Predicting self-regulated learning support needs during learning

Rick (S.H.E.) Dijkstra[1] [0003-3688-5456], Max Hinne[2], Eliane Segers[1,3], Josh Ring[1], and Inge Molenaar[1,3]

[1] Behavioural Science Institute, Radboud University, Nijmegen, The Netherlands
`{rick.dijkstra@, eliane.segers@, josh.ring@donders., inge.molenaar@}ru.nl`
[2] Donders Institute for Brain, Cognition, and Behaviour, Radboud University, Nijmegen, The Netherlands `max.hinne@donders.ru.nl`
[3] Dutch National Education Lab AI, Radboud University, Nijmegen, The Netherlands
https://www.ru.nl/en/nolai

Abstract. Adaptive Learning Technologies generate data traces as children interact with them, offering a unique opportunity to estimate self-regulated learning (SRL) support needs. Providing timely, data-driven support for these needs may enhance children's ability to self-regulate their learning and improve learning outcomes. While previous work has identified different levels of SRL support needs using Bayesian non-parametric clustering, these classifications were determined after completing a learning session, delaying potential support. In this study, we present a novel method for identifying children's SRL support needs *during* a learning session, utilizing a Dirichlet-process Gaussian-process mixture model (DPGP). We compared three novel variations of the DPGP model, which were based on the response count (model C), the response time (model T) and the joint combination of response count and response time (model J) of children working on arithmetic problems. Model C predicts SRL support needs with a Matthews' correlation coefficient of 0.75 after children solve the first 48 out of an average of 71 problems. Model T slightly improves on this score by needing 47 problems. However, combining response count and response time improves prediction efficiency, achieving the same performance after 33 problems - less than half the session. Our findings demonstrate that real-time identification of SRL support needs is feasible and effective. This work opens new possibilities for enhancing personalized, online learning experiences by enabling timely, data-driven support tailored to each child's needs.

Keywords: Clustering, Gaussian Processes, Self-Regulated Learning, Prediction, Personalized support

1 Introduction

The daily use of adaptive learning technologies (ALT) in primary schools is increasing [36]. This development creates advanced opportunities to use AI to optimize children's learning with personalized support [25, 3]. ALTs support children using data collected during learning, but current classroom applications primarily focus on improving children's cognitive abilities and knowledge development [14]. However, children also need self-regulated learning (SRL) support, which emphasizes monitoring and controlling their learning [37]. Recent studies have shown that data collected by ALTs can be leveraged to both accurately measure children's SRL [31, 40] and detect their SRL support needs [7].

There are two challenges regarding SRL support that this paper addresses: first, current ALTs estimate children's cognitive ability level during learning [14], but the ALTs do not translate these into SRL support needs. Second, SRL support needs are typically estimated only after learning has concluded, even though research shows that children benefit from personalized SRL support during learning [8, 13, 24, 27]. Such SRL support is currently not feasible with ALTs used in classroom settings, as these ALTs focus on the cognitive processes. To overcome these challenges, this study predicts SRL support needs with a Bayesian non-parametric clustering model based on ALT trace data. This novel model uniquely allows real-time, data-driven estimates of children's SRL support needs during learning.

1.1 Background

According to SRL theory, learning is a goal-directed process that involves conscious decision-making to achieve the desired outcomes [1]. To monitor and control their learning, children use cognitive actions, such as summarizing, rereading, and elaborating, to study a topic or skill and metacognitive actions, such as orienting, planning, monitoring, and evaluating [9]. The COPES model (Conditions, Operations, Performance, Evaluations and Standards) [38] provides a theoretical framework for SRL, suggesting that when children effectively regulate themselves, they monitor their actions to determine how these actions contribute to their goals [2]. When children notice that their actions do not contribute to their goals, they adjust their cognitive actions (strategies or tactics) accordingly using metacognitive actions [37]. Within the COPES model, cognitive and metacognitive actions are together indicative of self-regulated learning, which has been associated with improved learning outcomes [38]. However, children do not apply SRL naturally and need training and support to engage in appropriate cognitive and metacognitive actions [23, 17]. Research indicates that personalized support effectively improves SRL [26].

Personalized SRL support relies on accurate measurements of SRL during learning. SRL can be inferred from trace data from ALTs [7] collected during learning [29], such as response count and correctness. These trace data are used to estimate an ability score that quantifies learners' knowledge development. New problems are selected based on this score, ensuring challenges with the appropriate difficulty for each learner [14]. Previous research demonstrated that the development of these estimates of ability scores over time (the ability curve) is helpful in estimating the children's SRL support needs [7].

In addition to the ability curve, the time a child takes to answer each problem (response time) may also indicate their SRL support needs. Figure 1 shows two plots with the same three ability curves from children working on math problems. The three curves are plotted with the ability score estimates on the y-axis. The left plot has response count on the x-axis, the right plot has response time. When looking at the response count, the orange (dashed) line seems to be rising faster in the beginning, whereas the purple (dash-dot-dot) and pink (solid) lines rise at a similar pace. However, when looking at time spent solving problems, the purple and orange lines initially rise at a similar pace, with the purple line lagging. Additionally, the difference between the plots shows that problems are not solved at an average pace. For example, the child belonging to the purple line grows in spurts, and then is stuck for longer times.

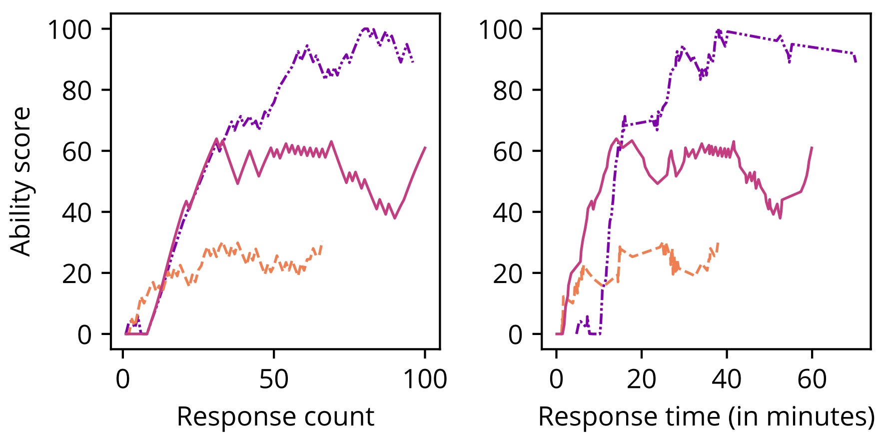


Fig. 1. Example of how changing the x-axis from response count to response time changes the shape of the ability curve. Where the orange (dashed) ability curve rises first when plotting over response count, it is the pink (solid) ability curve that rises first when plotting over response time.

Response time has been shown to correlate with children's affective and metacognitive state [33], predict a child's performance [14, 34, 28], and improve the accuracy of knowledge tracing models [19]. Changes in children's strategies or tactics can be inferred from the time they spend solving each problem; however, limited research has explored the relationship between response time and SRL processes [16, 35].

1.2 Study overview

In this study, we investigated whether it was possible to predict SRL support needs from children in real time, while using trace data from ALTs that are focussed on estimating cognitive abilities.

To estimate the SRL support needs of 5th-grade children during learning, we constructed ability curves from children working on arithmetic problems based on the number of problems solved (response count) and the time spent solving the problems (response time). We extended the model from [7], which estimates SRL support needs after learning. The original approach used in [7] is a Bayesian clustering model that uses the ability curves to find group-wise differences in the ability estimate score progression. SRL support needs for each cluster were determined in combination with learning outcomes and process measures such as accuracy and effort.

The current study enhanced the Bayesian clustering model in two ways. First, we revised the process of identifying the best-fitting cluster for an ability curve. Instead of relying on the entire ability curve – only available after learning—the model was modified to predict SRL support needs with only the part of the data that was available while learning was still taking place. Reducing the amount of data required for accurate cluster prediction enables earlier, real-time inference of the type of SRL support children need.

Response time was integrated as an additional feature to enhance the predictive power of the clustering model. Consequently, three distinct models were developed to capture the development of ability curves through different lenses: A count-based model (C), focused on the response count; a time-based model (T), centred on the response time; and a joint model (J), which integrates variables from both C and T. Since the three models were trained on different inputs, they produced different clusters. These clusters were interpreted on SRL support needed using the children's scores on a pre-test, a post-test and the learning gain; as well as their effort and accuracy the same learning metrics as in [7].

2 Methods

2.1 Data

The data have been collected in an experiment in a classroom setting. In total, 194 children between the ages of 10-13 in grade 5 of elementary school were asked to work with the ALT, which they work with daily [10]. In the first three days, each lesson consisted of instruction about a specific learning goal; subsequently, children practised with this learning goal. The three learning goals covered similar topics (simplifying fractions) but were of increasing difficulty. On the fourth day, the children could select which learning goals they would continue to practice for an additional 30 minutes. A pre-test was administered at the start of the first day, and a post-test was administered at the end of the fourth day. The total data set collected consists of 37,628 responses to problems, with an average of 70.8 problems solved per child per learning goal (23.5 sd.). We preprocess the response time in this data set by removing values that are further than two standard deviations away from the mean and resetting them to the mean value. A subset of these data has previously been used by [7] to obtain a post-hoc clustering of learners.

2.2 Clustering model

The foundation of our SRL support-needs prediction algorithm is a probabilistic clustering model known as a Dirichlet process-Gaussian process model (DPGP model; [22, 7]). The DPGP model identifies clusters of sequential data, i.e., the ability curves collected via the ALT, without pre-determining the number of clusters or their properties (e.g., slope, variance) of the sequential data.

The DPGP model is a mixture model, which means it assumes that data originate from a combination of different sources. A typical example of a mixture model is the Gaussian mixture model (GMM), the probabilistic equivalent of K-means clustering [4]. The GMM represents each cluster as a Gaussian distribution with a mean and a variance and a probabilistic assignment of each data point to a cluster.

In typical GMMs, data are numerical, but they do not a priori contain additional structure. In our case, however, our observations (that is, the ability curves) are sequential. Therefore, instead of using a Gaussian distribution for each cluster, we use a *Gaussian process* [22]. Now, the distribution of a single cluster is represented by a combination of mean and covariance *functions* [30].

More formally, the model is defined as follows. First, our data consist of children's ability curves. We use the following notation: For each ability curve a, we have N sequential observations of p dimensions. These dimensions reflect either the response count (for the count-based model, which we refer to as C), the response time (for the time-based model, T), or both (for the joint model, J). Let $\mathbf{x}_i^{(a)} \in \mathbb{R}^p$ be the input for problem i, and collectively let $\mathbf{X}^{(a)} \in \mathbb{R}^{N \times p} = \left(\mathbf{x}_1^{(a)}, \ldots, \mathbf{x}_N^{(a)}\right)^\top$. The response, consisting of the actual ability scores, is denoted by the vector $\mathbf{y}^{(a)} \in \mathbb{R}^N$. Lastly, we use the notation $z^{(a)} = c$ to indicate that ability curve a is assigned to cluster c. In the remainder, to avoid clutter, we will drop the superscript (a) when no confusion is likely to arise.

Modelling the expected ability curve The mean function for each cluster c describes the expected ability level. The function incorporates the assumption that children might not initially know the topic, but once they do, they will grow to a certain ability level with a certain speed. To reflect these assumptions, we introduce a halved and scaled sigmoid function, defined for a uni-dimensional input as:

$$\mu_c\left(x_i \mid \kappa_c, \lambda_c, \hat{y}_c\right) = \max\left(0, \hat{y}_c\left(2\left(1 + \exp\left(-\frac{(x_i - \kappa_c)}{\lambda_c}\right)\right)^{-1} - 1\right)\right) \quad . \tag{1}$$

Here, κ_c represents the moment when children start giving correct answers, which is the point at which their ability level starts to increase from 0; λ_c represents how fast a child increases in ability; $\hat{y}_c$ represents the maximum ability level that children reach in cluster c. In contrast to [7], we treat $\hat{y}_c$ as a latent variable. This way, not all clusters are expected to reach the maximum ability score in the end, more closely representing the situation in the classroom, where not all children reach the same level of ability. For the two-dimensional input, we sum the contribution of the mean function over each dimension and rescale them as $\mu_c(\mathbf{x}_i \mid \cdot) = \frac{1}{2}\mu_c(x_{ip} \mid \cdot) + \frac{1}{2}\mu_c(x_{it} \mid \cdot)$.

The kernel function for each cluster incorporates the assumption that inputs that are close together will have more similar outputs. This assumption is reflected in the Radial Basis Function:

$$k_c\left(x_i, x_j \mid \tau_c, \ell_c\right) = \tau_c^2 \exp\left(-\frac{\|x_i - x_j\|^2}{2\ell_c^2}\right) \quad . \tag{2}$$

Here, ℓ_c represents the correlation between responses y_i en y_j based on the Euclidean distance between the corresponding inputs x_i and x_j. When ℓ_c is small, only inputs that are very close result in correlated outputs, so that the ability curve can change rapidly. In contrast, a large ℓ_c means that outputs are strongly correlated, making the ability curves look smooth. Lastly, τ_c determines the amount of vertical fluctuation in the expected ability curve. For the two-dimensional input, we sum the contribution of the kernel for each dimension as $k_c(x_i, x_j \mid \cdot) = k_c(x_{ip}, x_{jp} \mid \cdot) + k_c(x_{it}, x_{jt} \mid \cdot)$

Using expressions (1) and (2), we can compute the probability of an observed ability curve $\mathbf{y}^{(a)}$ given a cluster c using

$$p\left(\mathbf{y}^{(a)} \mid \kappa_c, \lambda_c, \hat{y}_c, \tau_c, \ell_c, \mathbf{x}^{(a)}\right) = \mathcal{N}\left(\boldsymbol{\mu}_c, \mathbf{K}_c + \sigma_c^2\mathbf{I}\right) \quad , \tag{3}$$

where $\boldsymbol{\mu}_c = (\mu_c(x_1), \ldots, \mu_c(x_N))^\top$ with $\mu_c(x_i) = \mu_c(x_i^{(a)} \mid \kappa_c, \lambda_c, \hat{y}_c)$, $\{\mathbf{K}_c\}_{ij} = k_c\left(\mathbf{x}_i, \mathbf{x}_j \mid \tau_c, \ell_c\right)$ (for more details on this expression and its derivation, we refer to [30, 7]).

Modelling the clustering In standard mixture models like the GMM mentioned above, the number of clusters must be specified beforehand. Instead, we use a Dirichlet process as a nonparametric prior to learn the number of required clusters from our observations. Let $\mathbf{w} = (w_1, \ldots, w_C)^\top$, be a vector of weights for all C possible clusters, such that $\sum_{i=1}^{C} w_i = 1$ and $w_i \geq 0, \forall i$. However, as C should be adapted to the data, the vector $\mathbf{w}$ changes dimensionality during inference, which is problematic. Instead, we place a Dirichlet process prior on $\mathbf{w}$ and subsequently marginalise over this parameter, leaving us with the distribution of cluster assignments $p(z^{(a)} = c)$, which does not change dimensionality [22, 18].

Using Bayes' rule, we can calculate the probability that a curve a belongs to cluster c (that is, $P\left(z^{(a)} = c\right)$) instead of the other clusters using the weighted proportion of the probability of observing data from the ability curve if it is part of cluster c.

$$P\left(z^{(a)} = c \mid \mathbf{y}^{(a)}, \mathbf{X}^{(a)}, \theta, \mathbf{w}\right) = \frac{w_c p\left(\mathbf{y}^{(a)} \mid \theta_c, \mathbf{X}^{(a)}\right)}{\sum_{c'=1}^{C} w_{c'} p\left(\mathbf{y}^{(a)} \mid \theta_{c'}, \mathbf{X}^{(a)}\right)} \quad , \tag{4}$$

where $\theta = (\ell, \tau, \lambda, \kappa, \hat{y})$ Inferring the distribution of the values for the weights ($\mathbf{w}$) and the clustering parameters (θ) is achieved using adaptive tempered sequential Monte Carlo sampling [11, 6, 12].

2.3 Predicting cluster assignment

Real-time prediction means that in contrast to clustering, after all data have been collected, we predict the cluster label while the children are still working on their tasks. Let $\mathbf{y}_{t<h}^{(a)}$, $\mathbf{X}_{t<h}^{(a)}$ contain all data points for ability curve a up until time point h, the latest time point at which data came in during real-time learning. The probability that this partial ability curve will be assigned to a cluster is then computed using Equation 4 using only the observations available up until that point; $p\left(z_a = c \mid \mathbf{y}_{t<h}^{(a)}, \mathbf{X}_{t<h}^{(a)}, \theta, \mathbf{w}\right)$. We select the cluster with the highest probability $\hat{c}_h$ as the predicted cluster for that partial ability curve.

To evaluate the performance of these predictions, we interpret this as a classification task, where the true label is set as the cluster assignment given all data $\hat{c}_T$. We measure each cluster's performance using Matthews correlation coefficient (MCC) [21]. The MCC is a reliable statistical measure for unbalanced data sets, such as when clusters do not have to have the same size; it only gives high values when the sensitivity, specificity, precision, and the negative predictive value of the predictions are high [39, 5].

2.4 Interpreting clusters

After obtaining the clusters for the count-based, time-based and joint models, we select the model with the best performance on the prediction task. We link this model to SRL support needs, following [7], by comparing the learning process measures and learning outcomes of the children assigned to each cluster. The learning process measures are accuracy, effort, and the cluster's mean curve characteristics. Accuracy is measured as the percentage of problems solved correctly. Effort is calculated as the number of problems solved.

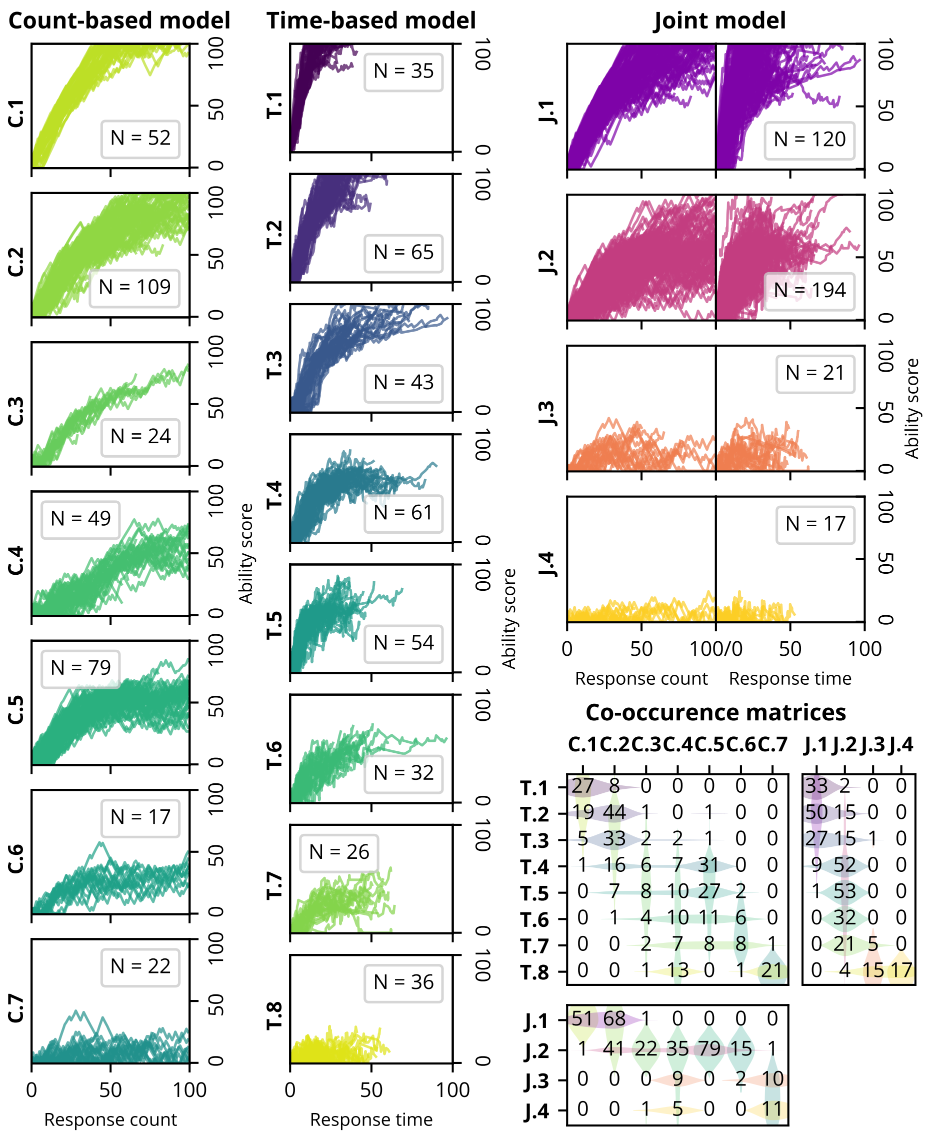
The learning outcomes are the children's scores on the pre-tests and post-tests, as well as the relative gain between the two tests, which we measure with the normalised learning change [20]. The normalised learning change is defined as $\frac{\text{Post}-\text{pre}}{\text{pre}_{\text{max}}-\text{pre}}$ when the child has improved, $\frac{\text{Post}-\text{pre}}{\text{pre}}$ if post-test scores are lower than pre-test scores, and 0 if the child already has the maximum score (of 8) on the pre-test, or if the pre-test and post-test are equal [20].

To see whether the learning outcomes and process measures of the children differ significantly between clusters, we run a linear mixed model in R using lmerTest [15] and, if so, t-tests with corrections using Satterthwaite's method [32] to determine which clusters differ significantly from each other on which measures.

3 Results

3.1 Number of clusters identified

Clusters were identified using the count-based (C), the time-based (T), and the joint model (J). The count-based model identified 7 clusters, the time-based



	C.1	C.2	C.3	C.4	C.5	C.6	C.7	J.1	J.2	J.3	J.4
T.1	27	8	0	0	0	0	0	33	2	0	0
T.2	19	44	1	0	1	0	0	50	15	0	0
T.3	5	33	2	2	1	0	0	27	15	1	0
T.4	1	16	6	7	31	0	0	9	52	0	0
T.5	0	7	8	10	27	2	0	1	53	0	0
T.6	0	1	4	10	11	6	0	0	32	0	0
T.7	0	0	2	7	8	8	1	0	21	5	0
T.8	0	0	1	13	0	1	21	0	4	15	17

	C.1	C.2	C.3	C.4	C.5	C.6	C.7
J.1	51	68	1	0	0	0	0
J.2	1	41	22	35	79	15	1
J.3	0	0	0	9	0	2	10
J.4	0	0	1	5	0	0	11

Fig. 2. From left to right, the clusters that were found for respectively the count-based (C), time-based (T) and joint (J) models. In the bottom right, a matrix describing how often curves co-occur in clusters. Since the joint model has two input dimensions, the plots are shown as two adjoined plots. The co-occurrence matrix shows, for each cluster in one model, how much of the ability curves assigned to that cluster are also assigned to the clusters of the other model.

model identified 8, and the joint model identified 4 clusters. These clusters are visualised in Figure 2, showing the full ability curves assigned to these clusters.

The clusters are sorted by their average ability score after solving 75 problems. The similarity between the S, T, and C model, is shown by the co-occurrence matrix in the bottom right in Figure 2. Each cell in the matrix represents the number of ability curves that occur in both clusters from the row and the column. For most clusters, the co-occurrence of their ability curves with clusters from the other models is spread over different clusters. The diagonal patterns in the matrix shows that the average accuracy score (since, that is what we sorted the clusters on) of ability curves is related to determining their cluster.

3.2 Prediction accuracy of cluster models

Prediction is successful when a model assigns a partial ability curve to the same cluster as the entire ability curve. We evaluate prediction performance both as a function of the response count and of the response time to determine how quickly the model converges to a stable cluster assignment. We use Matthews correlation coefficient (MCC), which ranges between -1 and 1 and gives a high score only if the sensitivity, specificity, precision and the negative predictive value of the predictions are high as well. We take the MCC value of 0.75 as reference point.

The top plots in Figure 3 show that the count-based model (C) takes 48 problems out of the average of 71 problems solved before reaching an MCC value of 0.75. For the time-based model (T), it takes 24 and 22 seconds out of an average of 35 minutes. However, the number of problems solved for the joint model (J) takes only 33 out of 71 problems, which is also reached after 15 minutes and 13 seconds out of 35 minutes. Halfway through the learning process (after 36 problems or 17.5 minutes), the model C reaches an MCC value of 0.61, the time-based model an MCC value of 0.61 as well, and the joint model an MCC value of 0.78 for half the problems solved and 0.77 for half the time spent solving problems, respectively. In general, the count-based and time-based models have similar accuracy development; however, the joint model outperforms the other two models.

Since the joint model is the best performing model, we further analyse the individual clusters of this model. As shown in the middle plots in Figure 3, the best performing cluster, J.1, reaches an MCC value of 0.75 after 24 problems out of the average of 72 problems, followed by cluster J.4 after 28 problems, cluster J.2 after 35 problems and cluster J.3 only reaches this score after 43 of the 72 average problems solved. This shows that the clusters have a different development in prediction performance. The bottom plots in Figure 3 show that the poorer performance of cluster J.3 is primarily due to the curves of this cluster being predicted as either cluster J.2 or J.4 early on.

3.3 Interpreting SRL support needs based on the joint model

The clusters in the joint clustering model can be related to the learning metrics collected from the children working on the three skills. These are pre and post-test scores, relative learning gain between these tests, accuracy as measured by the percentage of problems solved correctly, and effort as measured by the

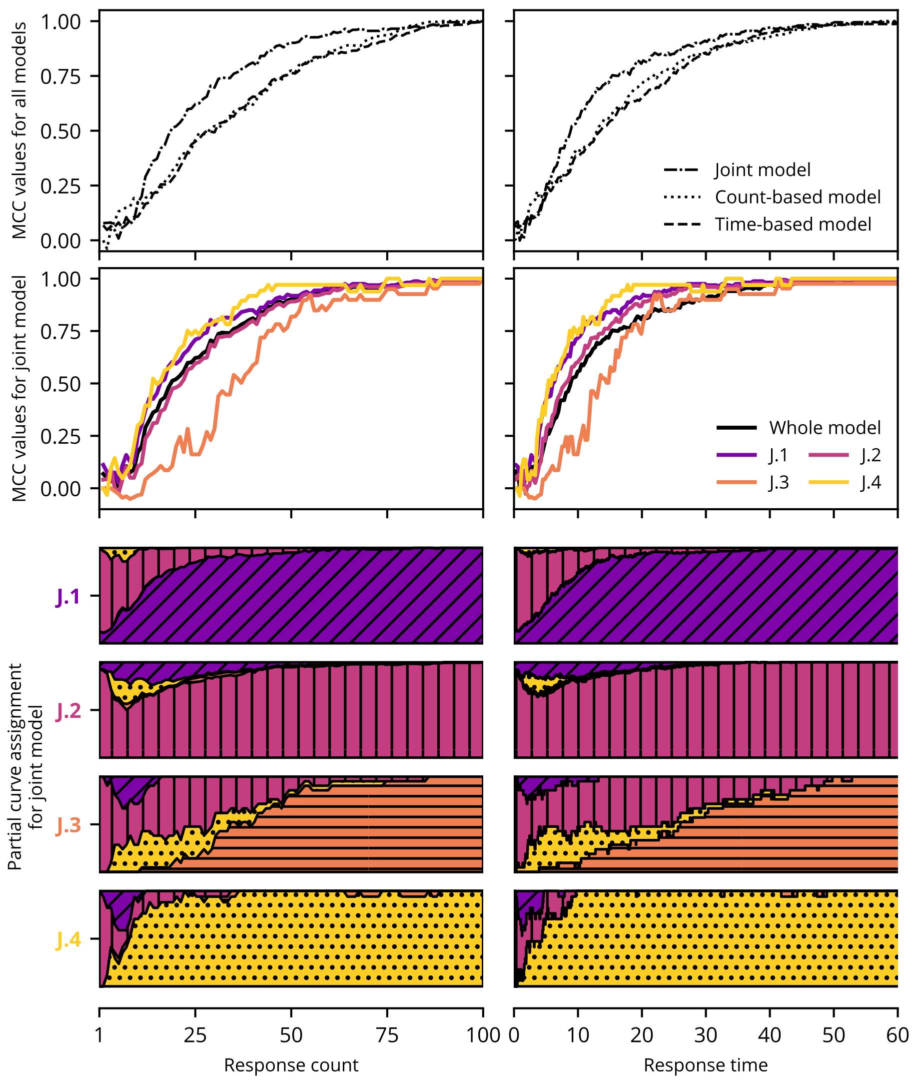


Fig. 3. The MCC values as a function of response count(left) and response time (right) for the three different models (top) show the joint model outperforming the rest. The MCC values for the individual clusters of the joint model (middle) show the performance of J.3 falling behind the other clusters. The proportion of the predicted cluster assignment (bottom) shows that most curves in J.3 are erroneously predicted as J.2 and J.4. All plots on the right have been cut off at 60 minutes as not much change is made afterwards.

number of problems solved. As shown in Figure 4, there is a clear and significant difference in accuracy between all clusters. The other metrics have a significant

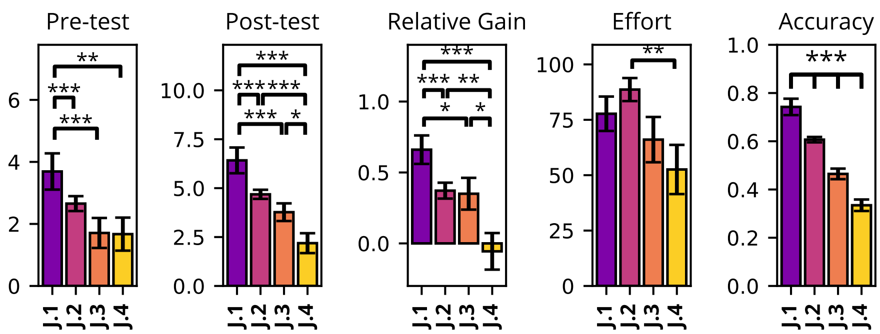


Fig. 4. The learning metrics for the clusters identified by the joint model. Significance level is indicated with asterisks: * = p<0.05; ** = p<0.01; *** = p<0.001. On pre-test score cluster J.1 outperforms the other clusters. On post-test J.1 outperforms the other clusters and J.2 and J.3 outperform J.4. The same is the case for Relative Gain. For effort only J.2 outperforms J.4, whereas for accuracy all earlier clusters outperform the later ones.

difference between some of the clusters. Detailed results of the tests are available on GitHub[4].

Cluster J.1 demonstrates the highest accuracy of all clusters. Its pre-test and post-test scores are significantly higher than all other clusters, showing that children in this cluster had mastered the skill to a large extent before learning began. The ability curves assigned to this cluster reach the system's highest possible ability level. Following [7], we call this cluster the *Masters* cluster. Children in this cluster appear not to need SRL support, as they either mastered the skill before practice or can master it independently.

Cluster J.2 has the second-best accuracy of the clusters, with the post-test scores and relative gains being significantly higher than those of cluster J.4 but lower than cluster J.1. It also exhibits significantly higher effort compared to cluster J.4. Generally, the ability curves assigned to this cluster show an upward trend in ability score, but do not reach the maximum score. Therefore, we refer to the children in this cluster as *Risers*. Risers may require moderate SRL support to achieve mastery.

Children in cluster J.3 have significantly lower pre-test scores, post-test scores, and relative gains than those in cluster J.1, as well as a significantly lower accuracy than children both in cluster J.1 and J.2. However, they achieve significantly higher post-test scores, relative gains, and accuracy than children in cluster J.4. Additionally, children in this cluster do not exceed above half of the maximum possible ability level. Therefore, we refer to these children as *Strugglers*, and they may need substantial SRL support.

[4] https://github.com/anonymousAIED2025/AIED2025_Appendix

Finally, the children in cluster J.4 have significantly lower post-test scores, relative learning gains, and accuracy than all other clusters. They also exhibit significantly lower pre-test scores than the children in cluster J.1 and significantly lower effort than those in cluster J.2. Additionally, their ability score does not exceed 20% of the maximum possible ability level. This stagnating ability score shows they are *Trailers* and may need extensive SRL support for this learning goal.

4 Discussion

In the current study, we investigated the use of a Dirichlet process Gaussian process model to identify clusters related to children's SRL support needs and evaluate the model's predictive power in the context of real-time data. Our findings showed that the model's predictive performance is best when incorporating both the response count and the response time. This model resulted in four clusters that indicate different SRL support needs, ranging from the *Masters* cluster that seems to need no SRL support to the *Trailers* cluster, which seems to require extensive support. Notably, these clusters can be accurately predicted after 33 responses or 15 minutes (out of an average of 71 responses or 35 minutes in a lesson).

The improved prediction due to integrating the response time implies that both data streams contain information needed to predict the child's SRL support needs. This improvement demonstrates, as shown in [19], that incorporating additional information, such as response time, increases the strength of models that estimate children's abilities.

Predicting the cluster best fitting the child's ability curve enables early intervention in children's learning. If we predict that a child may not reach their goal, the ALT can adjust the SRL support to their needs. This support can provide more monitoring information, suggest control strategies, or take over regulation by fine-tuning beyond the current adaptivity of the difficulty level used to select the following problem. If effective, directing support to children to improve self-regulated learning may impact their regulation of practice behaviour and learning outcomes [37]. Future studies can investigate different types of SRL support.

A limitation of this study is that response time is calculated as the interval between solving two problems. However, this interval only indirectly reflects response time. The child could have been receiving additional instructions, have been writing solutions on a scratch path to understand the problem further, have been distracted, or there could have been other reasons why they were not actively engaged in the task. Further work could be done to determine which response times are outliers and which are valid.

The prediction performance after 15 minutes shows that the joint model can be used in real-time in a classroom situation. Additionally, some clusters can already be predicted earlier, making it possible to target some children with earlier support. Support can be done, for example, with personalized visualizations to guide children through the phases of self-regulated learning described by the COPES model, resulting in personalized adaptive support for children when they need it.

5 Acknowledgement

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

References

1. Azevedo, R.: Understanding the complex nature of self-regulatory processes in learning with computer-based learning environments: an introduction. Metacognition and Learning **2**(2), 57–65 (2007). https://doi.org/10.1007/S11409-007-9018-5
2. Azevedo, R.: Theoretical, conceptual, methodological, and instructional issues in research on metacognition and self-regulated learning: A discussion. Metacognition and Learning **4**(1), 87–95 (2009). https://doi.org/10.1007/S11409-009-9035-7
3. Baker, R.S.: Artificial intelligence in education: Bringing it all together. Digital education outlook: Pushing the frontiers with AI, blockchain, and robots pp. 43–54 (2021). https://doi.org/10.1787/589b283f-en
4. Bishop, C.M., Nasrabadi, N.M.: Pattern recognition and machine learning, vol. 4. Springer (2006). https://doi.org/10.1117/1.2819119
5. Chicco, D., Jurman, G.: The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification. BioData Mining **16**(1), 4 (2023). https://doi.org/10.1186/s13040-023-00322-4
6. Chopin, N., Papaspiliopoulos, O., et al.: An introduction to sequential Monte Carlo, vol. 4. Springer (2020). https://doi.org/https://doi.org/10.1007/978-3-030-47845-2
7. Dijkstra, R., Hinne, M., Segers, E., Molenaar, I.: Clustering children's learning behaviour to identify self-regulated learning support needs. Computers in Human Behavior p. 107754 (2023). https://doi.org/10.1016/j.chb.2023.107754
8. Greene, J.A., Azevedo, R.: The measurement of learners' self-regulated cognitive and metacognitive processes while using computer-based learning environments. Educational Psychologist **45**(4), 203–209 (2010). https://doi.org/10.1080/00461520.2010.515935
9. Greene, J.A., Azevedo, R.: 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 (2007). https://doi.org/10.2190/G7M1-2734-3JRR-8033
10. Gynzy: Rekenen. Educational software (2025), https://www.gynzy.com/en, accessed February 17, 2025
11. Herbst, E., Schorfheide, F.: Sequential monte carlo sampling for dsge models. Journal of Applied Econometrics **29**(7), 1073–1098 (2014). https://doi.org/10.1002/jae.2397
12. Hinne, M.: An introduction to sequential monte carlo for bayesian inference and model comparison—with examples for psychology and behavioral science. Behavior Research Methods **57**(5), 125 (2025)
13. Järvelä, S., Järvenoja, H., Malmberg, J., Hadwin, A.F.: Exploring Socially Shared Regulation in the Context of Collaboration. Journal of Cognitive Education and Psychology **12**(3), 267–286 (2013). https://doi.org/10.1891/1945-8959.12.3.267

14. Klinkenberg, S., Straatemeier, M., Van Der Maas, H.L.J.: Computer adaptive practice of Maths ability using a new item response model for on the fly ability and difficulty estimation. Computers & Education **57**, 1813–1824 (2011). https://doi.org/10.1016/j.compedu.2011.02.003
15. Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B.: lmerTest Package: Tests in Linear Mixed Effects Models. Journal of Statistical Software **82**(13), 1–26 (2017). https://doi.org/10.18637/jss.v082.i13
16. Li, S., Zheng, J., Poitras, E., Lajoie, S.: The Allocation of Time Matters to Students' Performance in Clinical Reasoning. In: Nkambou, R., Azevedo, R., Vassileva, J. (eds.) Intelligent Tutoring Systems. Lecture Notes in Computer Science, vol. 10858, pp. 110–119. Springer International Publishing, Cham (2018). https://doi.org/10.1007/978-3-319-91464-0_11
17. Li, T., Fan, Y., Tan, Y., Wang, Y., Singh, S., Li, X., Raković, M., van der Graaf, J., Lim, L., Yang, B., Molenaar, I., Bannert, M., Moore, J., Swiecki, Z., Tsai, Y.S., Shaffer, D.W., Gašević, D.: Analytics of self-regulated learning scaffolding: effects on learning processes. Frontiers in psychology **14** (2023). https://doi.org/10.3389/fpsyg.2023.1206696
18. Li, Y., Schofield, E., Gönen, M.: A tutorial on dirichlet process mixture modeling. Journal of Mathematical Psychology **91**, 128–144 (2019). https://doi.org/10.1016/j.jmp.2019.04.004
19. Lin, C., Shen, S., Chi, M.: Incorporating student response time and tutor instructional interventions into student modeling. In: Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization. p. 157–161. UMAP '16, Association for Computing Machinery, New York, NY, USA (2016). https://doi.org/10.1145/2930238.2930291
20. Marx, J.D., Cummings, K.: Normalized change. American Journal of Physics **75**(1), 87–91 (2007). https://doi.org/10.1119/1.2372468
21. Matthews, B.: Comparison of the predicted and observed secondary structure of t4 phage lysozyme. Biochimica et Biophysica Acta (BBA) - Protein Structure **405**(2), 442–451 (1975). https://doi.org/https://doi.org/10.1016/0005-2795(75)90109-9
22. McDowell, I.C., Manandhar, D., Vockley, C.M., Schmid, A.K., Reddy, T.E., Engelhardt, B.E.: Clustering gene expression time series data using an infinite Gaussian process mixture model. PLoS Computational Biology **14**(1) (2018). https://doi.org/10.1371/journal.pcbi.1005896
23. Miller, C.J., Bernacki, M.L.: Training preparatory mathematics students to be high ability self-regulators: Comparative and case-study analyses of impact on learning behavior and achievement. High Ability Studies **30**(1-2), 167–197 (2019). https://doi.org/10.1080/13598139.2019.1568829
24. Molenaar, I.: Advances in temporal analysis in learning and instruction. Frontline Learning Research **2**(4), 15–24 (2014). https://doi.org/10.14786/flr.v2i4.118
25. Molenaar, I.: Towards hybrid human-AI learning technologies. European Journal of Education **57**(4), 632–645 (2022). https://doi.org/10.1111/ejed.12527
26. Molenaar, I., Horvers, A., Dijkstra, R., Baker, R.S.: Personalized visualizations to promote young learners' srl: The learning path app. In: Proceedings of the tenth international conference on learning analytics & knowledge. pp. 330–339 (2020). https://doi.org/10.1145/3375462.3375465
27. Molenaar, I., Järvelä, S.: Sequential and temporal characteristics of self and socially regulated learning. Metacognition and Learning **9**, 75–85 (2014). https://doi.org/10.1007/s11409-014-9114-2

28. Papamitsiou, Z., Economides, A.A., Pappas, I.O., Giannakos, M.N.: Explaining learning performance using response-time, self-regulation and satisfaction from content: an fsQCA approach. In: Proceedings of the 8th International Conference on Learning Analytics and Knowledge. pp. 181–190. ACM, Sydney New South Wales Australia (2018). https://doi.org/10.1145/3170358.3170397
29. Plass, J., Pawar, S.: Toward a taxonomy of adaptivity for learning. Journal of Research on Technology in Education **52**, 275–300 (2020). https://doi.org/10.1080/15391523.2020.1719943
30. Rasmussen, C., Williams, C.: Gaussian processes for machine learning. MIT press, Cambridge, MA (2006). https://doi.org/10.7551/mitpress/3206.001.0001
31. Saint, J., Fan, Y., Gašević, D., Pardo, A.: Temporally-focused analytics of self-regulated learning: A systematic review of literature. Computers and Education: Artificial Intelligence **3**, 100060 (2022). https://doi.org/10.1016/j.caeai.2022.100060
32. Satterthwaite, F.E.: An approximate distribution of estimates of variance components. Biometrics Bulletin **2**(6), 110–114 (1946), http://www.jstor.org/stable/3002019
33. Shih, B., Koedinger, K.R., Scheines, R.: A response-time model for bottom-out hints as worked examples. In: Handbook of Educational Data Mining, pp. 201–212. CRC Press (2010)
34. Susac, A.N., Bubic, A., Kaponja, J., Planinic, M., Palmovic, M.: Eye movements reveal students'strategies in simple equation solving. International Journal of Science and Mathematics Education **12**, 555–577 (2014). https://doi.org/10.1007/s10763-014-9514-4
35. Taub, M., Azevedo, R., Bradbury, A.E., Millar, G.C., Lester, J.: Using sequence mining to reveal the efficiency in scientific reasoning during STEM learning with a game-based learning environment. Learning and Instruction **54**, 93–103 (2018). https://doi.org/10.1016/j.learninstruc.2017.08.005
36. van Wetering, M., Booij, E., van Bruggen, W.: Education in an artificially intelligent world. Tech. rep., Kennisnet (2020)
37. Winne, P.H.: Learning analytics for self-regulated learning. Handbook of learning analytics pp. 241–249 (2017). https://doi.org/10.18608/hla17.021
38. Winne, P.H., Hadwin, A.F.: Studying as self-regulated learning. In: Metacognition in educational theory and practice, pp. 277–304. The educational psychology series, Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US (1998). https://doi.org/10.4324/9781410602350-19
39. Yao, J., Shepperd, M.: Assessing software defection prediction performance: Why using the matthews correlation coefficient matters. In: Proceedings of the 24th International Conference on Evaluation and Assessment in Software Engineering. pp. 120–129 (2020). https://doi.org/10.1145/3383219.3383232
40. Zhang, T., Taub, M., Chen, Z.: A Multi-Level Trace Clustering Analysis Scheme for Measuring Students' Self-Regulated Learning Behavior in a Mastery-Based Online Learning Environment. In: LAK22: 12th International Learning Analytics and Knowledge Conference. pp. 197–207. ACM, Online USA (2022). https://doi.org/10.1145/3506860.3506887