Computers in Human Behavior 145 (2023) 107754

Contents lists available at ScienceDirect

Computers in Human Behavior

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

Clustering children's learning behaviour to identify self-regulated learning support needs

S.H.E. Dijkstra [a,*], M. Hinne [b], E. Segers [a], I. Molenaar [a]

[a] Behavioural Science Institute, Radboud University, the Netherlands
[b] Donders Centre for Brain, Cognition and Behaviour, Radboud University, the Netherlands

ARTICLE INFO

Handling Editor: Prof. Paul Kirschner

Keywords:
Self-regulated learning
Learning behaviour
Bayesian nonparametric clustering

ABSTRACT

When children are learning using adaptive learning technologies (ALTs), the technology builds a learner model, which creates temporal trajectories providing insight into how children's knowledge develops. Based on this learner model, ALTs adjust the difficulty of problems for each child, yet children still need to regulate their practice behaviour and uphold effort and accuracy. The temporal trajectories are consequently likely to, besides showing children's knowledge development, be indicative of children's regulation. Therefore, we explore clusters of these trajectories to further identify failure in children's self-regulated learning (SRL) and potential support needs. We propose a data-driven approach to cluster 354 trajectories of 134 5th graders learning three skills with different complexity. The resulting 9 clusters were interpreted using practice accuracy and effort as indicators of regulation of practice behaviour and prior and post-knowledge and learning gain as indicators of knowledge development. The differences between clusters regarding these indicators signal there are different levels of SRL failure and, consequently, different SRL support needs: high accuracy and knowledge development indicate minimal support needs, whereas clusters with low accuracy, showing no knowledge development, indicate extensive SRL support needs. In conclusion: clusters of temporal patterns in children's learning data can identify SRL support is needed.

1. Introduction

Adaptive learning technologies (ALTs) are widely used by children in classrooms to support the learning of basic skills in math, arithmetic, and language (Aleven et al., 2016, van Wetering et al., 2020). During this learning, ALTs capture trace data from children, which are used to assess children's knowledge development of these skills. This prediction of children's current knowledge is used to adjust problems to the needs of individual children (Klinkenberg et al., 2011). In this way, problems are more adjusted to the children's zone of proximal development (Vygotsky, 1980). While this approach optimises children's learning compared to traditional non-adaptive technologies (Aleven et al., 2016), it also takes control away from the children (a phenomenon known as 'offloading'; Buxbaum-Conradi et al., 2016). This reduces the need for children to engage in self-regulated learning (SRL) (Martinez-Pons, 2002) and may negatively influence the development of self-regulated learning skills (Molenaar, 2022).

A potential solution lies in a system that gradually gives more control to the children ('onloading') so that they can increasingly self-regulate their learning (Molenaar et al., 2019). However, a prerequisite for developing such a system is a good indication of the SRL support needs of individual children. Therefore, the current study used data gathered by an ALT to identify how children regulate their learning and their consequent SRL support needs. Although the ALT takes over part of the regulation process, children continue to control their effort and, consequently, their accuracy (Molenaar, et al. 2019). The ALT's trace data provide insights into the temporal development of children's knowledge in open learner models (Kay et al., 2022; Long & Aleven, 2017). We propose that these temporal trajectories can also assess how children regulate effort and accuracy in these contexts. To further understand children's SRL support needs, we use a clustering model (McDowell et al., 2018) to detect similar temporal patterns in knowledge development among children working on arithmetic skills. The current study aimed to determine whether the identified clusters provide a meaningful method of identifying children's SRL during learning and consequent support needs. To do so, we examined how clusters relate to children's learning metrics in the form of practice accuracy and practice effort to assess regulation of practice behaviour and prior and

* Corresponding author. Behavioural Science Institute, Postbus 9104, 6500, HE, Nijmegen, the Netherlands.
E-mail address: rick.dijkstra@ru.nl (S.H.E. Dijkstra).

https://doi.org/10.1016/j.chb.2023.107754
Received 25 April 2022; Received in revised form 20 February 2023; Accepted 19 March 2023
Available online 25 March 2023

post-knowledge and learning gain to understand associations with knowledge development.

1.1. Self-regulated learning

SRL theory defines learning as a goal-directed process in which children make conscious decisions to reach their goals (Azevedo, 2007). Children use two types of processes while learning: cognitive processes, such as summarising, re-reading, and elaborating, which are used to study a topic or skill, as well as metacognitive processes, such as orienting, planning, monitoring, and evaluating. The latter are used to monitor and control the learning process and motivation, which drives children to maintain an appropriate level of effort and accuracy during learning (Greene & Azevedo, 2007). The COPES model (Winne & Hadwin, 1998) is a theoretical framework for SRL, according to which children, when they self-regulate effectively, monitor whether their actions contribute to reaching their goals (Azevedo, 2009) and subsequently enact control by adjusting their strategies or tactics to reach them (Winne, 2017).

1.1.1. The effect of adaptive learning technologies on self-regulated learning

When children are working with an ALT, part of their SRL processes is offloaded by the ALT (Molenaar et al. 2021). In the monitoring loop, the child still needs to monitor their accuracy by making "judgements of learning" to determine their progress during learning (Winne & Hadwin, 1998). The child's monitoring process is supported by the ALT with feedback, mostly direct knowledge of results, indicating whether a problem was solved correctly (Klinkenberg et al., 2011). This helps the child to assess progress and determine the need for control strategies.

In the controlling loop, the child is still in charge of adjusting effort when the need arises to uphold accuracy. This can be done by selecting new solution strategies, using additional support tools (e.g., a mini-whiteboard to write down intermediate steps), or by using a different heuristic to solve the problem. Eventually, they may ask the teacher or peers for help. The ALT does take over part of the control loop by selecting those problems that are estimated to be in the zone of proximal development for the child (Klinkenberg et al., 2011). Although children have multiple options to regulate their practice behaviour, the ALT is offloading regulation to some extent in this context. Advanced understanding of children's SRL support needs would allow us to develop a system that gradually gives more control to the children ('onloading') to increasingly allow them to self-regulate their learning.

1.1.2. SRL support needs

Research shows that many children find it difficult to effectively regulate their learning (Greene & Azevedo, 2010; Järvelä et al., 2013). SRL is often hampered by children's inability to initiate monitoring and control actions during learning, which leads to ineffective regulation and, hence, reduced learning (Winne & Baker, 2013). Furthermore, we know that SRL is contextual and influenced by children's prior knowledge and the complexity of skills to be learned (Taub et al., 2014). Consequently, many children need external support to successfully regulate their learning. This support may be personalized by a system that predicts children's SRL support needs (Molenaar, 2022).

However, measuring SRL has been challenging (Bannert et al., 2017; Veenman et al., 2006). For example, students' self-reported self--regulated learning is only poorly related to students' actual control and monitoring actions *during* learning (Schellings & Van Hout-Wolters, 2011; Winne & Jamieson-Noel, 2002). Therefore, it has been proposed to advance the online measurement of SRL during learning (Greene & Azevedo, 2010; Järvelä et al., 2013; Molenaar, 2014; Molenaar & Järvelä, 2014). But, existing methods such as video analysis (Azevedo, 2009) and think-aloud approaches are not scalable to educational contexts.

As a solution, using trace data from online learning environments may be a feasible approach (Azevedo & Gašević, 2019; Kay et al., 2022). An additional advantage is that measurement during learning could also improve our understanding of how SRL develops over time (Molenaar, 2014; Saint et al., 2022). Temporal characteristics of SRL are expected to help advance SRL theory on a micro level (Fan et al., 2021). These detailed insights will help to understand further how to support learners in their self-regulated learning processes.

When children are correctly self-regulating their learning, they monitor whether the actions they take result in the correct level of accuracy to reach their learning goal (Greene & Azevedo, 2010; Winne, 2010). If their accuracy drops, they increase their effort to regulate their accuracy to the correct level (Hadwin, 2011). Hence both accuracy and effort can be used as indicators of how well the SRL process is going and to identify if support is needed.

1.2. ALT trace data: ability curves

ALTs detect data when children solve problems successfully and unsuccessfully to learn a particular skill. Typically, these data are used to build a model of a learner's cognitive knowledge development of the skill (Kay et al., 2022). Many ALTs apply an adaptation of the Elo algorithm to build such a learner model (Elo, 1978; Klinkenberg et al., 2011). This algorithm was originally used in chess rankings to estimate player abilities when playing against other players (Elo, 1978). The player with the higher Elo score has a higher chance of winning than the player with the lower Elo score. After the chess match has been played, the Elo score is updated for both players, raising the winning player's score and lowering the losing player's score.

This Elo score has been adapted in the context of ALTs, where children are seen as "playing" against the problems (Klinkenberg et al., 2011). This adapted Elo score is also referred to as ability score when referring to the score for children and as difficulty score when referring to problems. If the child solves a problem correctly, it is seen as if the child has "won" against the problem and its ability score will be raised (and the difficulty score of the problem will be lowered). The ability score can, in this way, be seen as the chance that the child will solve the problem correctly. With this, the ALT can select problems that fit the children's current knowledge assessed by the child's ability score. This ability score represents the child's cognitive knowledge of the skill and indicates the difficulty needed for problems in the child's zone of proximal development (Vygotsky, 1980).

Next to individual ability scores calculated after each problem is solved, one can also look at how the ability score develops over time as the child learns the skill. This temporal process is captured in the form of a sequence of ability scores calculated by the ALT. We refer to this sequence as the ability curve. This curve visualises how a child's knowledge of a skill develops over time (Pelánek, 2016).

The ALT calculates the ability scores based on trace data that was measured when problems are solved by the children, which is related to the performance part of the COPES model (Winne & Hadwin, 1998). Indeed, previous research has shown that temporal trajectories of knowledge tracing can provide insight into children's learning (Baker et al., 2011) and related SRL support needs (Molenaar et al., 2021). Baker et al. showed that it is possible, using Bayesian knowledge tracing, to assess the probability that a child has learned a skill when solving a specific problem by constructing a moment-by-moment learning curve. In the initial research, Baker et al. showed the relation between different types of curves and children's learning (Baker et al., 2013). Molenaar et al. (2021) showed that there is merit to the assumption that these curves provide insight into children's SRL support needs: the form of the curves was associated with the children's regulation of accuracy and learning. Where regulation of accuracy means that children regulate their accuracy by, for example, rechecking whether they expect their answer to be correct before hitting the answer button.

The similarity between the moment-by-moment learning curves and the ability curves lies in the trace data they use. Both approaches use the information on problems solved as input to generate the temporal trajectories that represent children's knowledge development curves. Because of this similarity, we hypothesise that the shape of the ability curve, reflecting the ability score going up or down, may also reflect how children control and monitor their learning and can be used as a measurement of SRL support needs. A way to further asses this expectation is by clustering the ability curves and consequently relating them to children's learning and regulation of practice behaviour metrics.

1.3. Clustering SRL

Clustering is often applied to identify information about SRL in trace data (Saint et al., 2022). In previous studies, three different approaches to cluster trace data have been used: Aggregating trace data, mapping trace data to meaningful labels, and mapping trace data to temporal trajectories.

The first approach entails aggregating the trace data into variables and then performing clustering on these variables (e.g., Lust et al., 2011; Mirriahi et al., 2016). Examples of variables include the average time spent on solving a problem or the number of times a video is paused. This approach is used to find different tool-use (Lust et al., 2011) or learning profiles (Mirriahi et al., 2016) clusters. This strategy's drawback is that the trace data's temporal characteristics are lost in the analysis.

In the second approach, trace data are mapped into meaningful labels, sometimes called action labels. For example, by labelling logged navigation actions as particular behaviour such as planning or monitoring. Next, those labels are transformed into either counts of action sequences to find clusters of SRL profiles (Li et al., 2020) or action transition probabilities to find clusters of learning strategies (Fincham et al., 2019; Matcha et al., 2019; Saint et al., 2020; Uzir et al., 2020). A requirement for using this approach is rich enough trace data to capture different micro-level SRL processes, such as the planning and orientation process. However, trace data from ALTs are ontologically poor (Järvelä et al., 2019); as such, the labelling cannot be made, making it impossible to enact this approach.

The third approach used is mapping trace data to temporal trajectories of a learning process and clustering these temporal trajectories based on human judgement. One such approach can be performed using Bayesian Knowledge Tracing to generate moment-by-moment learning curves (mbmlcs; Baker et al., 2011). An mbmlc visualises the probability, for each problem solved, that the child has mastered the skill when working on that problem. These curves were then manually clustered (Baker et al., 2013). The mbmlcs were found to be indicative of children's learning (Baker et al., 2013) and regulation of practice behaviour (Molenaar et al., 2021). This approach does consider the temporal aspect of knowledge development but does not need detailed trace data like the second approach. A limitation is that the curves were clustered based on human judgement (Baker et al., 2013), possibly combined with coding rules based on these clusters (Molenaar et al., 2021), and this approach could only be applied to historic data, which limits its use to make predictions *during* learning.

1.4. Present study

The three approaches mentioned above, with the goal of identifying SRL support needs in mind, have drawbacks for the clustering of ability curves. Following the first approach, aggregating the ability curve into different variables will result in a loss of the temporal dynamics captured in the ability curve. Transforming trace data and ability curves into meaningful labels, following the second approach, is impossible, as the data is not rich enough. Lastly, the ability curves are too complex and diverse to cluster them manually, as was done in the third approach. Therefore, in the current study, we propose a clustering approach that captures the entire temporal pattern of the ability curve using the Dirichlet process Gaussian process mixture (DPGP) model developed by McDowell et al. (2018).

The DPGP model assumes that each cluster is expressed in a particular prototypical curve, and individual observations of an ability curve will then be closely similar to the prototypical curve. McDowell et al. (2018) used the DPGP model to find clusters of gene expression curves. In the current study, we modified the DPGP model such that the model finds clusters of ability curves.

A benefit of the DPGP model is that it is nonparametric (Orbanz & Teh, 2010). This means that no prior information is needed on the shape of the ability curves nor on the expected number of clusters. After the ability curves have been clustered, the identified clusters are interpreted with learning metrics such as practice accuracy, practice effort, and learning gain. These metrics are measured by the total amount of problems solved, the relative number of problems solved correctly, and the test scores on pre-tests and post-tests for the arithmetic skill.

In the present study, we are interested in identifying different types of temporal trajectories in ability curves and examining what these temporal trajectories can tell about children's SRL failure during learning and consequent support needs. The main research question was: To what extent can different temporal trajectories be used to identify the children's failure in self-regulated learning and consequent support needs? To determine how clusters indicate knowledge development and self-regulated learning, we relate them to the children's learning metrics, such as accuracy, practice effort, prior knowledge, post-knowledge, and learning gain. These associations allow us to assess the relationship between the clusters and learning metrics and consequently develop an understanding of how clusters indicate children's SRL and consequent support needs.

2. Methods

2.1. Participants

The study is based on data gathered in an earlier experiment where children learned with an adaptive learning technology (Horvers et al., Submitted). In total, 134 children participated in this study. The children were in grade 5 at four public primary schools in the Netherlands. There were 6 children with missing descriptive information. The average age for the rest of the children was 11 years and 0 months (sd = 7 months; Ages between 10 and 12, with one child age 13), with 47% boys and 53% girls.

The arithmetic levels of the children were determined using the Dutch national standardised mathematics assessment, CITO Mathematics. This assessment has 5 levels 22.4% of the children had the highest category, level 1. Level 2 had 17.2%, level 3 had 20.1%, level 4 had 22.4%, and level 5 had 11.2%. Data were missing for 6.7% of the children.

All children were already working with the ALT at their schools. However, they did not have prior training in the skills used in the experiments.

2.2. Design

The experiment was conducted in an authentic classroom situation with a quasi-experimental pre-test – post-test design. Children worked on three skills with different complexity over the course of four days. The children received instruction in the first three days and then practised the skill. On the fourth day, the children were allowed to choose which skill they continued to practice. On this fourth day, children had more space to regulate their learning and could choose to over-practice a skill. Before the lessons, children did a pre-test, and after the four days, they did a post-test (see Fig. 1).

2.3. Materials

The data used were gathered in an experiment using an ALT called Gynzy. This ALT is developed for children in primary education to learn

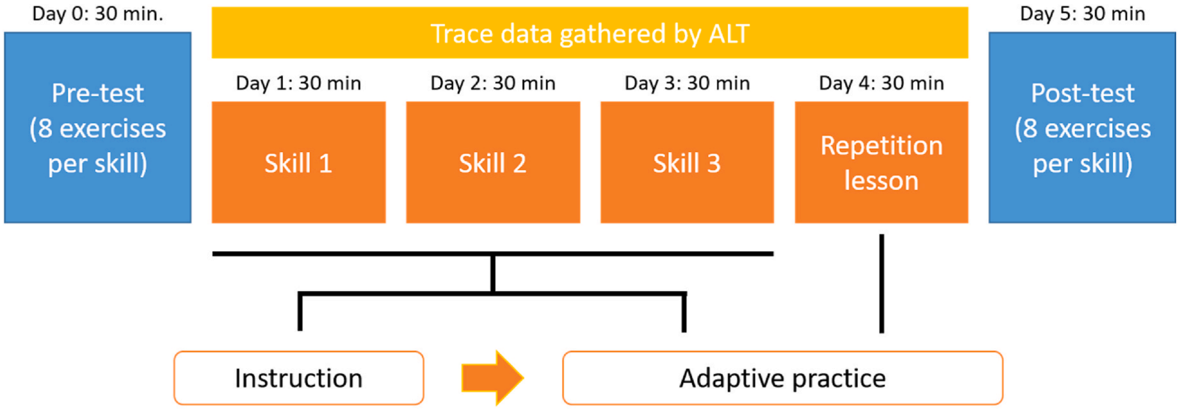

Fig. 1. Study design.

math and spelling in school and is used by over 700.000 young learners in The Netherlands. The ALT is used in blended classrooms. Children receive classical instructions about a skill and then work on the skill individually. The ALT uses the ability algorithm to select problems for each child, such that they have a chance of 75% to solve the problem correctly. When children practice a skill, they receive a block of 12 problems from the ALT. After each problem, a child receives direct feedback on whether the problem was solved correctly. If the problem is solved incorrectly, the child gets another chance directly. And if the answer is still incorrect, there is a third opportunity at the end of the block. After the block is finished, the child is shown the number of problems they correctly solved.

The data are collected from children working on three math skills on the ALT. These three skills were all on the topic of simplifying fractions and increased in complexity. The first skill, "Simplifying basic fractions", was the easiest of the three as it only used simple fractions, which were either smaller than one or were simplified to whole numbers without needing to divide by numbers larger than 10. The second skill, "Simplifying mixed fractions", expanded on this with fractions that contained both fractions and wholes. The third skill, "simplifying complex fractions", was the hardest, as the fractions now contained numbers that had to be divided by numbers bigger than 10. An example of each skill is given in Table 1.

2.4. Measurements

2.4.1. Learning metrics

The pre-test and post-test consisted of 24 problems each, 8 problems per skill. The difficulty of both tests was similar to each other by pre-selecting problems with similar ability scores. Learning gain was computed as normalised change (Marx & Cummings, 2007). This method was chosen over the normalised gain (Hake, 1998), as the normalised gain cannot handle perfect pre-test scores, which sometimes happens in this study. The equation for normalised change is given in Equation (1):

$$\hat{g} = \begin{cases} \text{drop} & \text{if } \ pre = post = 0 \ \text{ or } \ pre = post = 8 \\ \dfrac{post - pre}{8 - pre} & \text{if } \ post \geq pre \\ \dfrac{post - pre}{pre} & \text{if } \ pre > post \end{cases} \tag{1}$$

The overall Cronbach's alpha for the pre-test was 0.86, with 0.77 for the first skill, 0.70 for the second, and 0.80 for the third. The overall Cronbach's alpha for the post-test was 0.88, with 0.77 for the first skill, 0.85 for the second, and 0.72 for the third.

2.4.2. Regulation of practice behaviour metrics

Regulation of practice behaviour was measured with practice accuracy and effort, based on trace data from the ALT, for both unique problems (the first time a learner tries to solve a problem) and attempts (all times a learner tries to solve a problem, after solving problems incorrectly). Practice accuracy is defined as the percentage of problems made correctly, and practice effort is defined as the total number of problems solved.

While practice effort measured in this way does not directly reflect the mental effort required for solving each of the problems, we expect it to be a close enough indication: Each problem requires roughly the same amount of mental effort as the ALT tries to adapt the exercises to the level of the child.

2.4.3. Ability curve

After each problem-solving attempt, the ability score of the learner for the skill they are working on was updated. We call this progression of a learner's ability score over consecutive problems an ability curve. Examples of ability curves are shown in Fig. 2. The calculation for the ability score is done by the Gynzy ALT as a vital part of their adaptive mechanism. This means that the calculations are performed outside the scope of our research. For the Gynzy ALT, at the time of data collection, the assumption was made that children start with no previous knowledge of the skill, so their ability score, in the beginning, is set at 0.

From the 134 children, 402 ability curves were extracted, 3 for each child, as they all worked on 3 subskills. Some children had already worked on the first skill, resulting in incomplete ability curves. For some of those curves, there was data on how many problems the children had already solved so that earlier-made problems could be set as missing values. The ability curve was discarded if there was no data on how many problems were made before the experiment. This resulted in 48 discarded ability curves, so from the original 402, 354 ability curves were used. We cut the ability curves off at 100 problems to keep the lengths of the curves more consistent.

Table 1
Skills practised with different complexity and examples of each skill.

Skill	Example
Skill 1: Simplifying basic fractions	$\frac{4}{8} = \frac{1}{2}$
Skill 2: Simplifying mixed fractions	$\frac{12}{8} = 1\frac{1}{2}$
Skill 3: Simplifying complex fractions	$\frac{57}{38} = 1\frac{1}{2}$

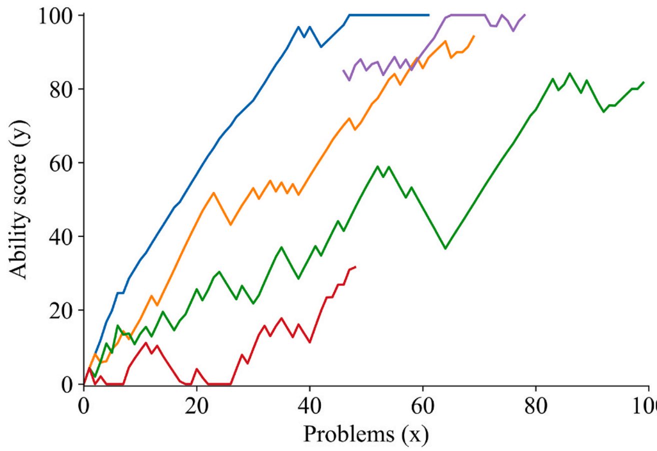

Fig. 2. Examples of ability curves. *Note:* Each line is a different ability curve. Children do not solve an equal number of problems and sometimes have solved problems before the experiment, resulting in missing values in the beginning.

2.5. Procedure

On the first day, the children completed the pre-test on all three skills in 30 min. After that, there were three days in a new week where children practised a new skill each day. First, they received instructions on the skill following a standardised protocol. Then they practised the skill for 30 min. After the three days, a day followed for repeated practice, where children could pick which skill or skills they wanted to improve on and could exercise the skills for 30 min. On the final day, they completed a post-test in 30 min.

2.6. Computational model

To cluster ability curves, we use the Dirichlet process Gaussian process (DPGP) mixture model (McDowell et al., 2018). The DPGP model is a Bayesian clustering model. The fact that this model is Bayesian means it does not provide the single most likely clustering but instead the probability distribution over all possible clusterings. For each possible clustering it gives the probability of how likely this possibility is given the trace data. Computing the probability distribution exactly is intractable. We, therefore, approximate it using Markov chain Monte Carlo (MCMC) sampling (Neal, 2000) and select one of them for further analysis, which we will further explain in section 2.6.3.

A clustering in the DPGP model is described in two parts: The first part describes which ability curve belongs to which cluster. This is represented in cluster assignment z. The second part describes the properties of each of the clusters, represented in the cluster parameters θ.

We use Bayes' rule to calculate the probability of a set of clusters given the trace data. Bayes' rule in its basic form (see Fig. 3) can be used to find the posterior probability that hypothesis H is true when evidence E is obtained. Bayes' rule states that this posterior probability is equal to the probability of E given H (known as the likelihood) times the probability of H on its own (called the prior) divided by the probability of E on its own (known as the marginal likelihood).

The probability of a clustering z with certain probabilities θ given trace data t is calculated with the right-hand formula in Fig. 3. To calculate this probability using Bayes' rule, we need the following probabilities:

- $p(z)$: How probable is it that when we cluster items, they end up together in this way? This is calculated with the Dirichlet Process prior (see section 2.6.1).
- $p(\theta)$: What are likely cluster properties, such as how smooth are curves in the cluster and how quickly do these curves rise? This is calculated with the Gaussian Process prior (see section 2.6.2).
- $p(t|z,\theta)$: What trace data do we expect to see for a certain clustering and cluster properties? This forms the likelihood (see section 2.6.3).
- $p(t)$: How probable is this trace data, given our modelling assumptions? This term is the most difficult to compute as it sums over all possible clusterings and all possible cluster probabilities. However, using MCMC, the calculation of *p(t)* is circumvented.

2.6.1. What goes where? The Dirichlet process prior

The Dirichlet process determines the prior probability of the assignment of ability curves to clusters. A common, intuitive analogy as to how the Dirichlet process prior works is the Chinese Restaurant process: In this metaphor, customers (ability curves) enter a Chinese restaurant with infinite tables (clusters) one by one and choose a table to sit at based on how many people are already sitting at each of the tables, with a chance to pick a new, empty, table. The process is influenced by hyperparameter α, which influences the probability of a new table being chosen (and thus affects the total number of clusters).

The prior probability is thus independent of the properties of the ability curve itself; instead, it assumes that the probability of assigning a curve to a cluster is determined only by the number of ability curves that have already been assigned. This means that there is a low probability that all ability curves end up in one cluster or that all ability curves end up in different clusters of size 1.

The probability of an ability curve getting assigned to a new cluster is $p(z_j = h|z_{-j}) = \frac{\alpha}{n+\alpha-1}$, with n being the number of ability curves already assigned and z_{-j} being the assignments of the other ability curves. The probability of the ability curve getting assigned to a cluster with m_h curves already assigned is $p(z_j = h|z_{-j}) = \frac{m_h}{n+\alpha-1}$.

2.6.2. What is it like there? The Gaussian process prior

The Gaussian process determines the shape of ability curves expected for any particular cluster. This shape, the prototypical curve f_h for cluster h, is characterised by two terms: the mean function, which roughly determines the expected ability curves, and the covariance function, which determines the temporal correlations in these curves

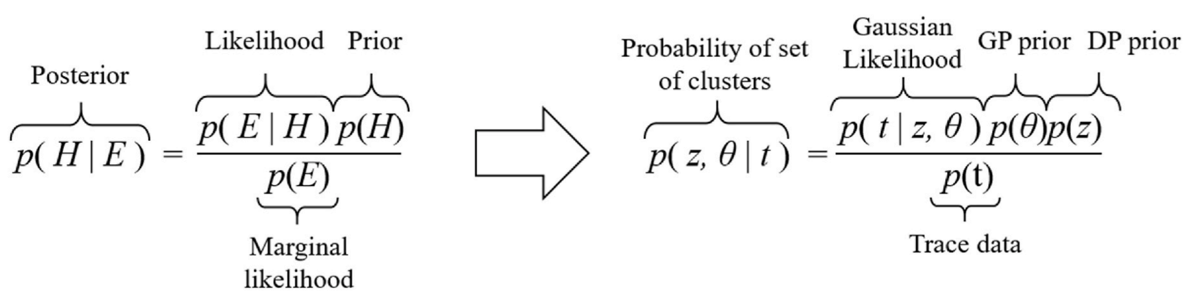

Fig. 3. Bayes' rules. *Note:* On the left is the general form of Bayes' rule, where the probability of a hypothesis H given evidence E is calculated. On the right is Bayes' rule, as implemented by the DPGP model, to calculate the probability of a clustering (z) with properties (θ) given the trace data (t).

(for example, we might expect smooth or irregularly shaped curves). Both elements have their own set of parameters, which are explained below.

2.6.2.1. Mean function. The mean function describes what we expect the trajectory of ability curves assigned to a cluster to look like. In the application of the DPGP model by McDowell et al. (2018), the mean is simply the zero function. In the application of ability curves, this would mean that we assume that the ability score of a learner at any moment in time is 0 and thus that they are (on average) not learning at all. This is generally not the case, so instead, we assume that there is a point in time when learning starts to happen and that from that point on, their ability score will first quickly increase until it slows down as their ability score reaches its maximum value. To represent these assumptions, we use a capped sigmoid function, that for every problem x, its expectation (μ) for each cluster h is given in equation (2),

$$\mu_h(x) = \max\left(0, \frac{2y_{\max}}{1+\exp\left(-\frac{x-\kappa_h}{\lambda_h}\right)} - y_{\max}\right), \tag{2}$$

where $y_{\max}$ is the maximum value the mean can take, which for this study is an ability score of 100. The parameter λ determines how fast the mean function rises, and the parameter κ determines when the function starts to rise. Examples of curves using different settings for λ and κ are given in Fig. 4.

2.6.2.2. Covariance function. The covariance function describes the covariance between two outputs of a function, $f(x)$ and $f(x')$, based on the distance between the inputs x and x'. Following McDowell et al. (2018), we use the common squared-exponential kernel as a covariance function. This covariance function is defined as:

$$k_h(x, x') = \tau_h^2 \exp\left(\frac{||x-x'||^2}{-2\ell_h^2}\right). \tag{3}$$

This covariance function has two parameters. The parameter τ determines the amount of vertical variance in a curve. A higher value for τ results in the trajectory of a curve deviating further from the mean. The parameter ℓ is known as the length scale and determines how smooth a curve is, with a higher ℓ resulting in a smoother curve. The influence of parameters τ and ℓ on possible prototypical ability curves f_h are shown in Fig. 5.

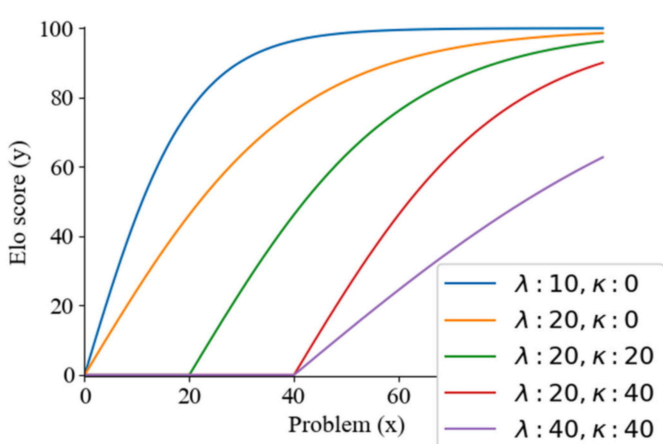

Fig. 4. Expected ability curves f_h. *Note:* The three middle lines show how κ influences the onset of the slope on the x-axis. The two outer lines show how a lower (on the left) and a higher λ (on the right) affect the slope of the curve.

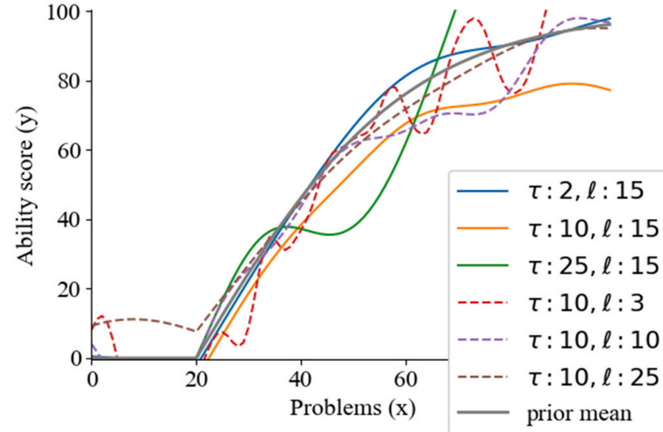

Fig. 5. Prototypical ability curves f_h. *Note:* The effect of τ is visible in the solid lines: a higher τ results in the lines being further from the mean (plotted in grey in the middle). The effect of different values for ℓ is visible in the dashed lines: a higher ℓ results in a smoother curve.

2.6.3. Gaussian likelihood

In the preceding sections, we explained how the Dirichlet process assigns curves to clusters (indicated by z) and how the Gaussian process results in prototypical curves for each cluster h (indicated by f_h). These form the prior of the DPGP model. The likelihood of the model, which describes how the latent variables of the prior result in actual observations (the ability curves), is a Gaussian distribution with mean f_h and noise variance σ^2.

The noise variance reflects to which degree observations deviate from the prototypical ability curve. With low observation noise, observations are expected to be closer to the curve of the cluster. In contrast, curves in a cluster with high observational noise have observations that deviate further from the prototypical curve. Examples of different levels of observation noise are shown in Fig. 6.

2.6.4. How do we get the optimal set of clusters?

With the priors and the likelihood defined, the application of Bayes' rule results in the posterior distribution. However, computing the distribution over all possible clusterings is computationally intractable. Instead, the distribution is approximated using MCMC sampling methods. The details of the sampling procedure are provided in

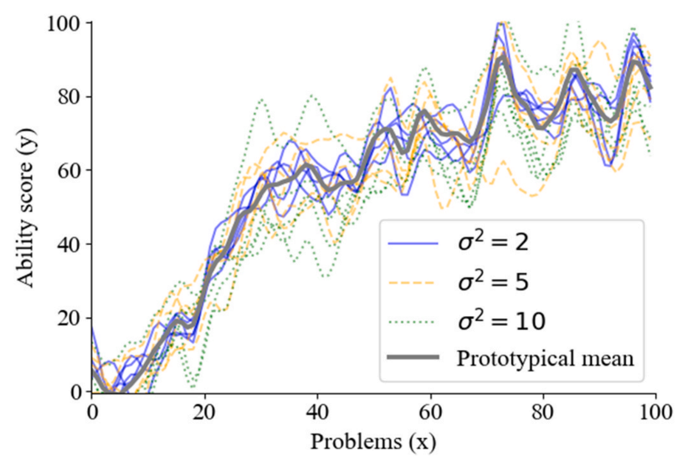

Fig. 6. Effect of observation noise. *Note:* Effect of observation noise on possible ability curves assigned to a cluster with a prototypical ability curve (plotted thickened and grey here). The effect of observation noise σ^2 is apparent in how with higher observation noise (solid blue for low, dashed yellow for medium and dotted green for high noise), the possible ability curves can get further away from the prototypical curve.

Appendix B.4.

The sampling methods produce a probability distribution over possible clusterings. From this distribution we select the clustering with the highest posterior probability for subsequent analysis. In this clustering, some clusters have only a small number of ability curves assigned. To improve interpretability and statistical power, we set the minimum number of ability curves assigned to a cluster to 10. The ability curves from clusters with a smaller size are then reassigned to the cluster with the next highest probability of assignment. This process started from the cluster with the least amount of ability curves assigned and ended when all clusters had more than 10 clusters assigned. After this process, the cluster parameters were resampled with MCMC sampling, which is further explained in Appendix B.4.

2.7. Analyses

2.7.1. Co-occurrence matrix

Cluster labels change during the sampling of the DPGP model, an issue known as 'label switching'. This makes it impossible to directly compare how often an ability curve is assigned to a particular cluster. Instead, we infer the probability of two ability curves being in the same cluster using a co-occurrence matrix. This matrix $\mathbf{C}$ is of size $m \times m$, where m is the number of ability curves and every $\mathbf{C}_{i,j}$ represents the probability that ability curve i is in the same cluster as ability curve j.

2.7.2. Analysing differences between clusters related to learning behaviour measures

Linear mixed models were used to analyse the data using the lmerTest package version 3.1–3 in R (Kuznetsova et al., 2017). This analysis was used to investigate the effect of the practice accuracy, practice effort, and learning metrics on cluster assignment while controlling for intraindividual and intraskill variability across the assignments.

Each model has a fixed effect for cluster and a random effect for child and for skill with a random intercept. The models were fit using REML (Restricted Maximum Likelihood), with the measure as a dependent variable, a random effect for learner and subskill, and a fixed effect for cluster using the lmerTest package in R (Kuznetsova et al., 2017). The significance of the parameters was determined through t-tests using Satterthwaite's method.

We report the conditional R^2 statistic for each model using the MuMIn package version 1.47.1 in R (Barton & Barton, 2022). This statistic can be interpreted as the variance explained by the entire model (Nakagawa & Schielzeth, 2013).

Assumptions of linear mixed models were verified by visual inspection of residual plots for deviations of normality and absence of homoscedasticity. While the learning gain measure and the practice effort for all attempts measure were not exactly normally distributed, this is not a problem, as linear mixed models are robust to violations in normality (Schielzeth et al., 2020). To inspect if there were influential data points, we inspected Cook's distance plots. Since the data points were all below 1 and there are no values substantially larger than the rest, we do not see them as influential (Fox, 2019).

To test the difference between the clusters, we used a post-hoc pairwise comparison of estimated means with Kenward-Roger adjustment of the degree of freedoms and Tukey correction for multiple comparisons using the emmeans package version 1.8.1-1in R (Lenth, 2022).

3. Results

3.1. Optimal clustering

The clustering resulted in 9 clusters of non-trivial size. The clusters are sorted on the average final ability score of the ability curves belonging to the clusters and subsequently labelled alphabetically based. The number of ability curves assigned to the clusters differs per cluster, with a few large clusters and a few clusters of minimal size: the *fast-masters* cluster (C) is the largest, with 70 ability curves assigned, followed by the *slow-masters* (E, 68) and *plateauing* cluster (F, 67). The smallest clusters are the *already-proficient* (B, 12), *masters* (D, 14), and *proceeding already-proficient* (A, 18) clusters. The ability curves per cluster and their mean curves are shown in Fig. 7.

3.2. Visualising cluster uncertainty with a co-occurrence matrix

To inspect the cluster uncertainty over the clusters resulting from the iteration with the optimal posterior probability, the assignments of the ability curves over all iterations resulted in the co-occurrence matrix shown in Fig. 8. The higher the probability of two ability curves being in the same cluster, the more yellow their co-occurrence (This results in a yellow diagonal where x_j equals $y_{j'}$, as ability curves are always in a cluster with themselves). The y-axis on the right indicates the cluster labels from the final selected clustering.

If all ability curves in the current clusters were always assigned to the same cluster, a diagonal with yellow squares with the size of the clusters would appear. However, the yellow squares will become more "jagged" as the cluster assignment uncertainty increases. While this jaggedness in Fig. 8 is often happening on the borders of the clusters, this does not extend much further, meaning that while it can be likely for some ability curves to belong to a cluster close by, it is not likely that they belong to the other clusters.

Ability curves assigned to cluster B (*already-proficient*) are frequently clustered together with ability curves assigned to clusters A and C (*proceeding already-proficient* and *fast-mastering*). However, ability curves assigned to A are rarely clustered together with ability curves assigned to cluster C. A similar story can be told for clusters B, C, and D (*mastering*).

Reassigned ability curves that were part of singular clusters are also distinguishable in the co-occurrence matrix as rows (and columns) that are mostly blue, meaning that they are clearly different from the other curves and thus unlikely to be assigned to the same cluster. Cluster I (*stagnators*) has the most of these reassigned singular clusters.

3.3. Learning metrics and clusters

3.3.1. Descriptive

The average pre-test score per child per skill was 3.01 ($sd = 2.32$), and the average post-test score per child per skill was 5.17 ($sd = 2.41$), resulting in an average learning gain of 2.24 ($sd = 2.24$). The learning metrics from the children whose ability curve was assigned to the different clusters can be found in Table 2.

3.3.2. Pre-test

Differences in children's pre-test scores on skills were compared for the clusters using a linear mixed model, see Fig. 9. Pre-test scores differed significantly overall between clusters, see Table 3.

The proceeding already-proficient (A) and the *already-proficient* (B) clusters were associated with higher level pre-test scores as they had significantly higher pre-test scores than the other clusters. The *fast-mastering* (*C*), *mastering* (D), and *slow-mastering* (E) were associated with medium-level pre-test scores: Cluster C had significantly higher pre-test than the scores of clusters F, G, H, and I. Cluster D scored higher than cluster I. Cluster E scored higher than clusters G, H, and I. The *rising strugglers* (G), *brief-risen strugglers* (H), and *stagnators* (I) clusters were associated with low-level pre-test scores and did not significantly differ from each other in pre-test scores. The *plateauing* (F) cluster bordered between medium and low-level pre-test scores, as it had only significantly different pre-test scores, lower than clusters A and B and higher than cluster I. See Table A1 for the post hoc comparison result details.

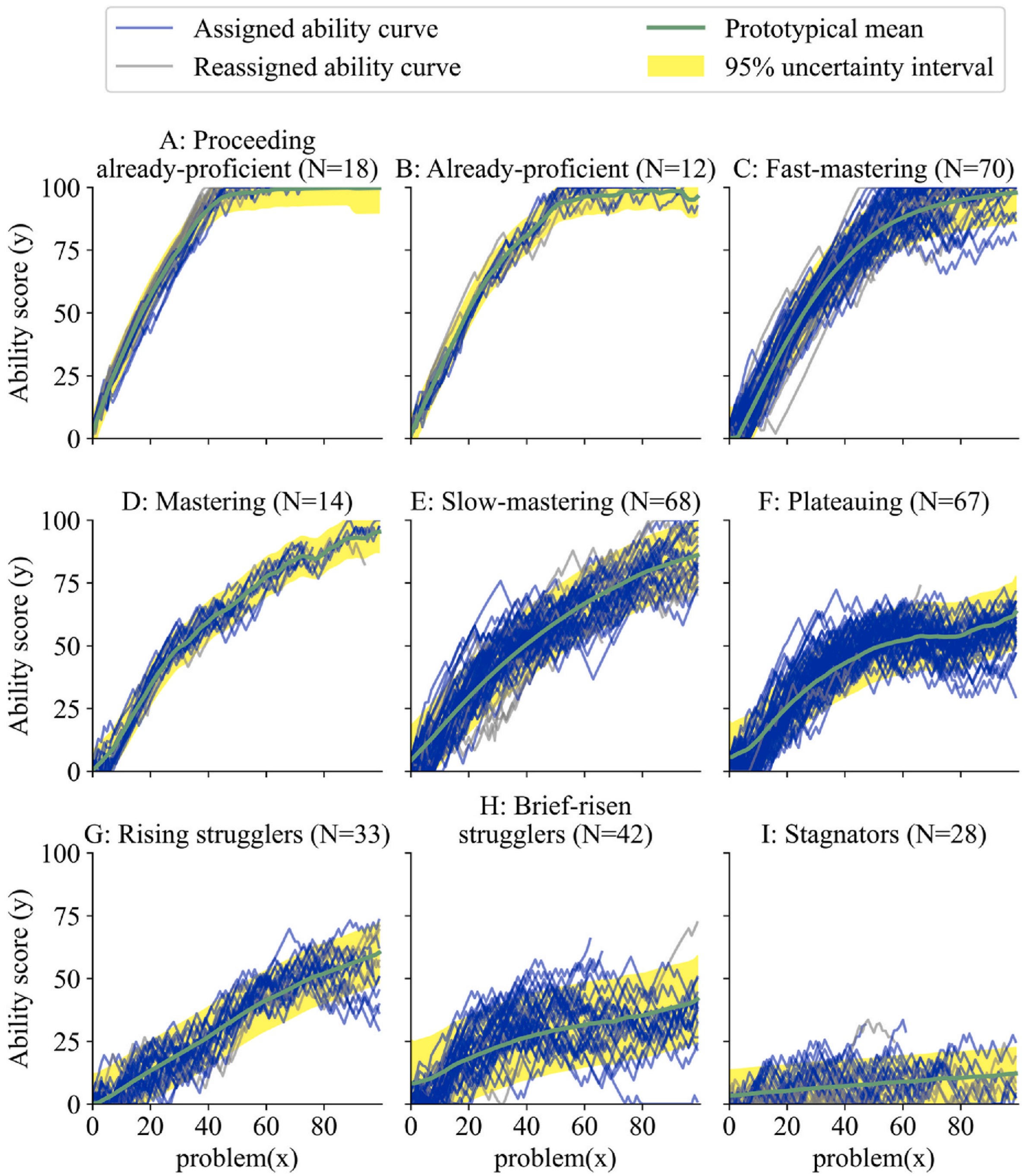

Fig. 7. Optimal clustering result. *Note:* There are 9 different non-trivial-sized clusters. The blue lines show the ability curves directly assigned to the cluster, whereas the grey lines are from ability curves reassigned after removing clusters of a smaller size. The green line shows the prototypical mean of each cluster, and the yellow interval shows the 95% uncertainty interval.

3.3.3. Post-test

Differences in post-test scores on skills were compared for the clusters using a linear mixed model, see Fig. 9. Post-test scores differed significantly between clusters; see Table 3.

Clusters A (*proceeding already-proficient*) to E (*slow-mastering*) were associated with high-level post-test scores as they had significantly higher scores than clusters F (*plateauing*) to I (*stagnators*). There were no significant differences in post-test scores between the clusters except that cluster C (*fast-mastering*) had a significantly higher post-test score than cluster E (*slow-mastering*).

Clusters F to I were associated with lower post-test scores as all their scores were significantly lower than the scores of clusters A to E. Between clusters F to I the post-test scores were not significantly different except for clusters F and I, where cluster F had a significantly higher post-test score than cluster I. See table A2 for the post hoc comparison result details.

3.3.4. Learning gain

Differences in learning gain on skills between pre-test and post-test scores were compared for the clusters using a linear mixed model, see Fig. 9. Overall, the learning gain differed significantly between clusters (see Table 3).

Cluster C (*fast-mastering*) had significantly higher learning gain than clusters E (*slow-mastering*) to I (*stagnators*), and cluster D (*mastering*) had

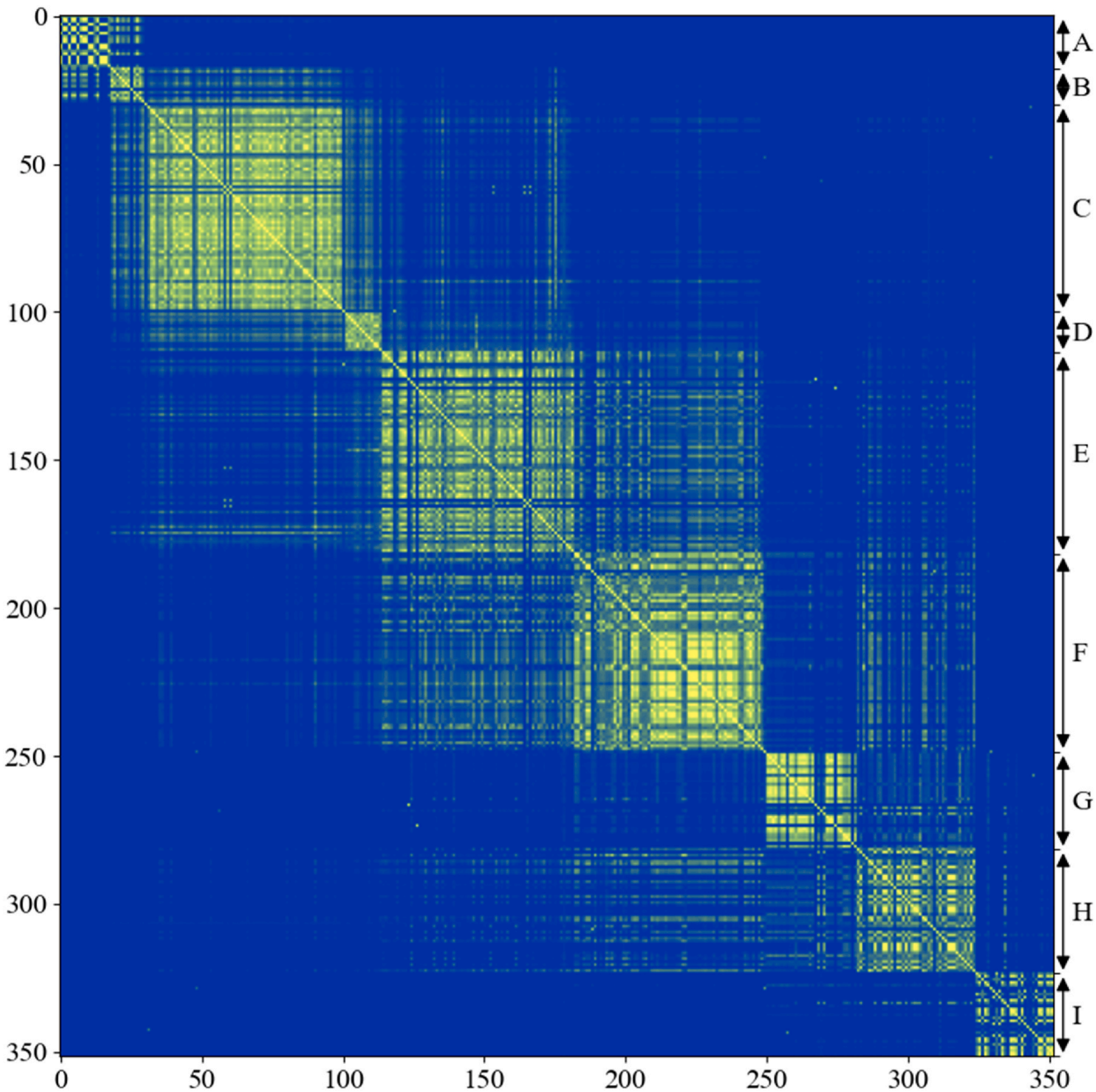

Fig. 8. Co-occurrence matrix of ability curves. *Note:* Blue indicates none to almost no co-occurrences, whereas yellow indicates that curves are always in the same cluster. The x-axis and the left y-axis indicate the indices of the ability curves, and the right y-axis indicates the boundaries of the final selected clusters. The clear diagonal block pattern illustrates that most clusters are clearly distinct.

Table 2
Descriptive of learning metrics for final clusters.

Cluster	Pre-test		Post-test		Relative learning gain	
	M	SD	M	SD	M	SD
A: proceeding already-proficient	7.17	0.86	7.80	0.56	0.67	1.18
B: already-proficient	6.42	1.73	7.42	0.79	1.00	1.95
C: fast-mastering	3.91	2.16	6.94	1.20	3.00	1.99
D: mastering	3.14	1.61	6.50	1.17	3.50	1.57
E: slow-mastering	3.20	1.76	5.53	2.06	2.40	2.42
F: plateauing	2.33	1.86	4.47	2.27	1.98	2.42
G: rising strugglers	1.45	1.55	4.07	2.12	2.62	2.34
H: brief-risen strugglers	1.78	1.71	3.66	2.23	1.90	2.04
I: stagnators	0.83	1.01	2.50	2.12	1.67	2.12
All	3.01	2.32	5.17	2.41	2.24	2.24

significantly higher learning gain than cluster I. See Table A3 for the post hoc comparison result details.

3.4. Practice accuracy and effort and clusters

3.4.1. Descriptive

The average practice accuracy on attempts per child per skill was 82.82 (sd = 43.55). The practice accuracy on unique problems is 53.92 (sd = 19.84). For the practice effort measure, the average correct percentage of attempts per child per skill was 0.63 (sd = 0.17). The practice effort average for the unique problems is 0.61 (sd = 0.18). The measures of practice accuracy and effort from the children working on the skill belonging to the ability curve in the clusters can be found in Table 4.

3.4.2. Practice effort measures

Differences in the number of attempted and unique problems solved were compared for the clusters using a linear mixed model, see Fig. 10. The total number of attempts differed significantly between Clusters, see Table 5. The *proceeding already-proficient* (A) cluster had a significantly

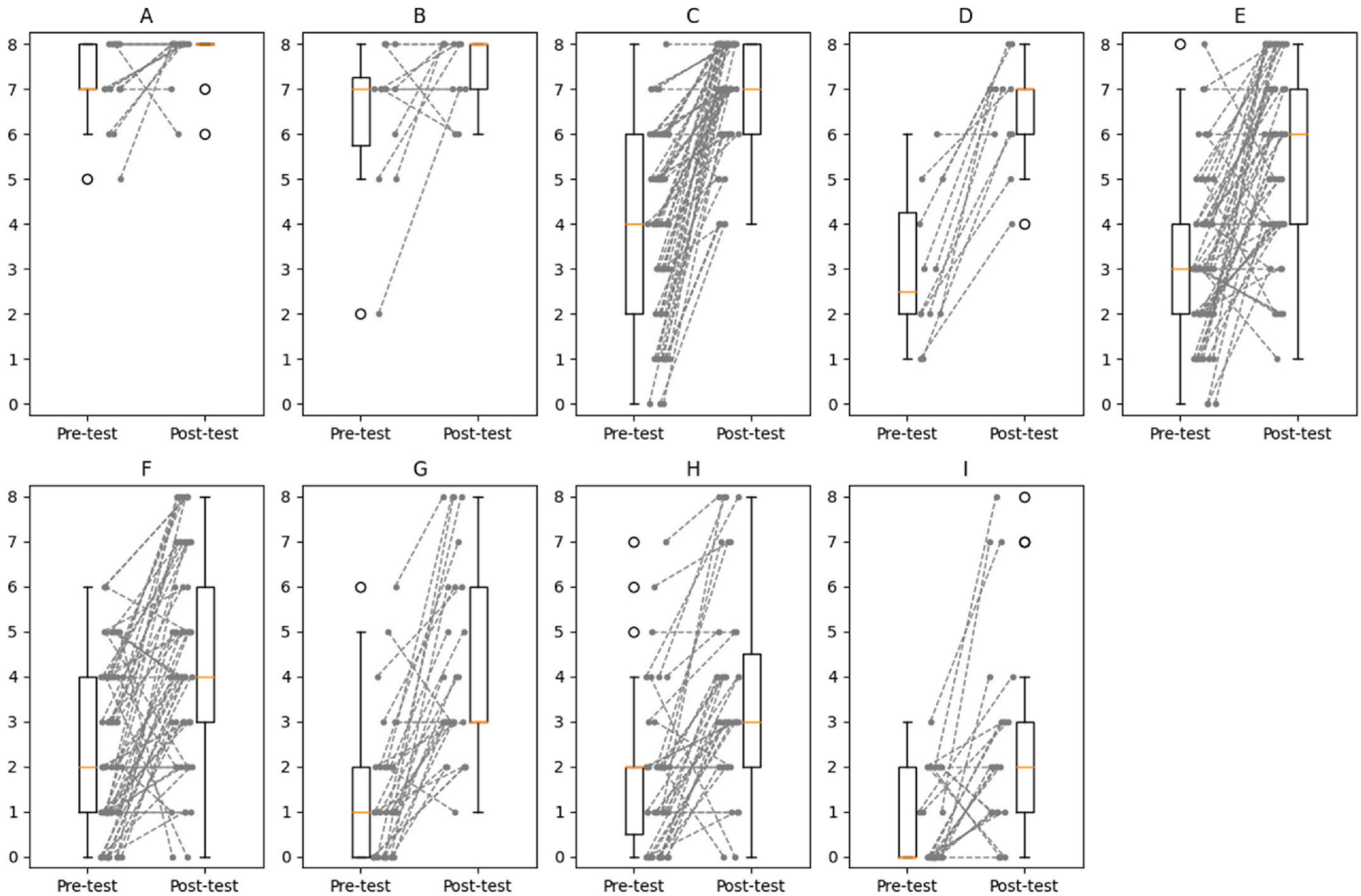

Fig. 9. Differences in learning metrics between clusters. *Note:* Both pre-test and post-test are plotted in boxplots. The learning gain is plotted as grey dotted lines between pre-test and post-test.

Table 3
Mixed linear model for pre- and post-test and learning gain.

Variable	Pre-test		Post-test		Learning gain	
Fixed part	B	SE	β	SE	β	SE
(Intercept) A: proceeding already-proficient	6.197***	0.583	6.737***	0.734	0.654	0.153
B: already-proficient	−0.366	0.599	−0.001	0.633	−0.164	0.175
C: fast–mastering	−2.545***	0.439	−0.127	0.486	0.062	0.137
D: mastering	−2.846***	0.590	−0.157	0.656	0.045	0.172
E: slow-mastering	−2.974***	0.451	−1.185*	0.500	−0.190	0.139
F: plateauing	−3.826***	0.471	−2.243***	0.517	−0.315	0.143
G: rising strugglers	−4.462***	0.517	−2.595***	0.563	−0.299	0.153
H: brief-risen strugglers	−4.395***	0.499	−2.97***	0.547	−0.360	0.149
I: stagnators	−5.269***	0.554	−3.953***	0.584	−0.493	0.161
Random part	σ	sd	Σ	Sd	σ	Sd
Intercept Skill	0.513	0.716	0.975	0.987	0.020	0.140
Intercept Child	0.710	0.843	0.885	0.941	0.041	0.203
Residual	2.126	1.458	2.128	1.459	0.128	0.358

Note: $N_{pre\text{-}test} = 336$, $N_{post\text{-}test} = 327$, $N_{learning\ gain} = 311$, $N_{children} = 134$, $N_{skills} = 3$, $R^2_{pre\text{-}test} = 0.57$, $R^2_{post\text{-}test} = 0.62$, $R^2_{learning\ gain} = 0.41$. More details are given in Table A8 for pre-test, A9 for post-test, and A10 for learning gain.
$^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

lower practice effort for attempts than the *plateauing* (F) cluster and the *brief-risen strugglers* (H) cluster. See Table A4 for all post-hoc pairwise comparison results.

There was no significant difference between clusters for unique problems solved. See Table 5.

3.4.3. Practice accuracy measures

Differences in practice accuracy on attempts were compared for the clusters using a linear mixed model, see Fig. 11. There were significant differences in practice accuracy for all clusters; see Table 6. The practice accuracy of attempts was significantly different between almost every cluster and was reducing in a downward trend. Clusters without significant differences in practice accuracy for attempts have been grouped in the same level of practice accuracy: Clusters A (*proceeding already-proficient*) and B (*already-proficient*) had a high level of practice accuracy. Clusters C (*fast-mastering*), D (*mastering*), and E (*slow-mastering*) had a semi-high level of practice accuracy. Cluster F (*plateauing*) had a medium level of practice accuracy. Clusters G (*rising strugglers*) and H (*brief-

Table 4
Descriptive of practice accuracy and effort measures for final clusters.

Cluster	Practice effort All		Practice accuracy All		Practice effort Unique		Practice accuracy Unique	
	M	SD	M	SD	M	SD	M	SD
A	61.67	13.49	0.93	0.03	51.50	9.41	0.93	0.04
B	68.08	26.22	0.85	0.04	52.50	15.03	0.85	0.04
C	72.04	26.84	0.77	0.06	54.80	15.12	0.77	0.07
D	75.57	27.65	0.72	0.03	57.29	16.18	0.72	0.05
E	78.74	33.49	0.66	0.05	55.65	19.92	0.65	0.06
F	96.76	48.33	0.58	0.06	58.51	21.19	0.53	0.09
G	91.58	65.60	0.48	0.14	52.52	29.10	0.43	0.14
H	96.51	58.49	0.50	0.08	52.53	22.17	0.47	0.10
I	78.96	39.93	0.34	0.09	40.82	13.12	0.31	0.09
All	82.82	43.55	0.63	0.17	53.92	19.84	0.61	0.18

Note: see i.a. Fig. 7 for the names corresponding to the clusters' labels.

risen strugglers) had a low level of practice accuracy. Finally, cluster I (*stagnators*) had a very low level of practice accuracy. See Table A.5 for all *t*-test results.

Differences in practice accuracy on unique problems were compared for clusters, see Fig. 11. The difference in practice accuracy between attempts and distinct problems was minimal, resulting in the same significant differences in practice accuracy between clusters for both attempts and distinct problems. See Table A6 for all post-hoc pairwise comparison results.

3.5. Cluster interpretation

This exploratory study examined how the clusters found were associated with learning metrics and metrics indicating regulation of practice behavior. The study investigated how data-driven clusters of ability curves can give us insight into how children regulate practice behaviour during learning. We found nine clusters with clear interactions between the clusters and the learning metrics. These clusters of ability curves are associated differently with prior knowledge, post-knowledge, and learning gains, as well as with practice accuracy and, to a lesser extent, with practice effort during learning. Note that each ability curve corresponds to the development of a child on one skill, and therefore, children may have ability curves of different skills in different clusters. Table 7 shows a brief summary of the support needs based on cluster properties.

The results showed two types of already-proficient clusters: the *proceeding already-proficient* and the *already-proficient*. These clusters were associated with high prior and post-knowledge compared to the other clusters, and at the same time, they were associated with the highest practice accuracy. This seems to indicate that the children were already proficient in this skill to a large extent before the lessons, and the high practice accuracy indicates they had little trouble regulating their learning. The *proceeding already-proficient* cluster showed lower practice effort than the *plateauing* and the *brief-risen strugglers* clusters, indicating that the children effectively regulated their practice effort and did not over-practice a skill they were already proficient in. In comparison to other research in a comparable context (Molenaar et al., 2021), both already-proficient clusters compare to the immediate drop moment-by-moment learning curves (mbmlc), which also showed high prior knowledge and high practice accuracy. Because children in this cluster have high prior and posterior knowledge combined with high practice accuracy, they seem to need minimal support concerning self-regulated learning for this skill.

Three types of master clusters were found: the *fast mastering*, the *mastering*, and the *slow-mastering*. These clusters were associated with medium prior knowledge and high post knowledge, a higher learning gain than other clusters, and a relatively high practice accuracy. The learning metrics indicated that children whose skill is assigned to these clusters already had some initial knowledge of this skill and managed to master the skill by practising. The relatively high practice accuracy indicated that the children were able to solve most problems correctly or were able to adjust their practice effort after errors and adjust their strategy to consequently uphold practice accuracy thereafter. The medium prior and post-knowledge, combined with the high level of practice accuracy, resembles the immediate peak from the mbmlc (Molenaar et al., 2021). Since the *fast mastering* and *mastering* clusters are able to reach the maximum ability score, they seem to need minimal SRL support for this skill. The *slow-mastering* cluster does not always reach the maximum ability score, indicating that they might need moderate SRL

Table 5
Mixed linear regression for practice effort measures by final clusters.

Variable	Practice effort all		Practice effort unique	
Fixed part	β	SE	β	SE
(Intercept) A: proceeding already-proficient	54.291***	10.766	48.438***	5.128
B: already-proficient	12.951	14.488	2.426	6.228
C: fast–mastering	17.854	10.648	5.693	4.616
D: mastering	23.201	14.315	8.081	6.203
E: slow-mastering	27.126*	10.988	9.373	4.810
F: plateauing	38.526***	11.444	8.862	5.072
G: rising strugglers	36.263**	12.521	4.906	5.547
H: brief-risen strugglers	43.096***	12.202	5.813	5.430
I: stagnators	23.482	13.243	−6.355	5.915
Random part	σ	Sd	σ	sd
Intercept Skill	38.4	6.2	17.91	4.23
Intercept Child	644.6	25.39	172.5	13.13
Residual	1172.69	34.24	205.17	14.35

Note: $N_{attempts} = 354$, $N_{unique\_problems} = 354$, $N_{children} = 134$, $N_{skills} = 3$, $R^2_{Effort\_all} = 0.41$, $R^2_{Effort\_unique} = 0.50$. More details are given in tables A11 for practice effort on all problems and A12 for practice effort on unique problems.
$^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

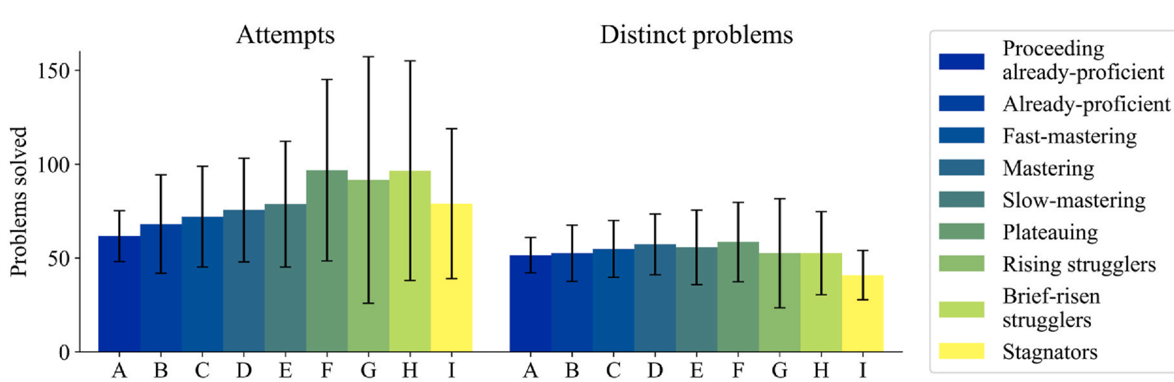

Fig. 10. Differences in practice effort measures between clusters.

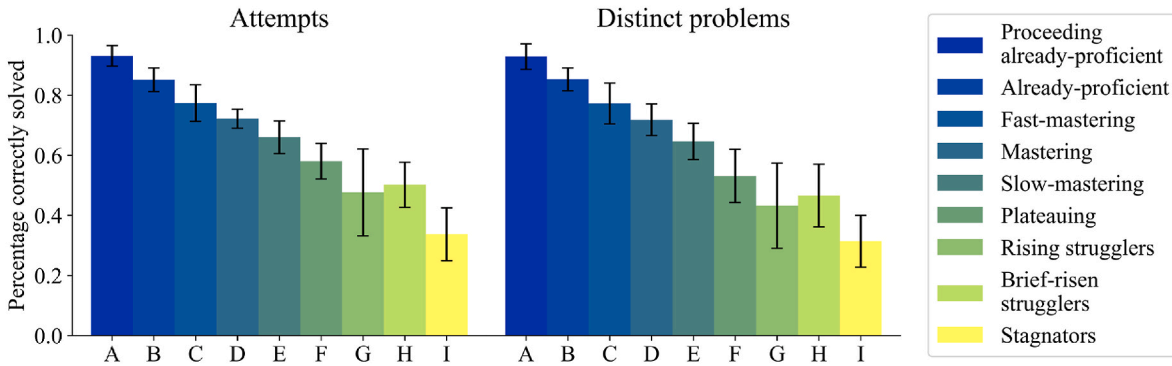

Fig. 11. Differences in practice accuracy measures between clusters.

Table 6
Mixed linear regression for practice accuracy measures by final clusters.

Variable	Practice accuracy all		Practice accuracy unique	
Fixed effects	B	SE	β	SE
(Intercept) A: proceeding already-proficient	0.905***	0.027	0.896***	0.034
B: already-proficient	−0.072**	0.025	−0.067*	0.028
C: fast–mastering	−0.144***	0.018	−0.142***	0.020
D: mastering	−0.192***	0.024	−0.189***	0.027
E: slow-mastering	−0.248***	0.018	−0.253***	0.020
F: plateauing	−0.31***	0.019	−0.342***	0.021
G: rising strugglers	−0.415***	0.021	−0.441***	0.023
H: brief-risen strugglers	−0.395***	0.020	−0.418***	0.022
I: stagnators	−0.55***	0.021	−0.555***	0.024
Random effects	Σ	SD	σ	SD
Intercept Skill	0.0014	0.037	0.0024	0.049
Intercept Child	0.00059	0.024	0.0010	0.032
Residual	0.0039	0.063	0.0048	0.069

Note: $N_{atempts} = 354$, $N_{unique} = 354$, $N_{children} = 134$, $N_{skills} = 3$, $R^2_{Practice\_accuracy\_all} = 0.84$, $R^2_{Practice\_accuracy\_unique} = 0.84$. More details are given in table A13 for practice accuracy on all problems and A14 for practice accuracy on unique problems.
*$p < .05$. **$p < .01$. ***$p < .001$.

Table 7
Interpretation of SRL support needs.

Cluster	Description of shape	Practice accuracy	Practice effort	SRL support needs
A: Proceeding already-proficient	Reaches maximum very quickly	Very high	Medium/ Low	Minimal
B: already-proficient	Reaches maximum very quickly	Very high	Medium	Minimal
C: Fast mastering	Reaches maximum quickly	High	Medium	Minimal
D: Mastering	Reaches maximum	High	Medium	Minimal
E: Slow mastering	Rises towards maximum	High	Medium	Moderate
F: Plateauing	Plateaus halfway after rising	Medium	Medium/ High	Moderate
G: Rising strugglers	Rises not further than half-way	Low	Medium	Substantial
H: Brief-risen strugglers	Plateaus very early	Low	Medium/ High	Substantial
I: Stagnators	Hardly departs from minimum	Very low	Medium	Extensive

support for this skill.

One *plateauing* cluster was found: Children in this cluster showed medium to low prior knowledge. The cluster had lower prior knowledge than the *fast-mastering* cluster but also significantly higher prior knowledge than the *stagnators* cluster. It was associated with low post-knowledge but had a significantly higher posterior knowledge than the *stagnators* cluster. It was also associated with a medium practice accuracy compared to other clusters and with a higher practice effort than the *proceeding already-proficient* cluster. This cluster started rising at the same pace as the *already-proficient* clusters and the *mastering* clusters, indicating that the children were initially learning. However, after approximately 50 problems, this learning plateaus, and children started to show more fluctuations, indicating that practice accuracy was no longer upheld. This indicated that children could not adjust their practice effort nor change their strategies effectively to support further learning. This aligns with the finding that these learners showed more practice effort than the *proceeding already-proficient* cluster, which potentially indicates that children were solving more problems and seemed aware of reduced progress but could not find a strategy to control their learning effectively.

Because of its lower post knowledge, the *plateauing* cluster is similar to the close multiple spikes category of the mbmlc (Molenaar et al., 2021). This seems to indicate that children showing this pattern when learning a skill need moderate support to enact effective SRL.

Two types of strugglers clusters were found: the *rising strugglers* and the *brief-risen strugglers* clusters. They were associated with low prior and post-knowledge and low practice accuracy. Additionally, the *brief-risen strugglers* cluster showed higher practice effort than the *proceeding already-proficient* cluster solving more problems. This indicates that children were practising the skill assigned to this cluster and may have been aware of the need to practise more but appear to be unable to uphold practice accuracy. These clusters are comparable to the close multiple spikes from the mbmlc (Molenaar et al., 2021). They seem to need substantial support for SRL while learning this skill.

Finally, an overall low-scoring cluster was found: the *stagnators* cluster. This cluster was associated with low prior and post-knowledge and very low practice accuracy. These measures indicated that the children showing this pattern hardly learned this skill and had trouble maintaining practice accuracy from the start. This resembles earlier found separate multiple peaks from the mbmlc (Molenaar et al., 2021). With their very low practice accuracy level, these children seem to need extensive regulation support in the form of advanced system regulation support.

4. Discussion

The different types of the resulting clusters are related to learning metrics in specific ways, which provides insights into how temporal trajectories in the form of ability curves are associated with learning and regulation of practice behaviour, as earlier proposed by Winne and

Baker (2013). Clearly, the DPGP clustering method also provides advanced insights into how different temporal trajectories are associated with each other. This gives us an automated way to detect the clusters. Previous work on the moment-by-moment learning curves (Baker et al., 2013) was used for visual clustering and rule-based automation (Molenaar et al., 2021). However, it did not provide detailed insights into the comparability of the clusters.

The clusters we found can be linked to the concept of a hybrid human-AI regulation system (Molenaar, 2022), which features a forward adaptivity to adjust the level of support to the need of the individual child. In this system, the degree of SRL support from the system is adjusted to the child's needs based on their detected ability to regulate their learning in the context of ALTs. At the same time, children are provided with advanced insights into their own SRL process through dashboards that show personalized visualisations. Four different degrees of hybrid regulation were proposed previously (Molenaar et al., 2019): The self-regulated learning degree, in which the AI only supports the child with dashboard information, mirroring how children regulate their learning to support understanding. This support seems appropriate for the *proceeding already-proficient* and *already-proficient* clusters. In the shared-regulation degree, the AI monitors the regulation and provides control action to the child. The dashboard provides insights into the learning trajectory in a simplified matter to scaffold the children's regulation. This level of support seems appropriate for the *fast-mastering, mastering, and slow-mastering* clusters as they have few SRL problems and could increase their learning efficiency by increasing the difficulty of the problems faster (Horvers et al., Submitted). In the co-regulation degree, the AI monitors and controls learners' regulation and automatically adjusts the difficulty level of problems. In a dashboard, children are shown how the system supports their learning to create awareness of its interventions. This level of support seems appropriate for the *plateauing, rising strugglers* and *brief-risen strugglers* clusters, as they have trouble regulating their practice accuracy and changing into different strategies themselves. The final degree is the AI regulation degree. Here the AI monitors and controls the child's learning and extensively adjusts the problems selected for the children. The child is made aware of the support given but only at a basic level to prevent cognitive overload. This level of support seems appropriate for the *stagnators* cluster as they are not making progress in the current situation.

In contrast to other approaches to clustering in SRL, that transform the trace data into meaningful action sequences (Fincham et al., 2019; Matcha et al., 2019; Saint et al., 2020; Uzir et al., 2020), we have directly clustered the temporal pattern of the trace data, namely the ability curve. Our approach results in clusters that are not as directly related to SRL processes as the other approach used in more open learning environments. Looking at the COPES model (Winne & Hadwin, 1998), the trace data from the ALT are directly reflecting the performance of the children, not their micro-level SRL processes, such as planning or monitoring. Yet, while the trace data coming from ALTs are not as rich, valuable information can be extracted using the temporal patterns in the data: using the clustering approach shown in this paper and relating learning metrics to those clusters to understand them in more detail, makes it possible to extract indications of whether SRL support is needed. Thus we have shown that, combined with SRL measures, there is more information in the ability curves than just an estimation of the children's cognitive abilities in a particular skill. The pattern of these curves can also be seen as an indicator of self-regulated learning and potential failure thereof. However, while we know that regulation can be optimized for particular clusters, it is still unknown which regulation problems these children have in this particular situation. Therefore, the current study is a first step towards understanding the value of temporal trajectories and ability curves for regulation. Future research is needed to examine more details of precise individual SRL support needs and the appropriate SRL support.

In future research, we also aim to address how children's cluster membership can be predicted early during learning. The Bayesian nature of the clustering model allows for predicting what cluster a child fits in best. At the same time, the ability curve is only used in early development and is not yet completed. By finding the best-fitting cluster for each child, we aim to make predictions to infer the needed degree of SRL support. This new approach to predict temporal patterns and related SRL support needs during learning is a first step towards developing hybrid human-AI regulation. This will provide advances for children to practise self-regulated learning in a forward adaptive matter.

The Bayesian model used in our study may be improved to better explain our ability curve observations by replacing certain assumptions. For instance, we have now used the popular squared-exponential covariance function in our prior on the cluster-specific prototypical ability curves. This assumption dictates smooth curves, while it may be possible that these curves are actually more irregular. In future work, we consider whether alternative covariance function choices, such as an exponential, rational-quadratic, or Ornstein-Uhlenbeck function (Rasmussen & Williams, 2006), improve the robustness of our clustering approach.

A limitation of our approach is that the ability algorithm needs to process several solved problems before it can accurately determine a child's ability level. One solution to incorporate the uncertainty of the correct ability score is to use a flexible noise observation parameter that decreases once children have solved more problems. Another solution is to implement an alternative to estimate the ability level of a child and the problems, for example, through knowledge tracing (Pardos & Heffernan, 2011; Piech et al., 2015; Zhang et al., 2017). Another limitation is that the model assumes independence between the ability curves, while, in fact, one child produces multiple ability curves, one for each skill it has worked on. This creates a correlation between the curves of a particular child, which is currently not reflected in our model. However, we notice that the curves from the same child may fall in different clusters, showing that different skill topics can lead to different behaviour. This indicates that the independence assumption is actually preferable.

To conclude, this study showed that it is possible to cluster temporal patterns derived from learner models from adaptive learning technologies in the form of ability curves. The resulting nine clusters are uniquely associated with knowledge development and regulation of practice behaviour metrics, indicating different support needs for different clusters. In the future, we will use these clusters to assess the children's current SRL support needs by predicting their cluster, making it possible to develop a system capable of supporting children's SRL development in a forward adaptive manner.

CRediT author statement

Rick Dijkstra: Methodology, Software, Validation, Formal analysis, Data curation, Writing – Original draft, Writing – Review & Editing. **Eliane Segers:** Writing – Review & Editing, Supervision. **Max Hinne:** Methodology, Software, Writing – Review & Editing, Supervision. **Inge Molenaar:** Conceptualization, Writing – Review & Editing, Supervision, Project administration, Funding acquisition.

Funding

This work was supported by the European Research Council [grant number 948786].

Data availability

The authors do not have permission to share data.

Appendix A

A.1 Post hoc pairwise comparison tables

Below are the tables for the post-hoc t-tests of the learning behaviour measures for the clusters. The bottom-left part of the table is the degrees of freedom, and the upper-right part is the t-score. Alpha values for significance are indicated with asterisks. Table A7 lists the matching names for the cluster labels.

Table A1
post-hoc comparisons for pre-test score per cluster

Comparison							
Condition		Condition	Mean Difference	SE	df	t	p_{Tukey}
Cluster A	-	Cluster B	0.366	0.602	303.0	0.608	1.000
Cluster A	-	Cluster C	2.545	0.442	319.7	5.760	<.001
Cluster A	-	Cluster D	2.846	0.594	317.7	4.791	<.001
Cluster A	-	Cluster E	2.974	0.455	325.8	6.542	<.001
Cluster A	-	Cluster F	3.826	0.476	321.1	8.041	<.001
Cluster A	-	Cluster G	4.462	0.522	321.9	8.555	<.001
Cluster A	-	Cluster H	4.395	0.504	313.5	8.723	<.001
Cluster A	-	Cluster I	5.269	0.559	306.1	9.427	<.001
Cluster B	-	Cluster C	2.178	0.497	284.5	4.384	<.001
Cluster B	-	Cluster D	2.479	0.641	303.7	3.870	.004
Cluster B	-	Cluster E	2.608	0.510	304.0	5.118	<.001
Cluster B	-	Cluster F	3.459	0.520	312.9	6.648	<.001
Cluster B	-	Cluster G	4.095	0.565	319.0	7.247	<.001
Cluster B	-	Cluster H	4.029	0.547	321.5	7.365	<.001
Cluster B	-	Cluster I	4.902	0.600	326.1	8.173	<.001
Cluster C	-	Cluster D	0.301	0.473	298.7	0.637	.999
Cluster C	-	Cluster E	0.430	0.286	316.6	1.503	.854
Cluster C	-	Cluster F	1.281	0.308	326.6	4.155	.001
Cluster C	-	Cluster G	1.917	0.376	326.9	5.097	<.001
Cluster C	-	Cluster H	1.851	0.353	315.9	5.242	<.001
Cluster C	-	Cluster I	2.724	0.430	304.6	6.341	<.001
Cluster D	-	Cluster E	0.129	0.477	297.3	0.270	1.000
Cluster D	-	Cluster F	0.980	0.486	294.1	2.015	.534
Cluster D	-	Cluster G	1.616	0.535	308.7	3.022	.067
Cluster D	-	Cluster H	1.550	0.516	314.6	3.003	.070
Cluster D	-	Cluster I	2.423	0.578	326.2	4.196	.001
Cluster E	-	Cluster F	0.851	0.295	302.6	2.889	.095
Cluster E	-	Cluster G	1.487	0.361	308.4	4.124	.002
Cluster E	-	Cluster H	1.421	0.338	324.1	4.204	.001
Cluster E	-	Cluster I	2.294	0.418	322.6	5.487	<.001
Cluster F	-	Cluster G	0.636	0.347	283.4	1.831	.662
Cluster F	-	Cluster H	0.570	0.331	316.8	1.723	.732
Cluster F	-	Cluster I	1.443	0.404	325.0	3.570	.012
Cluster G	-	Cluster H	−0.066	0.387	308.3	−0.171	1.000
Cluster G	-	Cluster I	0.807	0.446	315.1	1.809	.677
Cluster H	-	Cluster I	0.874	0.429	321.6	2.038	.518

Table A2
post-hoc t-tests for post-test score per cluster

Comparison							
Condition		Condition	Mean Difference	SE	df	t	p_{Tukey}
Cluster A	-	Cluster B	0.001	0.636	291.1	0.002	1.000
Cluster A	-	Cluster C	0.127	0.489	310.0	0.260	1.000
Cluster A	-	Cluster D	0.157	0.660	309.9	0.238	1.000
Cluster A	-	Cluster E	1.185	0.503	316.4	2.354	.313
Cluster A	-	Cluster F	2.243	0.522	313.5	4.301	<.001
Cluster A	-	Cluster G	2.595	0.567	314.2	4.579	<.001
Cluster A	-	Cluster H	2.970	0.551	306.2	5.386	<.001
Cluster A	-	Cluster I	3.953	0.589	300.7	6.716	<.001
Cluster B	-	Cluster C	0.126	0.504	268.6	0.249	1.000
Cluster B	-	Cluster D	0.156	0.672	287.8	0.231	1.000
Cluster B	-	Cluster E	1.183	0.518	291.6	2.283	.356
Cluster B	-	Cluster F	2.242	0.527	300.0	4.254	<.001
Cluster B	-	Cluster G	2.594	0.575	306.2	4.512	<.001
Cluster B	-	Cluster H	2.969	0.557	311.1	5.329	<.001

(continued on next page)

Table A2 (*continued*)

Comparison							
Condition		Condition	Mean Difference	SE	df	t	p_{Tukey}
Cluster B	-	Cluster I	3.952	0.596	316.3	6.634	<.001
Cluster C	-	Cluster D	0.030	0.513	280.1	0.058	1.000
Cluster C	-	Cluster E	1.058	0.301	307.0	3.512	.015
Cluster C	-	Cluster F	2.116	0.317	316.2	6.673	<.001
Cluster C	-	Cluster G	2.468	0.390	317.0	6.327	<.001
Cluster C	-	Cluster H	2.843	0.369	307.6	7.711	<.001
Cluster C	-	Cluster I	3.826	0.423	297.8	9.053	<.001
Cluster D	-	Cluster E	1.028	0.516	282.6	1.991	.551
Cluster D	-	Cluster F	2.086	0.523	277.3	3.988	.003
Cluster D	-	Cluster G	2.438	0.571	290.0	4.274	<.001
Cluster D	-	Cluster H	2.813	0.554	298.7	5.079	<.001
Cluster D	-	Cluster I	3.797	0.599	313.0	6.335	<.001
Cluster E	-	Cluster F	1.058	0.302	295.5	3.499	.016
Cluster E	-	Cluster G	1.411	0.370	294.5	3.816	.005
Cluster E	-	Cluster H	1.786	0.350	315.5	5.104	<.001
Cluster E	-	Cluster I	2.769	0.407	315.2	6.804	<.001
Cluster F	-	Cluster G	0.352	0.354	269.5	0.994	.986
Cluster F	-	Cluster H	0.727	0.335	305.5	2.168	.429
Cluster F	-	Cluster I	1.711	0.390	315.5	4.391	<.001
Cluster G	-	Cluster H	0.375	0.394	290.2	0.951	.990
Cluster G	-	Cluster I	1.358	0.438	302.3	3.103	.053
Cluster H	-	Cluster I	0.983	0.411	303.8	2.392	.292

Table A3
post-hoc *t*-test for normalised learning gain by final clusters

Comparison							
Condition		Condition	Mean Difference	SE	df	t	p_{Tukey}
Cluster A	-	Cluster B	0.164	0.176	266.8	0.928	.991
Cluster A	-	Cluster C	−0.062	0.138	283.9	−0.450	1.000
Cluster A	-	Cluster D	−0.045	0.173	284.7	−0.258	1.000
Cluster A	-	Cluster E	0.190	0.140	289.1	1.357	.913
Cluster A	-	Cluster F	0.315	0.145	292.9	2.175	.425
Cluster A	-	Cluster G	0.299	0.155	292.7	1.930	.593
Cluster A	-	Cluster H	0.360	0.151	290.4	2.383	.297
Cluster A	-	Cluster I	0.493	0.163	287.7	3.031	.065
Cluster B	-	Cluster C	−0.226	0.134	261.1	−1.688	.753
Cluster B	-	Cluster D	−0.208	0.171	272.8	−1.220	.952
Cluster B	-	Cluster E	0.026	0.136	273.9	0.192	1.000
Cluster B	-	Cluster F	0.151	0.138	283.1	1.093	.975
Cluster B	-	Cluster G	0.135	0.149	286.4	0.906	.993
Cluster B	-	Cluster H	0.196	0.145	288.8	1.356	.913
Cluster B	-	Cluster I	0.330	0.157	292.4	2.102	.474
Cluster C	-	Cluster D	0.017	0.125	263.4	0.139	1.000
Cluster C	-	Cluster E	0.252	0.075	285.7	3.376	.023
Cluster C	-	Cluster F	0.377	0.078	291.8	4.801	<.001
Cluster C	-	Cluster G	0.361	0.096	293.0	3.749	.007
Cluster C	-	Cluster H	0.422	0.090	280.4	4.664	<.001
Cluster C	-	Cluster I	0.555	0.108	276.1	5.120	<.001
Cluster D	-	Cluster E	0.234	0.126	265.6	1.863	.640
Cluster D	-	Cluster F	0.359	0.128	261.2	2.808	.118
Cluster D	-	Cluster G	0.343	0.140	273.7	2.454	.259
Cluster D	-	Cluster H	0.404	0.135	280.2	2.990	.073
Cluster D	-	Cluster I	0.538	0.150	291.3	3.585	.012
Cluster E	-	Cluster F	0.125	0.076	273.8	1.654	.773
Cluster E	-	Cluster G	0.109	0.092	275.6	1.182	.960
Cluster E	-	Cluster H	0.170	0.087	292.3	1.958	.574
Cluster E	-	Cluster I	0.304	0.106	290.6	2.868	.101
Cluster F	-	Cluster G	−0.016	0.088	251.1	−0.184	1.000
Cluster F	-	Cluster H	0.045	0.083	284.4	0.543	1.000
Cluster F	-	Cluster I	0.179	0.100	290.9	1.781	.694
Cluster G	-	Cluster H	0.061	0.098	275.4	0.625	.999
Cluster G	-	Cluster I	0.195	0.112	282.0	1.737	.723
Cluster H	-	Cluster I	0.134	0.107	286.1	1.251	.944

Table A4
Mixed linear model practice effort for all attempts by cluster

Comparison							
Condition		Condition	Mean Difference	SE	df	t	p_{Tukey}
Cluster A	-	Cluster B	−12.951	14.560	303.7	−0.889	0.993
Cluster A	-	Cluster C	−17.854	10.750	326.2	−1.661	0.770
Cluster A	-	Cluster D	−23.201	14.488	323.1	−1.601	0.804
Cluster A	-	Cluster E	−27.126	11.159	339.2	−2.431	0.271
Cluster A	-	Cluster F	−38.526	11.809	319.7	−3.262	0.033
Cluster A	-	Cluster G	−36.263	12.891	324.6	−2.813	0.116
Cluster A	-	Cluster H	−43.096	12.552	325.1	−3.434	0.019
Cluster A	-	Cluster I	−23.482	13.636	319.4	−1.722	0.733
Cluster B	-	Cluster C	−4.903	11.911	284.7	−0.412	1.000
Cluster B	-	Cluster D	−10.250	15.499	304.7	−0.661	0.999
Cluster B	-	Cluster E	−14.175	12.288	306.8	−1.154	0.965
Cluster B	-	Cluster F	−25.575	12.620	313.8	−2.027	0.526
Cluster B	-	Cluster G	−23.312	13.700	322.5	−1.702	0.745
Cluster B	-	Cluster H	−30.145	13.327	328.2	−2.262	0.369
Cluster B	-	Cluster I	−10.531	14.398	336.3	−0.731	0.998
Cluster C	-	Cluster D	−5.348	11.398	298.8	−0.469	1.000
Cluster C	-	Cluster E	−9.272	6.841	323.7	−1.355	0.913
Cluster C	-	Cluster F	−20.673	7.494	272.6	−2.759	0.133
Cluster C	-	Cluster G	−18.409	9.107	307.7	−2.021	0.530
Cluster C	-	Cluster H	−25.242	8.703	312.9	−2.901	0.093
Cluster C	-	Cluster I	−5.629	10.220	305.9	−0.551	1.000
Cluster D	-	Cluster E	−3.924	11.462	298.2	−0.342	1.000
Cluster D	-	Cluster F	−15.325	11.721	284.6	−1.308	0.929
Cluster D	-	Cluster G	−13.061	12.861	302.3	−1.016	0.984
Cluster D	-	Cluster H	−19.894	12.515	312.7	−1.590	0.810
Cluster D	-	Cluster I	−0.281	13.782	325.2	−0.020	1.000
Cluster E	-	Cluster F	−11.401	6.977	295.9	−1.634	0.785
Cluster E	-	Cluster G	−9.137	8.510	307.1	−1.074	0.978
Cluster E	-	Cluster H	−15.970	8.102	330.9	−1.971	0.565
Cluster E	-	Cluster I	3.644	9.704	337.0	0.375	1.000
Cluster F	-	Cluster G	2.264	8.093	284.1	0.280	1.000
Cluster F	-	Cluster H	−4.569	7.845	323.2	−0.582	1.000
Cluster F	-	Cluster I	15.044	9.316	338.7	1.615	0.796
Cluster G	-	Cluster H	−6.833	9.084	306.4	−0.752	0.998
Cluster G	-	Cluster I	12.781	10.210	316.4	1.252	0.944
Cluster H	-	Cluster I	19.614	9.752	320.1	2.011	0.537

Table A5
Mixed linear model practice accuracy for attempts by cluster

Comparison							
Condition		Condition	Mean Difference	SE	df	t	p_{Tukey}
Cluster A	-	Cluster B	0.067	0.028	329.5	2.421	.276
Cluster A	-	Cluster C	0.142	0.020	342.3	7.035	<.001
Cluster A	-	Cluster D	0.189	0.027	340.7	6.929	<.001
Cluster A	-	Cluster E	0.253	0.021	341.7	12.259	<.001
Cluster A	-	Cluster F	0.342	0.021	330.2	16.025	<.001
Cluster A	-	Cluster G	0.441	0.023	332.4	18.907	<.001
Cluster A	-	Cluster H	0.418	0.023	320.6	18.529	<.001
Cluster A	-	Cluster I	0.555	0.024	312.5	22.773	<.001
Cluster B	-	Cluster C	0.075	0.023	312.3	3.247	.035
Cluster B	-	Cluster D	0.122	0.030	330.5	4.111	.002
Cluster B	-	Cluster E	0.185	0.023	330.6	7.912	<.001
Cluster B	-	Cluster F	0.275	0.024	336.1	11.539	<.001
Cluster B	-	Cluster G	0.374	0.026	340.5	14.554	<.001
Cluster B	-	Cluster H	0.350	0.025	342.0	14.058	<.001
Cluster B	-	Cluster I	0.487	0.027	343.1	18.296	<.001
Cluster C	-	Cluster D	0.047	0.022	326.0	2.145	.445
Cluster C	-	Cluster E	0.111	0.013	340.6	8.585	<.001
Cluster C	-	Cluster F	0.200	0.014	344.3	14.575	<.001
Cluster C	-	Cluster G	0.299	0.017	342.5	17.950	<.001
Cluster C	-	Cluster H	0.276	0.016	321.9	17.631	<.001
Cluster C	-	Cluster I	0.412	0.018	307.4	22.670	<.001
Cluster D	-	Cluster E	0.064	0.022	325.4	2.903	.092
Cluster D	-	Cluster F	0.153	0.022	320.6	6.846	<.001
Cluster D	-	Cluster G	0.252	0.024	333.3	10.367	<.001
Cluster D	-	Cluster H	0.229	0.024	338.0	9.701	<.001
Cluster D	-	Cluster I	0.366	0.026	343.8	14.304	<.001
Cluster E	-	Cluster F	0.089	0.013	331.7	6.775	<.001

(*continued on next page*)

Table A5 (*continued*)

Comparison							
Condition		Condition	Mean Difference	SE	df	t	p_{Tukey}
Cluster E	-	Cluster G	0.189	0.016	334.0	11.728	<.001
Cluster E	-	Cluster H	0.165	0.015	343.5	10.954	<.001
Cluster E	-	Cluster I	0.302	0.018	333.9	17.041	<.001
Cluster F	-	Cluster G	0.099	0.016	311.2	6.341	<.001
Cluster F	-	Cluster H	0.076	0.015	340.9	5.103	<.001
Cluster F	-	Cluster I	0.212	0.017	341.5	12.269	<.001
Cluster G	-	Cluster H	−0.024	0.017	332.7	−1.368	.909
Cluster G	-	Cluster I	0.113	0.019	338.0	5.842	<.001
Cluster H	-	Cluster I	0.137	0.018	341.0	7.426	<.001

Table A6
Mixed linear model practice accuracy for unique problems by cluster

Comparison							
Condition		Condition	Mean Difference	SE	df	t	p_{Tukey}
Cluster A	-	Cluster B	0.072	0.025	334.7	2.912	.090
Cluster A	-	Cluster C	0.144	0.018	343.3	8.001	<.001
Cluster A	-	Cluster D	0.192	0.024	342.9	7.896	<.001
Cluster A	-	Cluster E	0.248	0.018	338.7	13.531	<.001
Cluster A	-	Cluster F	0.310	0.019	324.0	16.405	<.001
Cluster A	-	Cluster G	0.415	0.021	327.3	20.064	<.001
Cluster A	-	Cluster H	0.395	0.020	313.6	19.823	<.001
Cluster A	-	Cluster I	0.550	0.021	305.1	25.627	<.001
Cluster B	-	Cluster C	0.072	0.021	319.4	3.462	.017
Cluster B	-	Cluster D	0.120	0.026	335.9	4.519	<.001
Cluster B	-	Cluster E	0.175	0.021	335.5	8.362	<.001
Cluster B	-	Cluster F	0.237	0.021	339.6	11.159	<.001
Cluster B	-	Cluster G	0.342	0.023	342.6	14.925	<.001
Cluster B	-	Cluster H	0.322	0.022	343.2	14.504	<.001
Cluster B	-	Cluster I	0.478	0.024	341.9	20.190	<.001
Cluster C	-	Cluster D	0.048	0.020	332.2	2.458	.257
Cluster C	-	Cluster E	0.104	0.011	342.7	9.025	<.001
Cluster C	-	Cluster F	0.166	0.012	343.6	13.589	<.001
Cluster C	-	Cluster G	0.271	0.015	340.4	18.276	<.001
Cluster C	-	Cluster H	0.251	0.014	314.9	18.153	<.001
Cluster C	-	Cluster I	0.406	0.016	299.1	25.346	<.001
Cluster D	-	Cluster E	0.056	0.020	331.8	2.829	.111
Cluster D	-	Cluster F	0.118	0.020	327.1	5.873	<.001
Cluster D	-	Cluster G	0.223	0.022	338.0	10.222	<.001
Cluster D	-	Cluster H	0.203	0.021	341.3	9.621	<.001
Cluster D	-	Cluster I	0.358	0.023	342.9	15.762	<.001
Cluster E	-	Cluster F	0.062	0.012	336.3	5.267	<.001
Cluster E	-	Cluster G	0.167	0.014	338.0	11.608	<.001
Cluster E	-	Cluster H	0.147	0.013	342.9	10.968	<.001
Cluster E	-	Cluster I	0.303	0.016	328.3	19.301	<.001
Cluster F	-	Cluster G	0.105	0.014	318.3	7.457	<.001
Cluster F	-	Cluster H	0.085	0.013	342.7	6.421	<.001
Cluster F	-	Cluster I	0.240	0.015	338.5	15.647	<.001
Cluster G	-	Cluster H	−0.020	0.015	337.5	−1.291	.934
Cluster G	-	Cluster I	0.136	0.017	341.2	7.854	<.001
Cluster H	-	Cluster I	0.156	0.016	342.9	9.481	<.001

A.2 Cluster names

Table A7
Matching cluster names to cluster labels

Label	Name
A	Proceeding already-proficient
B	Already-proficient
C	Fast-mastering
D	Mastering
E	Slow-mastering
F	Plateauing
G	Rising strugglers
H	Brief-risen strugglers
I	Stagnators

A.3 Mixed model results

Table A8
Mixed linear model for pre-test

Variable					
Fixed part	B	SE	df	t	p
Intercept (A: proceeding already-proficient)	6.197	0.583	7.26	10.62582	<0.001
B: already-proficient	−0.366	0.599	299.06	−0.6112	.542
C: fast-mastering	−2.545	0.439	318.51	−5.79639	<0.001
D: mastering	−2.846	0.590	316.13	−4.82204	<0.001
E: slow-mastering	−2.974	0.451	325.77	−6.58771	<0.001
F: plateauing	−3.826	0.471	320.23	−8.12133	<0.001
G: rising strugglers	−4.462	0.517	321.24	−8.63329	<0.001
H: brief-risen strugglers	−4.395	0.499	311.53	−8.80175	<0.001
I: stagnators	−5.269	0.554	303.02	−9.5133	<0.001
Random part	σ	SD			
Intercept Skill	0.513	0.716			
Intercept Child	0.710	0.843			
Residual	2.126	1.458			

Note: $N_{pre\text{-}test} = 336$, $N_{children} = 134$, $N_{skills} = 3$, $R^2 = 0.57$.

Table A9
Mixed linear model for post-test including degrees of freedom

Variable					
Fixed part	β	SE	df	t	p
Intercept (A: proceeding already-proficient)	6.737	0.734	5.3	9.178	<.001
B: already-proficient	−0.001	0.633	287.5	−0.002	.998
C: fast-mastering	−0.127	0.486	309.0	−0.261	.794
D: mastering	−0.157	0.656	308.9	−0.239	.811
E: slow-mastering	−1.185	0.500	316.4	−2.370	.018
F: plateauing	−2.243	0.517	313.0	−4.335	<.001
G: rising strugglers	−2.595	0.563	313.8	−4.613	<.001
H: brief-risen strugglers	−2.970	0.547	304.7	−5.427	<.001
I: stagnators	−3.953	0.584	298.4	−6.768	<.001
Random part	σ	SD			
Intercept Skill	0.975	0.987			
Intercept Child	0.885	0.941			
Residual	2.128	1.459			

Note: $N_{post\text{-}test} = 327$, $N_{children} = 134$, $N_{skills} = 3$, $R^2 = 0.62$.

Table A10
Mixed linear model for learning gain

Variable					
Fixed part	β	SE	df	t	p
Intercept (A: proceeding already-proficient)	0.654	0.153	21.7	4.262	<.001
B: already-proficient	−0.164	0.175	266.5	−0.933	.352
C: fast-mastering	0.062	0.137	283.9	0.453	.651
D: mastering	0.045	0.172	284.7	0.259	.796
E: slow-mastering	−0.190	0.139	289.1	−1.367	.173
F: plateauing	−0.315	0.143	292.9	−2.201	.029
G: rising strugglers	−0.299	0.153	292.7	−1.951	.052
H: brief-risen strugglers	−0.360	0.149	290.3	−2.409	.017
I: stagnators	−0.493	0.161	287.6	−3.068	.002
Random part	σ	SD			
Intercept Skill	0.020	0.140			
Intercept Child	0.041	0.203			
Residual	0.128	0.358			

Note: $N_{learning\ gain} = 311$, $N_{children} = 134$, $N_{skills} = 3$, $R^2 = 0.41$.

Table A11
Mixed linear model for practice effort on all problems

Variable	β	SE	df	t	p
Fixed part					
Intercept (A: proceeding already-proficient)	54.291	10.766	68.1	5.043	<0.001

(*continued on next page*)

Table A11 (*continued*)

Variable	β	SE	df	t	p
B: already-proficient	12.951	14.488	299.5	0.894	.372
C: fast-mastering	17.854	10.648	323.9	1.677	.095
D: mastering	23.201	14.315	320.4	1.621	.106
E: slow-mastering	27.126	10.988	338.6	2.469	.014
F: plateauing	38.526	11.444	320.2	3.367	<0.001
G: rising strugglers	36.263	12.521	325.0	2.896	.004
H: brief-risen strugglers	43.096	12.202	325.4	3.532	<0.001
I: stagnators	23.482	13.243	319.6	1.773	.077
Random part	σ	SD			
Intercept Skill	38.4	6.2			
Intercept Child	644.6	25.39			
Residual	1172.69	34.24			

Note: $N_{problems} = 354$, $N_{children} = 134$, $N_{skills} = 3$, $R^2 = 0.41$.

Table A12
Mixed linear model for practice effort on unique problems

Variable					
Fixed part	β	SE	df	t	P
Intercept (A: proceeding already-proficient)	48.438	5.128	28.6	9.447	<0.001
B: already-proficient	2.426	6.228	282.9	0.390	.697
C: fast-mastering	5.693	4.616	306.8	1.233	.218
D: mastering	8.081	6.203	304.4	1.303	.194
E: slow-mastering	9.373	4.810	329.6	1.949	.052
F: plateauing	8.862	5.072	338.0	1.747	.081
G: rising strugglers	4.906	5.547	339.3	0.884	.377
H: brief-risen strugglers	5.813	5.430	342.0	1.071	.285
I: stagnators	−6.355	5.915	340.2	−1.074	.283
Random part	σ	SD			
Intercept Skill	17.91	4.23			
Intercept Child	172.5	13.13			
Residual	205.17	14.35			

Note: $N_{problems} = 354$, $N_{children} = 134$, $N_{skills} = 3$, $R^2 = 0.50$.

Table A13
Mixed linear model for practice accuracy on all problems

Variable					
Fixed part	β	SE	df	t	p
Intercept (A: proceeding already-proficient)	0.905	0.027	4.8	33.611	<0.001
B: already-proficient	−0.072	0.025	334.0	−2.927	.004
C: fast-mastering	−0.144	0.018	343.3	−8.050	<0.001
D: mastering	−0.192	0.024	342.9	−7.945	<0.001
E: slow-mastering	−0.248	0.018	338.5	−13.617	<0.001
F: plateauing	−0.310	0.019	322.9	−16.530	<0.001
G: rising strugglers	−0.415	0.021	326.4	−20.205	<0.001
H: brief-risen strugglers	−0.395	0.020	312.0	−19.962	<0.001
I: stagnators	−0.550	0.021	303.1	−25.805	<0.001
Random part	σ	SD			
Intercept Skill	0.0014	0.037			
Intercept Child	0.00059	0.024			
Residual	0.0039	0.063			

Note: $N_{problems} = 354$, $N_{children} = 134$, $N_{skills} = 3$, $R^2 = 0.84$.

Table A14
Mixed linear model for practice accuracy for unique problems

Variable					
Fixed part	β	SE	df	t	p
Intercept (A: proceeding already-proficient)	0.896	0.034	4.0	26.603	<0.001
B: already-proficient	−0.067	0.028	328.6	−2.433	.016
C: fast-mastering	−0.142	0.020	342.2	−7.076	<0.001
D: mastering	−0.189	0.027	340.5	−6.969	<0.001
E: slow-mastering	−0.253	0.020	341.7	−12.333	<0.001
F: plateauing	−0.342	0.021	329.8	−16.140	<0.001
G: rising strugglers	−0.441	0.023	332.1	−19.032	<0.001
H: brief-risen strugglers	−0.418	0.022	319.8	−18.653	<0.001
I: stagnators	−0.555	0.024	311.3	−22.924	<0.001
Random part	σ	SD			

(*continued on next page*)

Table A14 (*continued*)

Variable		
Intercept Skill	0.0024	0.049
Intercept Child	0.0010	0.032
Residual	0.0048	0.069

Note: $N_{problems} = 354$, $N_{children} = 134$, $N_{skills} = 3$, $R^2 = 0.84$.

Appendix B

In this appendix, we describe the mathematical details of the adapted version of the Dirichlet process Gaussian process mixture (DPGP) model. For more details, we refer the reader to the work by McDowell et al. (2018). First, we clarify the notation used: This paper uses lower-case boldface Roman symbols to denote vectors and upper-case boldface symbols to denote matrices. Scalars are indicated in lower-case italics and sets with upper-case italics.

When a child practices a particular skill, an ability curve is created for that skill for the child. This ability curve is indexed with j. The problems solved by the child are the x-values for the ability curve and are denoted with vector $\mathbf{x}_j$. All problems solved for all skills and all students are denoted in set X. The ability scores for the child working on the particular skill are the y-values of the ability curve and are denoted in vector $\mathbf{y}_j$. All ability scores for all students are denoted in set Y. A cluster is indexed with h. The assignments of all ability curves to the clusters are denoted with vector z, where $z_j = h$ if ability curve j is assigned to cluster h. The assignment of all ability curves except for ability curve j is denoted with vector $z_{\neg j}$. The hyperparameters of all clusters are denoted with set Θ, with θ_h being the hyperparameters for cluster h and θ_{z_j} being the hyperparameters of the cluster assigned to ability curve j.

B.1 Dirichlet process

The Dirichlet process prior defines a distribution over the distribution of cluster parameters. It is defined as follows:

$$G|\alpha, G_0 \sim DP(\alpha, G_0); \tag{B1}$$

$$\theta_h|G \sim G; \tag{B2}$$

$$\boldsymbol{y}_j|\theta_h \sim p(\cdot|\theta_h); \tag{B3}$$

Distribution G is drawn from a Dirichlet process with base distribution G_0 and concentration parameter α. From G, the latent variables θ_h are drawn for each cluster h. Observations (the ability curves) are then drawn from those parameters.

B.2 Gaussian Process

The Gaussian process defines how, in an ability curve j, problems $\boldsymbol{x}_j$ are mapped to ability scores $\boldsymbol{y}_j$ according to cluster h. A Gaussian process is defined by its mean function and its covariance function (see section 2.6.2), which depend on parameters. These parameters are assumed to follow these distributions, specific for each cluster h:

$$\kappa_h \sim \mathcal{N}(15, 25) \tag{B4}$$

$$\lambda_h \sim \ln \mathcal{N}(1, 30) \tag{B5}$$

$$\ell_h \sim \ln \mathcal{N}(0, 3) \tag{B6}$$

$$\tau_h \sim \ln \mathcal{N}(0, 3) \tag{B7}$$

$$\sigma_h^2 \sim \text{InverseGamma}(2, 4) \tag{B8}$$

Using κ_h and λ_h defines the mean of each cluster:

$$\mu_h(x) = \max\left(0, \frac{2\, y_{\max}}{1 + \exp\left(-\frac{x-\kappa_h}{\lambda_h}\right)} - y_{\max}\right). \tag{B9}$$

The parameter $y_{\max}$ stands for the maximum value an observed ability score can take, κ_h corresponds to the point where the mean starts to rise, and λ_h corresponds to how fast the mean rises. Note that this approach to modelling the cluster mean curve deviates from the approach by McDowell et al. (2018), where a constant zero mean function is used instead.

Parameters ℓ_h and τ_h in the cluster-specific squared-exponential covariance function, which is given by:

$$k_h(x, x') = \tau_h^2 \exp\left(\frac{||x - x'||^2}{-2\ell_h^2}\right) \tag{B10}$$

Parameter ℓ_h corresponds to the smoothness of the curve (its length scale), and parameter τ_h for the vertical variance of the curve. We define covariance kernel function K, which applies the covariance function k on each pairwise combination of x's in $\boldsymbol{x}_j$ to obtain matrix $\boldsymbol{K}$ where each element $\boldsymbol{K}_{pq} = k(x_p, x_q)$ for all points $x_p \in \boldsymbol{x}_j$ and $x_q \in \boldsymbol{x}_j$.

B.3 Likelihood

Combining the mean and covariance function of the Gaussian process results in the prototypical curve of each cluster f_h:

$$f_h|\theta_h \sim GP(\mu_h, K_h). \tag{B11}$$

The likelihood of observing the trace data results then results in the following formula:

$$p(\boldsymbol{y}_j|\boldsymbol{x}_j, f_h, z_j = h) = \mathcal{N}(f_h(\boldsymbol{x}_j), \sigma_h^2). \tag{B12}$$

Here $\mathcal{N}$ stands for the Gaussian distribution. Each cluster has its own noise variance parameter σ_h^2, which determines how far observations can deviate from the prototypical curve of a cluster. Since the GP prior is also Gaussian, we can also write the likelihood without using f_h.

$$p(\boldsymbol{y}_j|\boldsymbol{x}_j, \theta_h, z_j = h) = \mathcal{N}\left(\mu_h(\boldsymbol{x}_j), K_h\left(\boldsymbol{x}_j, \boldsymbol{x}_j^{'}\right) + \sigma_h^2 I\right), \tag{B13}$$

with I being the identity matrix.

B.4 Markov chain Monte Carl (MCMC) to estimate the posterior distribution

MCMC (Neal, 2000) is used to calculate the posterior probability that ability curve j belongs to cluster h according to the DP prior and that the mean curve for cluster h is distributed according to the GP distribution. To estimate the posterior distribution of the ability curve assignments, Neal's Gibbs sampling "Algorithm 8" was used (Neal, 2000). To estimate the posterior distribution of the parameters of each cluster, we use Metropolis-Hastings sampling (Hastings, 1970).

Using Bayes' Rule, the probability of ability curve j belonging to cluster h (notated as z_j=h) conditioned on the data and all other cluster assignments is:

$$p(z_j = h|\boldsymbol{y}_j, \boldsymbol{x}_j, z_{\neg j}, \theta_h, \alpha) \propto p(z_j = h|z_{\neg j}, \alpha) p(\boldsymbol{y}_j|\boldsymbol{x}_j, z_j = h, \theta_h). \tag{B14}$$

The first term on the right hand is the probability of assigning the ability curve to cluster h. This probability is given by:

$$p(z_j = h|z_{\neg j}, \alpha) \propto \begin{cases} \dfrac{\frac{\alpha}{m}}{\alpha + n - 1} & \text{if } h \text{ is empty.} \\ \dfrac{n_h}{\alpha + n - 1} & \text{otherwise.} \end{cases} \tag{B15}$$

In this formula, m stands for the number of empty clusters, which we have set to 4; n stands for the number of curves already clustered. n_h stands for the number of curves assigned to cluster h. The likelihood term in equation (B14) is calculated using the cluster-specific GPs with Equation (B13).

At initialisation, a cluster assignment for each curve is drawn from the DP by repeating equation (B15) for each ability curve. Then, sampling is done for 2000 MCMC iterations, where we sample both the parameters from the Gaussian processes and the Dirichlet process in every iteration. The cluster parameters are drawn for each new cluster from their respective distributions given in equation (B4) to (B8).

Before each iteration, 4 new empty clusters are generated, with cluster parameters drawn following equations (B4-B8). Next, the ability curves get reassigned to the clusters, one after another, following the probability given in Equation (B14). After this, we can calculate the posterior probability of the mean μ_h^* and the covariance matrix K_h^* with the following formula's (Rasmussen & Williams, 2006):

$$\bar{\boldsymbol{y}}_h = \frac{\boldsymbol{y}_1 + \ldots + \boldsymbol{y}_k}{n_\mathrm{h}} \tag{B16}$$

$$\mu_h^*(\mathbf{x}) = \mu_h(\mathbf{x}) + K_h(\mathbf{x}_h, \mathbf{x})\left[K_h(\mathbf{x}_h, \mathbf{x}_h) + \sigma_h^2 \mathbf{I}\right]^{-1}(\bar{\mathbf{y}}_h - \mu_h(\mathbf{x}_h)); \tag{B17}$$

$$K_h^*(\mathbf{x}) = K_h(\mathbf{x}, \mathbf{x}) - K_h(\mathbf{x}, \mathbf{x}_h)\left[K_h(\mathbf{x}_h, \mathbf{x}_h) + \sigma_h^2 I\right]^{-1} K_h(\mathbf{x}, \mathbf{x}_h); \tag{B18}$$

After the sampling, the Gaussian process hyperparameters for the most optimal selection were sampled for an additional 2000 iterations while keeping the cluster assignments **Z** fixed, so that given the cluster assignment at the last iterations, the approximate inference of the corresponding hyperparameters has sufficient time to converge.

Convergence was assessed following McDowell et al. (2018), based on the change in posterior likelihood. Further details about using MCMC to obtain the posterior distribution of DPGP can also be found in McDowell et al. (2018).

References

Aleven, V., McLaughlin, E., Glenn, R. A., & Koedinger, K. R. (2016). Instruction based on adaptive learning technologies. *Handbook of Research on Learning and Instruction, 2*, 522–560.

Azevedo, R. (2007). Understanding the complex nature of self-regulatory processes in learning with computer-based learning environments: An introduction. *Metacognition and Learning, 2*(2), 57–65. https://doi.org/10.1007/S11409-007-9018-5, 2007 2:2.

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., & Gašević, D. (2019). Analyzing multimodal multichannel data about self-regulated learning with advanced learning technologies: Issues and challenges. *Computers in Human Behavior, 96*, 207–210 (Elsevier).

Baker, R. S., Goldstein, A. B., & Heffernan, N. T. (2011). Detecting learning moment-by-moment. *International Journal of Artificial Intelligence in Education, 21*(1–2), 5–25.

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). Relevance of learning analytics to measure and support students' learning in adaptive educational

technologies. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 568–569).
Barton, K., & Barton, M. K. (2022). *MuMIn: Multi-Model Inference* (1.47.1) [R].
Buxbaum-Conradi, S., Redlich, T., & Branding, J. H. (2016). In *Conceptualizing hybrid human-machine systems and interaction. 2016 49th Hawaii International conference on system sciences (HICSS)* (pp. 551–559). https://doi.org/10.1109/HICSS.2016.75
Elo, A. E. (1978). *The rating of chessplayers, past and present.* BT Batsford Limited.
Fan, Y., Saint, J., Singh, S., Jovanovic, J., & Gašević, D. (2021). A learning analytic approach to unveiling self-regulatory processes in learning tactics. In *LAK21: 11th International learning analytics and knowledge conference* (pp. 184–195).
Fincham, E., Gašević, D., Jovanović, J., & Pardo, A. (2019). From study tactics to learning strategies: An analytical method for extracting interpretable representations. *IEEE Transactions on Learning Technologies, 12*(1), 59–72. https://doi.org/10.1109/TLT.2018.2823317
Fox, J. (2019). *Regression diagnostics: An introduction.* Sage publications.
Greene, J. A., & Azevedo, R. (2007). Adolescents' use of self-regulatory processes and their relation to qualitative mental model shifts while using hypermedia. *Journal of Educational Computing Research, 36*(2), 125–148. https://doi.org/10.2190/G7M1-2734-3JRR-8033
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
Hadwin, A. F. (2011). Self-regulated learning. In *21st century education: A reference handbook* (pp. 175–183). Sage.
Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*(1), 64–74.
Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika, 57*(1), 97–109. https://doi.org/10.1093/BIOMET/57.1.97
Horvers, A., Kooi, R., Knoop-van Campen, C., Dijkstra, S.H.E., & Molenaar, I. (Submitted). *Cue Utilization Prompts to Support Self-Regulated Learning in Adaptive Learning Technologies* [Paper submitted for publication].
Järvelä, S., Järvenoja, H., Malmberg, J., & Hadwin, A. F. (2013). Exploring socially shared regulation in the context of collaboration. *Journal of Cognitive Education and Psychology, 12*(3), 267–286. https://doi.org/10.1891/1945-8959.12.3.267
Järvelä, S., Malmberg, J., Haataja, E., Sobocinski, M., & Kirschner, P. A. (2019). What multimodal data can tell us about the students' regulation of their learning process. *Learning and Instruction, 72*(7), 4.
Kay, J., Bartimote, K., Kitto, K., Kummerfeld, B., Liu, D., & Reimann, P. (2022). Enhancing learning by Open Learner Model (OLM) driven data design. *Computers & Education: Artificial Intelligence, 3*, Article 100069. https://doi.org/10.1016/j.caeai.2022.100069
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
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. (2022). *emmeans: Estimated marginal means, aka least-squares means*. https://CRAN.R-project.org/package=emmeans.
Li, S., Chen, G., Xing, W., Zheng, J., & Xie, C. (2020). Longitudinal clustering of students' self-regulated learning behaviors in engineering design. *Computers & Education, 153*, Article 103899. https://doi.org/10.1016/J.COMPEDU.2020.103899
Long, Y., & Aleven, V. (2017). Enhancing learning outcomes through self-regulated learning support with an Open Learner Model. *User Modeling and User-Adapted Interaction, 27*(1), 55–88. https://doi.org/10.1007/s11257-016-9186-6
Lust, G., Vandewaetere, M., Ceulemans, E., Elen, J., & Clarebout, G. (2011). Tool-use in a blended undergraduate course: In Search of user profiles. *Computers & Education, 57* (3), 2135–2144. https://doi.org/10.1016/J.COMPEDU.2011.05.010
Martinez-Pons, M. (2002). Parental influences on children's academic self-regulatory development. *Theory Into Practice, 41*(2), 126–131. https://doi.org/10.1207/S15430421TIP4102_9
Marx, J. D., & Cummings, K. (2007). Normalized change. *American Journal of Physics, 75* (1), 87–91.
Matcha, W., Gašević, D., Uzir, N. A., Jovanović, J., & Pardo, A. (2019). Analytics of learning strategies: Associations with academic performance and feedback. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, 461–470.
McDowell, I. C., Manandhar, D., Vockley, C. M., Schmid, A. K., Reddy, T. E., & Engelhardt, B. E. (2018). Clustering gene expression time series data using an infinite Gaussian process mixture model. *PLoS Computational Biology, 14*(1). https://doi.org/10.1371/JOURNAL.PCBI.1005896
Mirriahi, N., Liaqat, D., Dawson, S., & Gašević, D. (2016). Uncovering student learning profiles with a video annotation tool: Reflective learning with and without instructional norms. *Educational Technology Research & Development, 64*(6), 1083–1106. https://doi.org/10.1007/S11423-016-9449-2
Molenaar, I. (2014). Advances in temporal analysis in learning and instruction. *Frontline Learning Research, 2*(4). https://doi.org/10.14786/flr.v2i4.118. Article 4.
Molenaar, I. (2022). Towards hybrid human-AI learning technologies. *European Journal of Education*.
Molenaar, I., Horvers, A., & Baker, R. (2019). Towards hybrid human-system regulation: Understanding children's SRL support needs in blended classrooms. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge* (pp. 471–480). https://doi.org/10.1145/3303772.3303780.
Molenaar, I., Horvers, A., & Baker, R. S. (2021). What can moment-by-moment learning curves tell about students' self-regulated learning? *Learning and Instruction, 72*, Article 101206.
Molenaar, I., & Järvelä, S. (2014). Sequential and temporal characteristics of self and socially regulated learning. *Metacognition and Learning, 9*(2), 75–85. https://doi.org/10.1007/s11409-014-9114-2.
Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution, 4*(2), 133–142.
Neal, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational & Graphical Statistics, 9*(2), 249–265.
Orbanz, P., & Teh, Y. (2010). Bayesian nonparametric models. *Encyclopedia of Machine Learning, 1*.
Pardos, Z. A., & Heffernan, N. T. (2011). KT-IDEM: Introducing item difficulty to the knowledge tracing model. In J. A. Konstan, R. Conejo, J. L. Marzo, & N. Oliver (Eds.), *User modeling, adaption and personalization* (pp. 243–254). Springer. https://doi.org/10.1007/978-3-642-22362-4_21.
Pelánek, R. (2016). Applications of the Elo rating system in adaptive educational systems. *Computers & Education, 98*, 169–179. https://doi.org/10.1016/J.COMPEDU.2016.03.017.
Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., & Sohl-Dickstein, J. (2015). Deep knowledge tracing. *Advances in Neural Information Processing Systems, 28*. https://proceedings.neurips.cc/paper/2015/hash/bac9162b47c56fc8a4d2a519803d51b3-Abstract.html.
Rasmussen, C., & Williams, C. (2006). *Gaussian processes for machine learning.* MIT press.
Saint, J., Fan, Y., Gašević, D., & Pardo, A. (2022). Temporally-focused analytics of self-regulated learning: A systematic review of literature. *Computers & Education: Artificial Intelligence, 3*, Article 100060. https://doi.org/10.1016/j.caeai.2022.100060
Saint, J., Whitelock-Wainwright, A., Gašević, D., & Pardo, A. (2020). Trace-SRL: A framework for analysis of microlevel processes of self-regulated learning from trace data. *IEEE Transactions on Learning Technologies, 13*(4), 861–877.
Schellings, G., & Van Hout-Wolters, B. (2011). Measuring strategy use with self-report instruments: Theoretical and empirical considerations. *Metacognition and Learning, 6* (2), 83–90. https://doi.org/10.1007/s11409-011-9081-9
Schielzeth, H., Dingemanse, N. J., Nakagawa, S., Westneat, D. F., Allegue, H., Teplitsky, C., Réale, D., Dochtermann, N. A., Garamszegi, L. Z., & Araya-Ajoy, Y. G. (2020). Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in Ecology and Evolution, 11*(9), 1141–1152. https://doi.org/10.1111/2041-210X.13434
Taub, M., Azevedo, R., Bouchet, F., & Khosravifar, B. (2014). Can the use of cognitive and metacognitive self-regulated learning strategies be predicted by learners' levels of prior knowledge in hypermedia-learning environments? *Computers in Human Behavior, 39*, 356–367. https://doi.org/10.1016/j.chb.2014.07.018
Uzir, N. A., Gašević, D., Jovanović, J., Matcha, W., Lim, L.-A., & Fudge, A. (2020). Analytics of time management and learning strategies for effective online learning in blended environments. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 392–401.
Veenman, M. V., Hout-Wolters, V., Bernadette, H. A. M., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning, 1*(1), 3–14.
Vygotsky, L. S. (1980). *Mind in society: The development of higher psychological processes.* Harvard University Press.
van Wetering, M., Booij, E., & van Bruggen, W. (2020). *Education in an artificially intelligent world.* Kennisnet.
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. (2017). Learning analytics for self-regulated learning. *Handbook of Learning Analytics*, 241–249.
Winne, P. H., & Baker, R. S. J. D. (2013). The potentials of educational data mining for researching metacognition, motivation and self-regulated learning. *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. (1998). Studying as self-regulated learning. In *Metacognition in educational theory and practice* (pp. 277–304). Lawrence Erlbaum Associates Publishers.
Winne, P. H., & Jamieson-Noel, D. (2002). Exploring students' calibration of self reports about study tactics and achievement. *Contemporary Educational Psychology, 27*(4), 551–572. https://doi.org/10.1016/S0361-476X(02)00006-1
Zhang, J., Shi, X., King, I., & Yeung, D.-Y. (2017). Dynamic key-value memory networks for knowledge tracing. In *Proceedings of the 26th International conference on world wide web* (pp. 765–774). https://doi.org/10.1145/3038912.3052580