy. Close multiple spikes indicated students that were experiencing initial issues in regulating their accuracy, but with the help of system regulation these students were able to regulate their learning more efficiently. Finally, separate multiple spikes indicated students that were unable to regulate their learning effectively over time.

In the introduction, we argue that learner control and monitoring supports the alignment between learners' actions, effort, accuracy and learning goals. Accuracy, in this loop, has an important signaling role for learners to become aware of the need to adjust their effort. When accuracy drops due to factors other than lack of knowledge, students should increase effort to maintain accuracy. Hence accuracy can be viewed as a function of knowledge and effort and can be regulated by the learner by adjusting his effort or by the system reducing the difficulty of problems. For example, when a learner makes a mistake, he or she could adjust effort by using a different strategy to solve problems. A possible new strategy could be to write difficult calculations on paper and hereby off-load short-term memory. This new strategy increases effort and is likely to support accuracy. Consequently, the transition to this new strategy will be visible in the MbMLC as a peak co-occurring with the change in the students' effort. In this case, the student increased effort by using a different strategy which resulted in increased accuracy.

An important limitation of the current study is that the number of problems solved is not completely indicative of the student's effort as indicated in the example above. In order to capture subtle differences in effort, it may be useful to improve our future measurement of effort with three important elements. First, harder problems may demand more effort compared to easier problems. One valuable area of future work would therefore be to include problem difficulty in the analysis of effort to better control for this aspect of performance when measuring students' effort on each problem. Second, the measure of effort could be improved by classifying errors in more detail. This would allow us to better distinguish between errors that result from lack of knowledge and errors that results from lack of effort. Third, current measures, i.e. unique problems solved and problem attempts, are both global measures. Efficient learners adjust their effort problem-by-problem, which is not captured by global measures. Hence, a local measure of effort that is related to each problem solved would be desirable to further conceptualize the interaction between effort, knowledge, and accuracy.

Future research questions to address are also the inclusion of error classification as slips or mistakes to improve the accuracy of the adaptive system's estimate of the students' knowledge level. This could improve alignment between the instructional material given and the students' needs which optimizes the functioning of system regulation. Moreover, this would facilitate the targeting of direct interventions on students' regulation with, for example, stop and think interventions – an intervention that has proven effective in other studies (Sande, Segers, & Verhoeven, 2018). Another possible future question is to explore how moment-by-moment learning curves could be used as an assessment tool for SRL. Our conclusions indicated that curves could potentially be used to distinguish good regulators from weak regulators and hence identify which learners are in need of additional support to effectively regulate their learning.

The purpose of this special issue is to further our understanding how multiple data streams can enhance our insights into SRL (Järvelä & Bannert, this issue). The implications of this study for SRL research are multiple. First, we now have a more advanced understanding of the diversity in how students regulate their learning with ALTs. Large differences in students' effort and accuracy were found in this study, which indicates broad variety in how students regulate their learning. Second, we have shown that the "probability just learned" metric can provide valuable insights into how students regulate the accuracy of their learning with ALTs over time. This shows that data derived from ALTs can indeed also be used to inform us about how students regulate their learning as proposed by Winne and Baker (2013). Third, contemporary perspectives view SRL as a process that unfolds over time (Greene & Azevedo, 2007; Molenaar & Järvelä, 2014; Paans, Moelnaar, Segers & Verhoeven, 2019). The moment-by-moment-learning curves show how much the learner is likely to have learned at each problem-solving opportunity, which is a representation of a student' progress over time. Combined with data indicating instructional phases during lessons which put different regulatory demands on students, this progress provides us with insights to understand when students are in need of additional support. This allows for further understanding of how regulation unfolds over time and when particular students face which regulation problems (Molenaar, Horvers & Baker, 2019). Fourth, SRL is often depicted as a social process (Miller, Hadwin & Järvelä, 2017; Molenaar, Sleegers, van Boxtel, 2014) that involves interaction of learners with others. Technologies also interact with learners and intelligent technologies could be potentially engage in social interactions around SRL with learners (Azevedo et al., 2016; Lester, Taylor, Sawyer, Culbertson, & Roberts, 2018). As discussed above, the results of this study further our understanding between the technical functioning of the ALT and the development of students' own regulation over time, this can help determine when and to what extent particular students need support from the system. Hence in light of the purpose of this special issue we have increased our understanding of when SRL processes take place during learning, how hybrid technology-human regulation takes form and how enactment of SRL over time is related to learning performance.

To conclude, this study showed large differences between learners in their effort, accuracy and learning while working with an ALT. Moment-by-moment learning curves were found to be related to learner accuracy and learning, and therefore providing insight into how students regulate their accuracy and learning over time while learning with ALTs. At the same time, variations in students' effort were not associated with different forms of learning curves. These findings increase our understanding of self-regulated learning with ALTs and highlight the importance to further study hybrid human-system regulation in detail. In this light, students' moment-by-moment learning curves provide insights into human-system regulation and can further enhance our understanding on how human-system interaction takes place around regulation. Going forward, we may be able to incorporate moment-by-moment learning curves into personalized visualizations designed to support students' inference about their accuracy and learning, and ultimately to develop richer self-regulated learning skills. Also, teachers could potentially use the curves as a diagnostic tool to analyze which students are in need of support.

Acknowledgements

This research was funded by a VENI grant awarded to dr. Molenaar by the Netherlands Organization for Scientific Research (project # 451-16-017). The authors would like to thank the schools, teachers and students for their participation and Jorieke Willems, Rick Dijkstra and Fred Hasselman for their support in the data collection, processing and analysis.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.learninstruc.2019.05.003.

References

Aleven, V., McLaughlin, E. A., Glenn, R. A., & Koedinger, K. R. (2017). Instruction based on adaptive learning technologies. In R. E. Mayer, & P. Alexander (Eds.). *Handbook of research on learning and instruction* (pp. 522–560). (2nd ed.). New York: Routledge.

Arroyo, I., Ferguson, K., Johns, J., Dragon, T., Mehranian, H., Fisher, D., et al. (2007). Repairing disengagement with non invasive interventions. *International Conference on Artificial Intelligence in Education* (pp. 195–202). August 2007.

Azevedo, R. (2009). Theoretical, conceptual, methodological, and instructional issues in research on metacognition and self-regulated learning: A discussion. *Metacognition and Learning, 4*(1), 87–95. https://doi.org/10.1007/s11409-009-9035-7.

Azevedo, R., Landis, R. S., Feyzi-Behnagh, R., Duffy, M., Trevors, G., Harley, J. M., ... Hossain, G. (2012). The effectiveness of pedagogical agents' prompting and feedback in facilitating co-adapted learning with MetaTutor. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): 7315*, (pp. 212–221). LNCS. https://doi.org/10.1007/978-3-642-30950-2_27.

Azevedo, R., Martin, S. A., Taub, M., Mudrick, N. V., Millar, G. C., & Grafsgaard, J. F. (2016). Are pedagogical agents' external regulation effective in fostering learning with intelligent tutoring systems? In A. Micarelli, J. Stamper, & K. Panourgia (Eds.). *Intelligent tutoring systems* (pp. 197–207). Cham: Springer International Publishing.

Azevedo, R., Moos, D. C., Greene, J. A., Winters, F. I., & Cromley, J. G. (2008). Why is externally-facilitated regulated learning more effective than self-regulated learning with hypermedia? *Educational Technology Research & Development, 56*(1), 45–72. https://doi.org/10.1007/s11423-007-9067-0.

Baker, R. S. J. d., Corbett, A. T., & Aleven, V. (2008). More accurate student modeling through contextual estimation of slip and guess probabilities in bayesian knowledge tracing. *Intelligent Tutoring Systems,* 406–415. https://doi.org/10.1007/978-3-540-69132-7_44.

Baker, R. S. J. D., Goldstein, A. B., & Heffernan, N. T. (2011). Detecting learning moment-by-moment. *International Journal of Artificial Intelligence in Education, 21*(1–2), 5–25. https://doi.org/10.3233/JAI-2011-015.

Baker, R. S., Hershkovitz, A., Rossi, L. M., Goldstein, A. B., & Gowda, S. M. (2013). Predicting robust learning with the visual form of the moment-by-moment learning curve. *The Journal of the Learning Sciences, 22*(4), 639–666. https://doi.org/10.1080/10508406.2013.836653.

Bannert, M., Molenaar, I., Azevedo, R., Järvelä, S., & Gašević, D. (2017, March). Relevance of learning analytics to measure and support students' learning in adaptive educational technologies. *Proceedings of the Seventh International Learning Analytics &; Knowledge Conference* (pp. 568–569). ACM.

Bannert, M., Hildebrand, M., & Mengelkamp, C. (2009). Effects of a metacognitive support device in learning environments. *Computers in Human Behavior, 25*(4), 829–835. https://doi.org/10.1016/j.chb.2008.07.002.

Bannert, M., Sonnenberg, C., Mengelkamp, C., & Pieger, E. (2015). Short- and long-term effects of students' self-directed metacognitive prompts on navigation behavior and learning performance. *Computers in Human Behavior, 52*, 293–306. https://doi.org/10.1016/j.chb.2015.05.038.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., & Rcpp, L. (2015). Package 'lme4'. *Convergence, 12*(1).

Boekaerts, M. (1999). Self-regulated learning: Where we are today. *International Journal of Educational Research, 31*(6), 445–457. https://doi.org/10.1016/S0883-0355(99)00014-2.

Corbett, A., & Anderson, J. (1995). Knowledge-tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction, 4*, 253–278. https://doi.org/10.1007/BF01099821.

Eggen, T. J. H. M., & Verschoor, A. J. (2006). Optimal testing with easy or difficult items in computerized adaptive testing. *Applied Psychological Measurement, 30*, 379–393.

Elo, A. (1978). *The rating of chessplayers, past and present. New York, Arco*.

Faber, J. M., Luyten, H., & Visscher, A. J. (2017). The effects of a digital formative assessment tool on mathematics achievement and student motivation: Results of a randomized experiment. *Computers & Education, 106*, 83–96. https://doi.org/10.1016/j.compedu.2016.12.001.

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends, 59*(1), 64–71. https://doi.org/10.1007/s11528-014-0822-x.

Greene, J. A., & Azevedo, R. (2007). A theoretical review of Winne and hadwin's model of self-regulated learning: New perspectives and directions. *Review of Educational Research, 77*(3), 334–372. https://doi.org/10.3102/003465430303953.

Greene, J. A., & Azevedo, R. (2010). The measurement of learners' self-regulated cognitive and metacognitive processes while using computer-based learning environments. *Educational Psychologist, 45*(4), 203–209. https://doi.org/10.1080/00461520.2010.515935.

Greller, W., & Drachsler, H. (2012). Translating learning into Numbers : A generic framework for learning analytics author contact details . *Educational Technology & Society, 15*(3), 42–57. https://doi.org/hdl.handle.net/1820/4506.

Hadwin, A. F. (2011). Self-regulated learning. In T. L. Good (Ed.). *21st century education: A reference handbook* (pp. 175–183). Thousand Oaks, California: Sage.

Harley, J. M., Taub, M., Azevedo, R., & Bouchet, F. (2018). "Let's set up some subgoals": Understanding human-pedagogical agent collaborations and their implications for learning and prompt and feedback compliance. *IEEE Transactions on Learning Technologies*. *11*. *IEEE Transactions on Learning Technologies* (pp. 54–66). . https://doi.org/10.1109/TLT.2017.2756629 1.

Jansen, B. R., Hofman, A. D., Savi, A., Visser, I., & van der Maas, H. L. (2016). Self-adapting the success rate when practicing math. *Learning and Individual Differences, 51*, 1–10.

Kennisnet (2014). *Onderwijs met een eigen device.* Retrieved from https://www.kennisnet.nl/fileadmin/kennisnet/publicatie/Onderwijs_met_eigen_device.pdf.

Klinkenberg, S., Straatemeier, M., & Van Der Maas, H. L. J. (2011). Computer adaptive practice of Maths ability using a new item response model for on the fly ability and difficulty estimation. *Computers & Education, 57*(2), 1813–1824. https://doi.org/10.1016/j.compedu.2011.02.003.

Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The knowledge-learning-instruction framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive Science, 36*(5), 757–798. https://doi.org/10.1111/j.1551-6709.2012.01245.x.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13.

Lenth, R. V. (2007). Statistical power calculations. *Journal of Animal Science, 85*(suppl_13), E24–E29.

Lester, J., Taylor, R., Sawyer, R., Culbertson, K., & Roberts, C. (2018, June). MetaMentor: A System Designed to Study, Teach, Train, and Foster Self-regulated Learning for Students and Experts Using Their Multimodal Data Visualizations. *Intelligent Tutoring Systems: 14th International Conference, ITS 2018, Montreal, QC, Canada, June 11–15, 2018, Proceedings: Vol.10858*, (pp. 411–). Springer.

van Loon, M. H., de Bruin, A. B. H., van Gog, T., & van Merriënboer, J. J. G. (2013). The effect of delayed-JOLs and sentence generation on children's monitoring accuracy and regulation of idiom study. *Metacognition and Learning, 8*(2), 173–191. https://doi.org/10.1007/s11409-013-9100-0.

Molenaar, I., Horvers, A., & Baker, R. (2019). Towards Hybrid Human-System Regulation: Understanding Children' SRL Support Needs in Blended Classrooms. *proceedings of the 9th International learning analytics & knowledge conference* (pp. 471–480). ACM.

Molenaar, I., & Knoop-van Campen, C. (2016). Learning analytics in practice: the effects of adaptive educational technology Snappet on students' arithmetic skills. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (LAK '16)* (pp. 538–539). New York, NY, USA: ACM.

Molenaar, I., & Järvelä, S. (2014). Sequential and temporal characteristics of self and socially regulated learning. *Metacognition and Learning, 9*(2)https://doi.org/10.1007/s11409-014-9114-2.

Molenaar, I., van Boxtel, C. A. M., & Sleegers, P. J. C. (2014). Metacognitive Scaffolding during Collaborative Learning: A Promising Combination. *Metacognition and learning, 9*(3), 309–332.

Molenaar, I., van Boxtel, C. A. M., & Sleegers, P. J. C. (2011). Metacognitive scaffolding in an innovative learning arrangement. *Instructional Science, 39*(6), 785–803. https://doi.org/10.1007/s11251-010-9154-1.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.). *Cognitive skills and their acquisition* (pp. 1–55). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Paans, C., Molenaar, I., Segers, E., & Verhoeven, L. (2019). Temporal variation in children's self-regulated hypermedia learning. *Computers in Human Behavior, 96*, 246–258.

Papamitsiou, Z., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Educational Technology & Society, 17*(4), 49–64.

Roebers, C. M. (2017). Executive function and metacognition: Towards a unifying framework of cognitive self-regulation. *Developmental Review, 45*, 31–51. https://doi.org/10.1016/j.dr.2017.04.001.

Sande, E. Van De, Segers, E., & Verhoeven, L. (2018). The role of executive functions for dyadic literacy learning in kindergarten. *Early Education & Development, 29*(2), 192–206. https://doi.org/10.1080/10409289.2017.1393739.

Winne, P. H. (2010). Improving measurements of self-regulated learning. *Educational Psychologist, 45*(4), 267–276. https://doi.org/10.1080/00461520.2010.517150.

Winne, P. H., & Baker, R. S. J. d. (2013). The potentials of educational data mining for researching metacognition, motivation and self-regulated learning. *JEDM - Journal of Educational Data Mining, 5*(1), 1–8. https://doi.org/10.1037/1082-989X.2.2.131.

Winne, P. H., & Hadwin, A. F. (2013). nStudy: Tracing and supporting self-regulated learning in the Internet. In R. Azevedo, & V. Aleven (Eds.). *International handbook of metacognition and learning technologies* (pp. 293–308). New York: Springer.

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. *Handbook of Self-Regulation, 13–39*https://doi.org/10.1016/B978-012109890-2/50031-7.