

A Multi-Level Trace Clustering Analysis Scheme for Measuring Students' Self-Regulated Learning Behavior in a Mastery-Based Online Learning Environment

Tom Zhang

Tom.Zhang@ucf.edu

University of Central Florida
Orlando, Florida, USA

Michelle Taub

Michelle.Taub@ucf.edu

University of Central Florida
Orlando, Florida, USA

Zhongzhou Chen

Zhongzhou.Chen@ucf.edu

University of Central Florida
Orlando, Florida, USA

ABSTRACT

This study introduces a new analysis scheme to analyze trace data and visualize students' self-regulated learning strategies in a mastery-based online learning modules platform. The pedagogical design of the platform resulted in fewer event types and less variability in student trace data. The current analysis scheme overcomes those challenges by conducting three levels of clustering analysis. On the event level, mixture-model fitting is employed to distinguish between abnormally short and normal assessment attempts and study events. On the module level, trace level clustering is performed with three different methods for generating distance metrics between traces, with the best performing output used in the next step. On the sequence level, trace level clustering is performed on top of module-level clusters to reveal students' change of learning strategy over time. We demonstrated that distance metrics generated based on learning theory produced better clustering results than pure data-driven or hybrid methods. The analysis showed that most students started the semester with productive learning strategies, but a significant fraction shifted to a multitude of less productive strategies in response to increasing content difficulty and stress. The observations could prompt instructors to rethink conventional course structure and implement interventions to improve self-regulation at optimal times.

CCS CONCEPTS

• Applied computing → Distance learning; E-learning; Physics.

KEYWORDS

Self-regulated learning, Online learning environments, Click-stream data

ACM Reference Format:

Tom Zhang, Michelle Taub, and Zhongzhou Chen. 2022. 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 (LAK22), March 21–25, 2022, Online, USA*. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3506860.3506887>



This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike International 4.0 License](https://creativecommons.org/licenses/by-nc-sa/4.0/).

LAK22, March 21–25, 2022, Online, USA
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9573-1/22/03.
<https://doi.org/10.1145/3506860.3506887>

1 INTRODUCTION

Studying students' self-regulated learning (SRL) behavior in online learning environments through analysis of trace data has become one of the focuses of learning analytics research, especially since the COVID pandemic forced educators around the world to abruptly switch to online learning. SRL identifies students as playing active roles in their learning by engaging in planning, monitoring, and reflection strategies during learning [22, 25], which is imperative for online learning as students are less reliant on instructors for in-the-moment assistance and are required to engage in more independent, structured, and regimented learning to keep up with class activities and content mastery. A number of recent studies have employed multiple data mining techniques such as sequence pattern analysis, process mining, and hierarchical clustering to identify, visualize and compare students' SRL strategies from trace data [5, 14], with the latest example being the Trace-SRL framework developed by Saint et. al. [18]. The Trace-SRL framework first maps event sequences onto micro-SRL processes based on an SRL model [26], then identifies SRL strategies by conducting two levels of sequence clustering on SRL action traces, and finally creates process models of SRL learning strategies for different groups of students.

The data set analyzed in Saint et. al. contains a rich variety of events from attempting problems to accessing the scheduling page or the dashboard. The online learning environment also provided students with a relatively high level of freedom to access different course contents in their preferred order. Under those conditions, different micro-level SRL actions will likely result in distinct event sequences, which allowed researchers to create a "dictionary" that uniquely maps event sequences onto SRL actions. The relatively high variability in event sequences also enables the dissimilarity between different traces to be calculated based on transition rates and state frequencies observed in the data set, which is the default option provided in mainstream trace clustering tools [7].

However, online learning systems or online learning data sets are oftentimes more restrictive, meaning that they provide fewer event types and less variability in event sequences. This could happen for several reasons, first, certain restrictions could be beneficial for learning from a pedagogical standpoint. For example, in a mastery-based learning system designed based on the principles of deliberate practice [4], students would be required to complete a sequence of tasks in a pre-determined order. From a "preparation for future learning" perspective [20], requiring students to attempt a problem first could improve the level of learning from subsequent study of the instructional materials. Second, online homework or intelligent tutoring systems are often integrated into an online course hosted

on a separate learning management system such as Canvas. In those cases, data from the sub-system may only contain a subset of events such as attempts and page access, whereas other types of events such as checking the dashboard are stored in a different data set that may not be available to the researcher.

For those more restrictive data sets, direct application of existing analysis schemes faces two prominent challenges. First, it will be challenging to reliably identify one-to-one correspondence between event sequences and microscopic SRL strategies. This is because, on one hand, students adopting different learning strategies may end up producing similar or even identical event sequences due to restrictions imposed by the system. On the other hand, each event or event sequence may be reflective of more than one SRL actions. For example, if a student accessed the learning material after a failed problem-solving attempt, the decision could indicate that the student reflected on the previous attempt, but may also be part of the students' initial plan to learn the skills needed to solve the problem or a combination of both. Second, sequence clustering based entirely on the frequencies of events in the data set could be insufficient in identifying key differences in students' learning strategies in a restricted online system. For example, a student who first attempted to solve a problem and then guessed on subsequent attempts likely had different strategies than those who started guessing on attempt one. Yet there is no guarantee that the frequency differences contained in the data itself will be sufficient to separate those two states, especially since the variability in trace sequences in a restricted system is markedly smaller. Both of those challenges are present in the analysis of all types of online learning trace data sets but are more prominent in limited data sets from restrictive systems.

Therefore, in this study we will try to answer the following questions: can we adopt and modify similar strategies and algorithms such as those used in the Trace-SRL framework to analyze students' learning data from a restrictive learning system? To what extent can this approach still provide valuable insight into students' self-regulation strategies, learning behaviors, as well as goals and motivations, without the need for directly mapping event sequences onto unique SRL actions?

Below, we will first briefly introduce the design of the mastery-based online learning modules (OLM) system being analyzed in the current study. We then outline our attempt to develop an analysis scheme suitable for trace data produced from the OLM system, by employing three different clustering algorithms on three consecutive levels of data granularity.

1.1 Design of OLM and OLM Sequences

Designed based on principles of mastery-learning [2] and deliberate practice [20], an OLM combines assessment, instruction, and practice into a single online learning unit. Each OLM is focused on explaining one or two basic concepts, or developing the skills to solve one kind of problem, designed to be completed in 5 to 30 minutes. Each individual OLM consists of an assessment component (AC) which tests students' content mastery in 1-2 questions, and an instructional component (IC) with instructional text and practice problems on the topic (see Figure 1). Upon accessing a module, students are shown the learning objectives of the current module

and are required to make an initial attempt on the AC before being allowed to access the IC. Students can make additional attempts on the AC at any time after the first attempt and are not required to access the IC. This design is motivated by both the "mastery-learning" format that allows students who are already familiar with the content to proceed quickly to the next assignment and by the concept of "preparation for future learning" intending to improve students' learning from the IC by exposing them to the questions first. It also allows researchers to more easily measure knowledge transfer between subsequent modules [21].

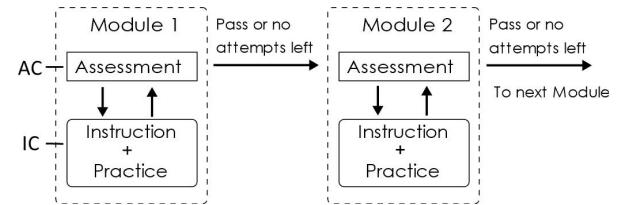


Figure 1: Schematic illustration of OLM design, adapted from [3]

A number of OLM modules form an OLM sequence on a more general topic (e.g., conservation of mechanical energy) and students are required to pass the AC or use up all attempts before moving onto the next OLM in that sequence. A typical OLM sequence consists of 5-12 modules and are assigned as self-study homework for students to complete over a period of one to two weeks.

1.2 Analysis of OLM Trace Data via Multi-Level Clustering

The current analysis scheme involves three consecutive clustering operations performed on the trace data from the OLM system at the event level, module level, and sequence level.

1.2.1 Level I: Clustering of Individual Events. Prior research on OLMs has shown that students' use of learning strategies may not be reflected by the order of events, but rather by the quality of events [3, 8]. More specifically, an abnormally short assessment attempt is likely the result of random guessing or answer copying and may be indicative of the student adopting a performance-avoidance goal. Therefore, the main goal of the event level clustering is to distinguish between abnormally short guessing attempts and longer problem-solving attempts. This is achieved by fitting the log-duration distribution with finite mixture modeling (FMM) [16], which can be seen as clustering based on a single continuous variable (event duration). The same technique is also applied to identify and exclude very short study events which likely originated from a student clicking through the instructional contents without meaningfully interacting with them.

1.2.2 Level II: Clustering Interaction Traces on Individual Modules. To identify the main strategies that students adopt when interacting with individual OLM modules, we partition students' event traces on a single module into multiple "module-level clusters", or m-clusters, using hierarchical agglomerative clustering. We will test and compare three different techniques for determining the dissimilarity between different traces:

- (1) A purely data-based strategy using optimal matching distance similar to the method used in [12, 18], and relying solely on frequencies observed from the data.
- (2) A hybrid strategy with which part of the dissimilarity calculation is informed by the design of the platform.
- (3) A purely theory-based strategy that determines the dissimilarity based on a set of features that are deemed to be important in the context of an OLM, informed by an SRL framework adapted for restrictive learning systems.

For each resulting m-cluster, a process map is generated via exploratory process mining using the package *bupaR* [10] with the top 60% of most frequent traces, as an intuitive visual representation of the main characteristics of each interaction strategy.

1.2.3 Level III: Clustering of Interaction Traces on OLM Sequences. As a result of module-level clustering, students' use of learning strategies for each OLM sequence can be captured as a trace of m-cluster memberships. We can then partition those traces into sequence-level clusters, or s-clusters, to capture the main strategies employed by each student to interact with the entire OLM sequence. In Trace-SRL, student-level clustering is based on the frequency of adopting different strategies using the Euclidean distance. In the current analysis, the s-clusters are identified based on the optimal matching distance between traces calculated using the *TraMineR* package [7]. Using the optimal matching distance preserves the temporal information in trace data, which will enable us to investigate when students change their interaction strategies in response to different factors such as content difficulty.

Due to the computational complexity of this analysis scheme, the current study will analyze trace data collected from three representative OLM sequences selected from the beginning, middle, and end of the semester, to answer the following three research questions:

- RQ1: Which of the three clustering strategies (Data, Hybrid, Theory) will produce the most well organized and interpretable m-clusters?
- RQ2: What are the main strategies that students adopted to interact with OLMs, and how are those strategies related to their assessment performance?
- RQ3: How do students' choice of strategies change over time, and how can we interpret their use of strategy in the framework of SRL?

2 METHODS

2.1 OLM Design and Implementation

The OLM modules are created and hosted on the Obojobo learning objects platform, an open-source online learning platform developed by the Center for Distributed Learning at the University of Central Florida. In the current iteration, the assessment component of each OLM contains 1-2 multiple choice problems and permits a maximum of 5 attempts. The first 3 attempts are sets of isomorphic problems assessing the same content knowledge with different surface features or numbers. On the 4th and 5th attempts, students are presented with the same problems in the 1st and 2nd attempts respectively and are awarded 90% of credit. The instructional component of each module contains a variety of learning resources

including text, figures, videos, and practice problems. Each OLM sequence contains 3-12 OLMs, which students must complete in the order given, with completion defined as either passing the assessment or using up all 5 attempts. Readers can access example OLMs at <https://canvas.instructure.com/courses/1726856>.

2.2 Instructional Conditions and Data Collection

Data used in this study was collected from a calculus-based university introductory physics course taught in the Fall 2020 semester. A total of 251 students were initially enrolled in the class, of which 67 were females, and 98 were historically underrepresented minorities. The course was taught asynchronously using pre-recorded lecture videos as the main method for content delivery as a result of the COVID pandemic. Students and instructors interacted via messages, posts, and short video conferences via Microsoft Teams (for more information on the course design, see <https://www.aaas-iuse.org/resource/course-design/>). Students were required to take a total of 7 quizzes during the semester, with each quiz lasting 20 minutes and proctored by a teaching assistant.

A total of 70 OLMs consisting of 9 sequences were assigned as the only form of online homework. Each OLM sequence was assigned for students to complete over 1-2 weeks, with all OLMs in a sequence due on the same day. The OLM modules accounted for 24% of course credit, with no additional homework assignments. Students could earn extra credits by completing some OLMs earlier than the due date, as explained in more detail in [6].

2.3 Collection and Analysis of Student Log Data

Students' click-stream log data collected from the Obojobo platform is first processed into attempt events and study events. An attempt event starts when the student enters the assessment page of the module, and ends when the student clicks the submit button on the assessment page. During this period the student is unable to navigate to any other pages in the current module or to other modules. The duration of the attempt event is the time between those two clicks minus the duration of 1) when the browser window is either closed or minimized, or when another window is in focus and 2) any non-active duration beyond 10 minutes. An attempt event is labeled as "pass" only if the student correctly answers all questions in the assessment, and "fail" otherwise.

A study event starts when the student clicks on any page in the instructional component of the module, and ends when the student clicks on the last record before a new attempt event is initiated, or the last record for the student on the module. In other words, a study event includes all the interaction with the instructional component between two attempt events. The duration of the study event is calculated as the sum of all the time spent interacting with each instructional page, minus the duration of inactive periods as explained above.

For the current study, we selected event data from Sequence 1: Motion in 1 Dimension, Sequence 6: Mechanical Energy, and Sequence 9: Angular Momentum. The three sequences consist of a total of 26 modules, and the resulting data set contains a total of 5960 traces. In addition, all records after the first passing attempt or

after the last attempt were truncated for simplicity of analysis, since there were significantly fewer records after passing or using up all attempts, and most of those events took place before an exam [3]. The topics of the three sequences are sufficiently different so that the learning outcome of earlier sequences should not significantly impact the learning behavior of later sequences.

2.4 Clustering Level I: Categorization of Events

At level I, abnormally short attempts on the AC of a given OLM are distinguished from normal AC attempts, by fitting the log duration distribution of all attempts on a single module using FMM. FMM is a model-based clustering algorithm that divides a population into subgroups according to one or more observable characteristics, by fitting the distribution of characteristics with a finite mixture of normal or skewed probability distributions. When two or more distinct problem-solving behaviors are present, the log attempt duration distribution can be fitted with the sum of two or more distributions, with the shortest distribution corresponding to abnormally short attempts. In the current study, we fit the log-duration of each assessment attempt using either normal or skewed distribution models using the R package *mixsmsn* [16], following the fitting procedure described in detail in a previous study [3]. In the case when a single component distributed is the best fit for the duration, then the cutoff is set as either 2 standard deviations below the mean duration, or 15 seconds, whichever is longer. This is because a previous clinical study indicated that attempts under 15s are likely to arise from complete random guessing [9].

We also conducted mixture-model fitting of the combined log duration of all study events from all modules in the data set, to determine the cutoff time between normal study events and very short study events that are likely the result of students clicking through the instructional pages. Study events that are shorter than the cutoff are excluded from the data set since those events are more likely the result of accidental or unintentional clicks rather than a deliberate decision to engage with the material.

2.5 Level II: Module-Level Trace Clustering

As a result of Level I clustering, each students' interaction with a given OLM is represented by a trace of either normal or short attempt events and study events that are longer than the minimal duration. Each attempt is treated as a separate event and labeled as "Attempt_N" with N being the attempt number. Short attempts are labeled as "Attempt_N_S" to distinguish them from normal attempts. For example, a trace of {Attempt_1_S, Study, Attempt_2, Attempt_3} indicates that the student took 3 attempts on the OLM, with the first attempt being a short attempt, and took a study session (longer than the minimum cutoff) between attempts 1 and 2.

Hierarchical agglomerative clustering using Ward's method is performed via the R package *cluster* [13] on traces from all three selected OLM sequences, with each trace treated as a data entry. We investigate and compare three different kinds of strategies for calculating the distance between two traces: purely data-driven strategies, hybrid strategies, and purely theory-based strategies, explain in detail below. The resulting module-level clusters, or m-clusters, reflect different learning strategies that students adopt to interact with individual OLMs.

2.5.1 Data-Driven Trace Distances. The distance metric used for data-driven strategy is the Levenshtein distance, defined as the minimum number of operations needed to transform a given string a into another b via single letter operations (insertion, deletion, and substitution). The matrices generated by the cost generation methods determine the relative weight of substitution between letters in the alphabet¹, while the indel costs refer to the relative weight of insertion/deletion of letters in the alphabet.

There exists three cost generation methods native to the *TraMineR* package: the INDEL, TRATE, and FUTURE methods which utilize *only* information observed from the data. The INDEL method generates the cost matrix by calculating the relative frequency of each state, the TRATE method by transition rates between adjacent states in a given trace, and the FUTURE method by the Chi-Square distance for the state probability vectors. A more detailed description of the methods can be found in the documentation of the *TraMineR* package [7]. All three methods are being tested for the current analysis, and the one that generates the most well-formed clusters (explained in more detail below) is selected as the result of the pure data-driven strategy.

2.5.2 Hybrid Trace Distance. The local structure in the TRATE most closely matches the theorized actions associated with SRL behaviors out of the three data-based methods. As such it requires the least modification of the dissimilarity inputs to better match and subsequently separate such predicted behaviors. To do this, we constructed a vector of the insertion/deletion costs of the alphabet with the following constraints in mind:

- (1) Study events after Attempts 1, 2, and 3 are significantly different from study events after Attempts 4 and 5.
- (2) Brief Attempts are significantly different from Normal Attempts.
- (3) The first Brief Attempt is significantly different from the first Normal Attempt.

2.5.3 Theory-Based Trace Distance. For each OLM, students are presented with two tasks: a required task to solve the problem in the assessment component, and an optional task to study the learning material. Therefore, students will need to make two types of decisions:

- (1) Whether to seriously engage in problem solving on a given attempt (resulting in a normal length attempt) or to make a guess (usually resulting in a short attempt).
- (2) Whether to engage with the study material if the first attempt fails.

Using the three macro-level self-regulatory phases presented in Zimmerman's model of SRL [24], we propose six features that capture students' interactions with the OLMs and their associated SRL processes, based partly on two earlier studies [3, 23]:

- (1) Total Number of Assessment Attempts (nA): This reflects the quality of **performance** on both the problem-solving task and the studying task. In general, passing on fewer attempts reflects either high incoming knowledge or successful learning, or both.

¹Here, alphabet refers to the available range of letters in the state space.

- (2) Number of Attempts before Study² (nY): When or whether to access learning materials can be influenced by either **planning** or **reflection**. While accessing the learning materials after the first mandatory attempt is more likely the result of planning, accessing after multiple failed attempts is more likely the result of reflecting on previous attempt performance.
- (3) Fraction of Short Attempts Among All Attempts (fS): Since most short attempts likely originate from either guessing or copying behaviors, a high fraction of short attempts could indicate low prior knowledge, low self-efficacy, low effort, or limited execution strategies like time management.
- (4) Is the First Attempt Short (1S): The first attempt is of particular significance as it reflects students' **planning** of the time and effort to spend on the mandatory first attempt. A short first attempt can be a sign that the student is trying to conserve time and effort on the task by guessing. It could also be influenced by the student's **reflection** on experience in the previous modules and their perceived self-efficacy within the course. A student with lower confidence, or is aware of their low prior knowledge, will be more likely to make a short first attempt to access the learning materials.
- (5) Is the Last Attempt Short (1S): The last attempt is also of particular significance since it is the passing attempt in all event traces, except for those with five failed attempts. A short final passing attempt signifies lower **performance** on both learning and problem-solving tasks, and probably lower levels of **reflection** on learning.
- (6) Did the Student Abort the Module (Ab): This feature represents a small number of (22 out of 5960) event traces that ended on a failed attempt that is not attempt 5 from the rest of the event traces. Those traces exist either because the student aborted the module, or because of corrupted data logs.

It is worth noting that the number of times a student accessed the instructional materials is not included as a feature. This is because, in 90% of cases, students interact with all the learning materials at once and in most other cases, one study event is significantly longer than the others. It is also much less clear whether a higher degree of access reflects higher or lower quality of learning.

Since features 1, 2, and 3 are numeric while features 4, 5, and 6 are binary, the distance metric between two event traces is computed using the Gower dissimilarity coefficient. We tested multiple sets of feature weights for calculating the Gower coefficient. The best cluster structure, as explained below, was produced when the weights for nA and 1S are set to 0.5 and all other weights set to 1.0, which emphasizes the forethought phase.

2.5.4 Determining the optimal number of m-clusters. Since agglomerative clustering produces a tree structure of all possible numbers of clusters, we choose to determine the optimum number of m-clusters based on the average silhouette value of each cluster. A

²If a student never accessed the instructional materials, this feature has a value of 5, which is equivalent to having studied after the 5th attempt. Since event logs after the passing attempt of 5th attempt are excluded from the current data set, a value of 5 always represents a no-study trace. The choice of coding amplifies the decision of whether to access the study materials when interacting with the OLMs.

silhouette is a measure of the ratio of intra- and inter-cluster variability which is described in [17]. The term "average silhouette" refers to the average silhouette value for every point in the clustering algorithm. Theoretically, the optimal number of clusters is chosen to maximize the average silhouette, as it indicates that the variability within clusters is very small compared to the variability between clusters, thus being well defined.

However, in practice, the current data set of 5960 traces contains only 53 unique traces. As a result, the average silhouette will always reach the global maximum at or near 53 clusters. Several clustering strategies resulted in local maxima of average silhouette in under 10 clusters, indicating a relatively well-defined cluster structure with a small number of clusters. In those cases, the number of m-clusters is chosen using the local maximum average silhouette.

2.5.5 Visualizing m-Cluster Structure using Process Maps. The main characteristic of each m-cluster is visualized by creating process maps (PMs) using the R package processmapR [11] for event traces in each cluster. A "PASS" event is appended to the end of the trace if the last attempt is a passing attempt. Otherwise, the sequence ends with the last attempt. The top 60% of most frequent event traces in each m-cluster are included in the map to reduce "spaghetti effects" caused by rare traces.

2.6 Level III: Sequence-Level Trace Clustering

Following module-level clustering, a student's event trace interacting with each OLM is classified as belonging to one of several m-clusters. Thus, the student's interaction with an OLM sequence of n modules can be captured by a sequence-level trace of n elements in the form of $\{m_1, m_2, \dots, m_n\}$, with each element m_i being a number representing the m-cluster that the student's event trace on module i belongs to.

We again perform hierarchical agglomerative trace clustering on the sequence-level traces for each of the three OLM sequences separately using the Ward method. The distance between two traces is calculated using the optimal matching distance via the TRATE method, as it takes into account the local ordering of states. Since each student contributes one trace per sequence to the data set, the s-clusters are a reflection of the strategy adopted by individual students.

The number of s-clusters for each sequence is determined by the maximal average silhouette value between 2 and 10 clusters. In the case that the maximum average silhouette is 2 clusters, but a local maximum exists for a higher number of clusters, then the higher number of clusters is selected in order to better elicit relatively rare but distinct interaction strategies.

2.7 Correlation With Assessment Outcomes

To explore the correlation between identified s-clusters and students' performance on assessment, we divided the students into three tertiles of near equal size along quiz score performance (High, Medium, and Low) following the practice in [12]. Quiz scores were chosen over exam scores since quizzes were being proctored and exams were not during the Fall 2020 semester. We then compared the frequency distribution of s-cluster membership between the three cohorts using Fisher's exact test. When a significant difference was observed for a given sequence, we then performed post hoc

analysis via pairwise Fisher exact tests for every s -cluster between every two performance cohorts, using the Benjamini, Hochberg, and Yekutieli ("fdr") method [1] for p-value adjustment.

3 RESULTS

3.0.1 Event Level Clustering. Of the 26 module assessments included in this data set, the attempt duration distribution of 8 assessments were fitted with 1 component FMMs, and the rest are all fitted with 2 or more components FMMs. For 4 assessments, the short versus normal cutoffs as determined by FMM modeling were less than 15s and were thus adjusted to 15s. Twenty-one modules had short versus normal cutoffs between 15 and 120s, and 2 modules had cutoffs beyond 120s. Assessments involving numerical calculation problems had longer cutoffs compared to those involving conceptual questions.

3.1 Module-Level Trace Clustering

3.1.1 Comparison of Different Clustering Methods. The silhouette plot for each of the three methods (data-based, hybrid, and theory-driven) is shown in Figure 2. The theory-based method outperformed the other two methods with a local maximum average silhouette of over 0.8 at 7 clusters. The purely data-driven method had a local maximum of 0.7 at three clusters, while the average silhouette of the hybrid method did not reach a local maximum in under 10 clusters.

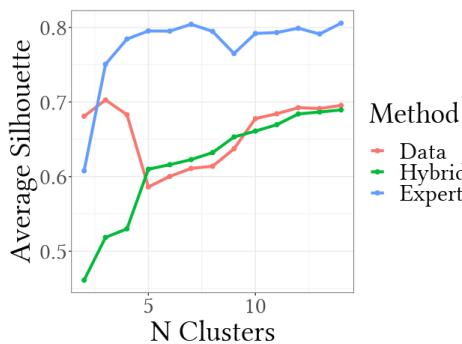


Figure 2: Plots of the Average Silhouettes for each distance generation method for m-Clusters

For all the clusters identified by either the pure data or the hybrid methods, the corresponding process maps (PMs) show that traces both with and without a study event are grouped into the same cluster, as shown in the example PM in Figure 3. Furthermore, the pure data method also cannot distinguish between normal and short assessment attempts. Since both methods failed to distinguish features important for interpreting students' SRL strategies, we chose to only show one example PM for each method.

The PMs for the m-clusters generated by the theory-based method are shown in figure 4, labeled using the shorthand notation explained in the figure caption. The PMs can be divided into two

groups based on the "lengths" of the PMs: Traces in PMs 1, 2, 4, and 7 all reach the PASS state within 3 attempts. PMs 3, 5, and 6 all contain traces that pass on either the 4th or 5th attempt, or did not reach the pass state after the 5th attempt.

In PMs 1, 2, 4, and 7, the majority of cases passed the module on the 1st or 2nd attempt, with a smaller fraction (< 20%) passing on the 3rd attempt. As shown in the PMs, most traces in m-cluster 1 (A) passed the module without studying the learning material whereas those in m-cluster 2 (AYAn) passed the module in a few attempts after studying. m-cluster 7 (SYAn) is similar to m-cluster 2, except that the initial attempt is a short one. The passing attempts in all three PMs are all normal attempts. All of the traces in m-cluster 4 (S), on the other hand, passed the module on short attempts with no study event, and at least 60% of the traces passed on only one short attempt, as shown in the PM. It is likely that the traces in m-cluster 4 (S) originated from the student either making a lucky guess or obtaining the answer from another source.

A significant common feature of PMs 3, 5, and 6 is that all the passing attempts are short attempts. M-cluster 3 (Sn) started with a normal attempt, followed by multiple short attempts without a study event in most traces. In m-clusters 5 (AYSn) and 6 (SYSn), a study event followed either a normal or a short first attempt. In those three clusters, it is likely that students are guessing on the assessments after either having taken the first attempt (3) or after studying the instructional materials (5, 6).

Since the theory-driven distance metric produced m-clusters with both better structure and better interpretability, we chose to use those results in Level III clustering analysis.

3.2 Sequence-Level Clusters

A total of 5, 4, and 6 s -clusters are identified for sequences 1, 6, and 9 respectively according to the average silhouette scores. In figure 5, we visualize the characteristics of each identified s -cluster by plotting the distribution of m-clusters for each OLM in the sequence using stacked bar charts. Each bar chart corresponds to one identified s -cluster and each column corresponds to one OLM in the given sequence. The height of the bars represents the fraction of event traces that belong to one of the seven m-clusters. Here, m-cluster 0 is used to indicate that the student did not interact with the given module. Based on the frequency of observing different m-clusters within each sequence, the s -clusters can be roughly categorized into five types.

Type I: Initial Pass. Traces classified into s -clusters 1-1, 6-1, and 9-1, share the common feature of having predominantly m-cluster 1 (A) on all or most of the modules. S-clusters 6-1 and 9-1 also had a smaller fraction of m-cluster 4 (S) which in the case of s -cluster 6-1 became dominant in the last 4 modules.

Type II: Pass or Study. Traces classified into s -clusters 1-2, 1-3, 1-4, 6-2, and 9-2 share the common feature of having a combination of m-clusters 1 (A) and 2 (AYAn). S-clusters 1-2 and 1-3 are different in that 1-2 had higher fractions of m-cluster 1 (A) on the first couple of modules whereas the distribution in 1-3 is roughly uniform and a bit higher in the later modules. S-cluster 1-3 is also smaller than 1-2 and 1-4. S-cluster 1-4 consists of predominantly m-cluster 2 (AYAn) across all modules, which is also the case for s -cluster 6-2. However,

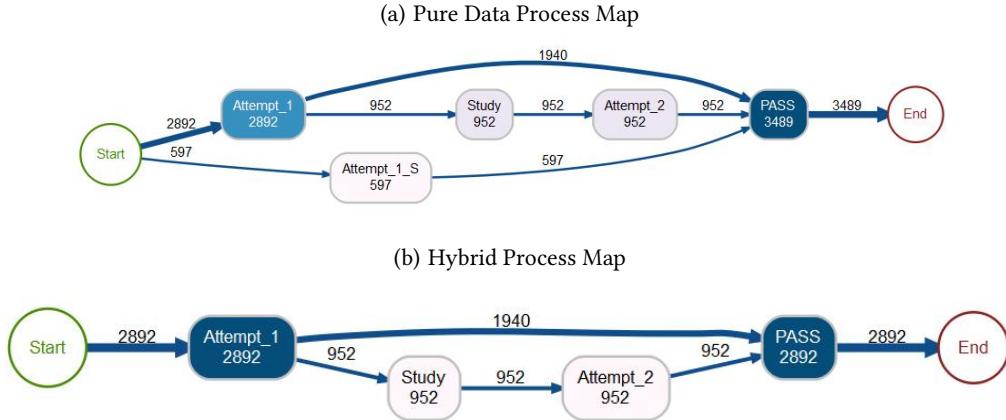


Figure 3: Process maps generated from the (a) INDEL and (b) Hybrid cost matrix generation methods.

on the last 4 modules, of s-cluster 6-2, a significantly higher fraction of m-clusters 5, 6, and 7 are observed.

Type III: Varied Strategies. For traces in s-clusters 6-3, 9-3, and 9-4, there isn't a single dominant m-cluster, but m-clusters 3 (Sn), 5 (AYSn), and 7 (SYAn) are observed at a higher frequency when compared to other s-clusters. m-clusters 5, 6, and 7 all involve both a study event followed by one or more short attempts. A notable difference between those three clusters is that s-cluster 6-3 has a higher fraction of m-cluster 6 (SYSn) and 4 (S) on the last four modules, suggesting ineffective study or giving up without studying. S-clusters 9-3 has a higher fraction of m-cluster 7 (SYAn), which is study and normal attempts after an initial guessing attempt on the mandatory first attempt.

Type IV: Short Pass. Traces in s-clusters 1-5 and 9-5 consist of predominantly, m-cluster 4 (S), short passing on the first attempt, on all modules in the sequence, which strongly suggests some form of answer copying or quick guessing. While s-cluster 1-5 consists of only 2 students, s-cluster 9-5 consists of 91 students and is the single largest cluster in sequence 9.

Type V: Incomplete. The small number of traces in s-clusters 6-4 and 9-6 do not have records on the last couple of modules in the sequences. Note that there are also several unfinished traces in cluster 1-3, but the missing data is predominantly concentrated on the last couple of modules.

3.3 S-cluster membership and Assessment Performance

Fisher tests conducted for each sequence detected significant differences in s-cluster membership among high, medium, and low quiz score cohorts ($p < 0.0001$ for each sequence). For each sequence, the distribution of s-cluster memberships among each quiz score cohort is plotted in figure 6 whereas the result of post hoc analysis is listed in table 1.

4 DISCUSSION

The multi-level hierarchical clustering analysis scheme produced relatively well-organized and highly interpretable clusters at both

s-cluster	H:M	H:L	M:L
1-1	0.0003***	0.037*	ns
1-4	ns	0.0022**	ns
6-2	ns	0.000003***	0.002**
6-3	0.0013*	0.00003***	ns
9-2	0.014*	0.0008***	ns
9-5	0.024*	0.00053***	ns

Table 1: Adjusted p-values for significant differences in s-cluster membership between average quiz score tertiles.

the module level and the sequence level. As one of the centerpieces of the analysis, the purely theory-based distance metric at the module level (Level II) outperformed both data-based and hybrid distance metrics, by producing both better-structured clusters (as judged by average silhouette) and more interpretable outcomes under the framework of SRL (RQ1). A possible implication is that in a restrictive system such as the OLMs, instructional conditions imposed by the system design play a major role in students' decision-making process, but may not be well reflected in the naturally occurring event frequencies. Therefore a strong input of expert knowledge is required to produce well-organized and interpretable clusters.

Results from module-level clustering answered RQ2 by revealing three major m-clusters accounting for about 86% of all traces, together with four smaller m-clusters. The largest m-cluster 1 (A), passing on the initial attempt before study, could indicate that a student entered the OLM with high prior content knowledge, possibly from learning the content from video lectures or in previous courses. However, it is also possible that in some cases students might have engaged in undesirable behaviors like conducting internet searches for problem solutions, which is indistinguishable from a normal attempt based on duration data alone. The second-largest m-cluster is 2 (AYAn), passing after study. Both m-clusters 1 and 2 represent the kinds of effective self-regulatory student behaviors that the

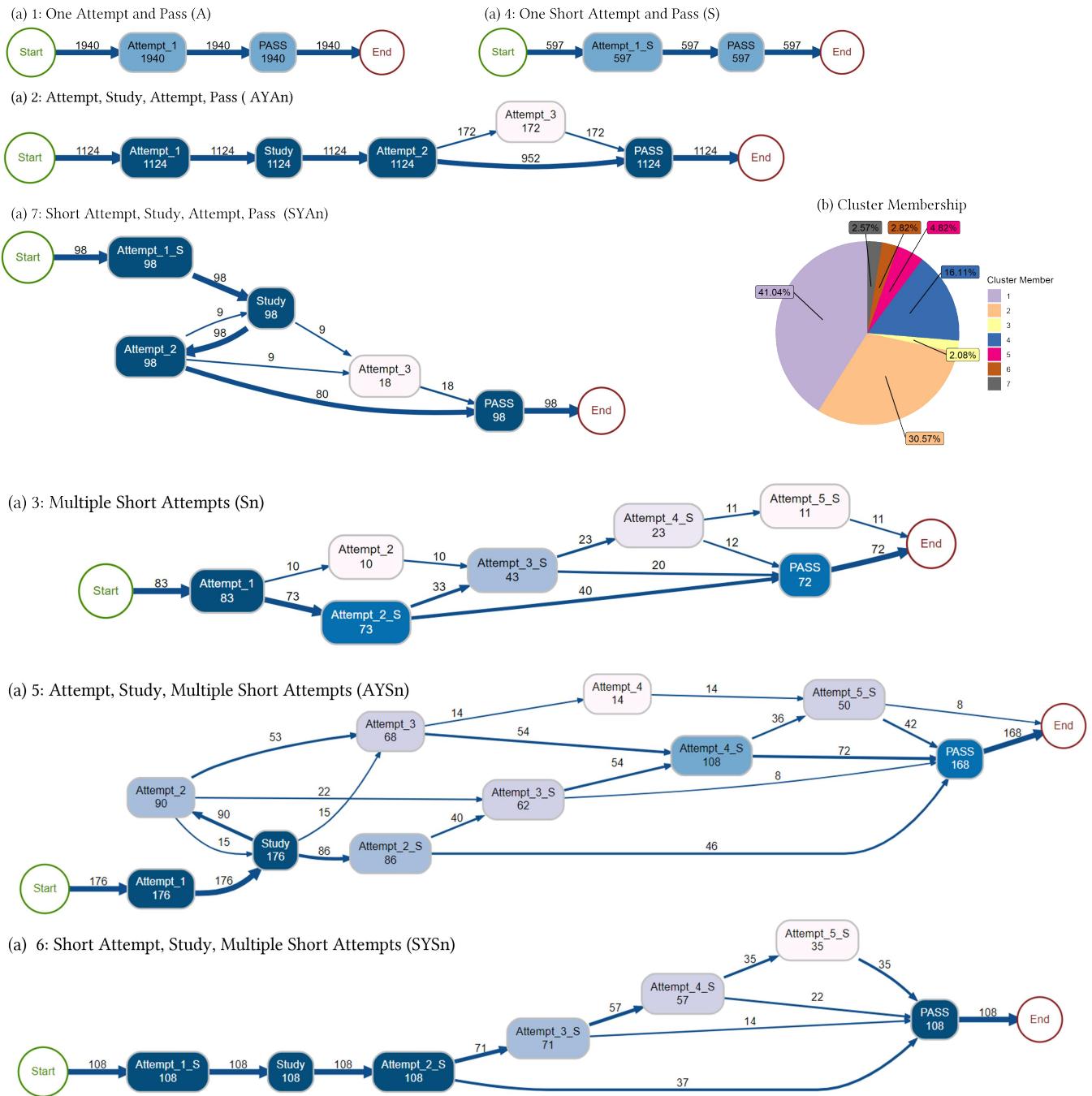


Figure 4: (a) 1-7: The most common 60% of traces in each m-cluster. The shorthand name for each PM captures its main characteristics, using the following convention: A, S, and Y represent an Attempt, a Short Attempt, and a Study event respectively. A letter followed by an n represents multiple occurrences of the event represented by the letter. (b): Pie chart showing the relative cluster membership of the student traces.

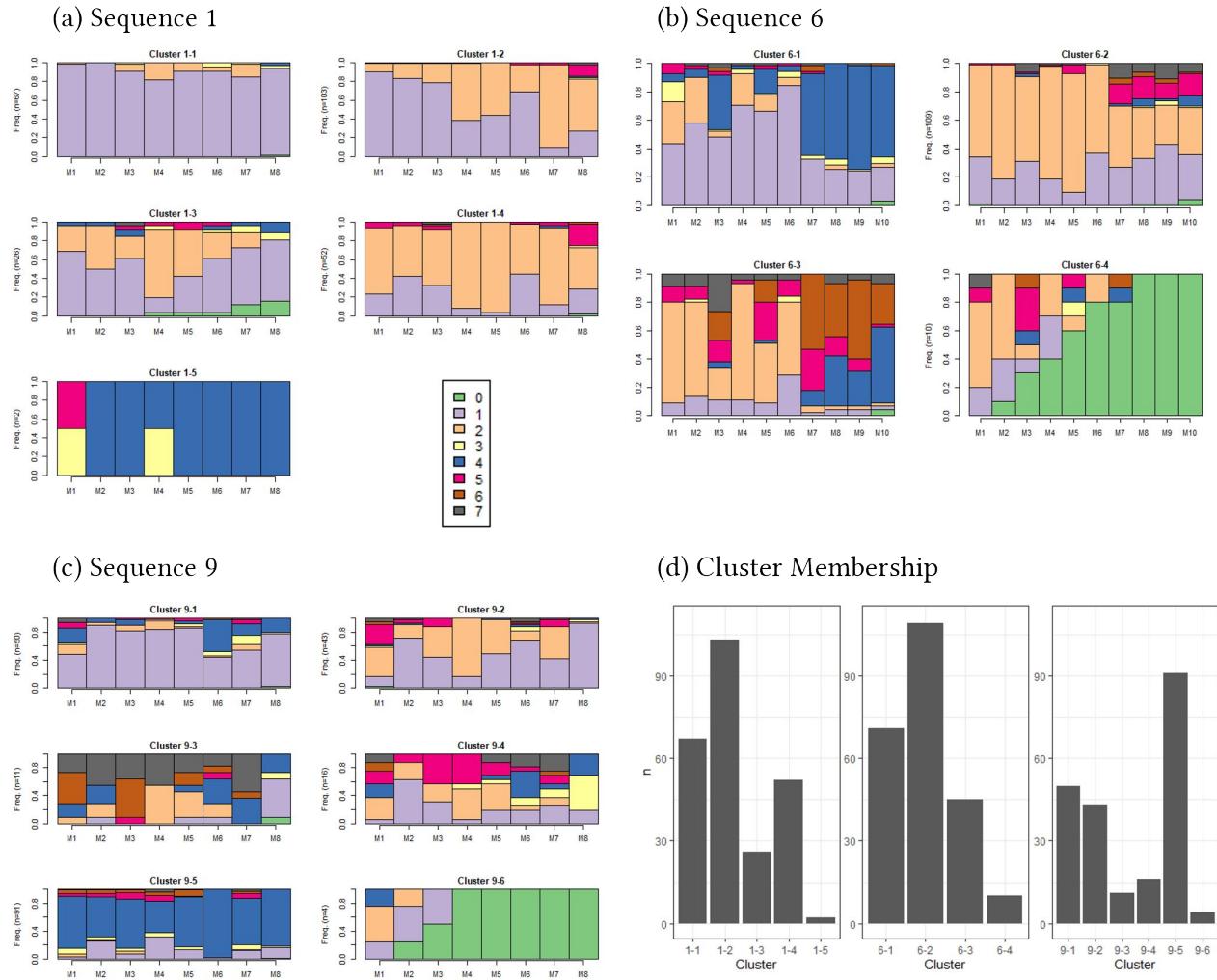


Figure 5: Visualization of the s-clusters of each sequence and overall cluster membership by sequence.

OLMs were designed to elicit. The majority (74.3%) of study events in this cluster took place immediately after the first attempt, showing that students either actively reflected on their problem-solving performance, or planned their learning strategy according to the instructor's recommendation. The third major m-cluster is 4 (S), pass on one brief attempt cluster, which suggests possible answer copying and counter-productive self-regulation during learning. These three m-clusters represent learning strategies frequently observed in all three sequences, while the four minor m-clusters reflect strategies mostly adopted by some students on specific modules and sequences.

The outcomes of the sequence level clustering provided the most insight into students' SRL strategies across multiple modules and how those strategies changed over time (RQ3). As shown in Figure 5, learning strategies employed by students changed significantly across the three sequences. For sequences 1 and 6, the most populous s-clusters were 1-2 and 6-2, both of which contain a mixture of

m-clusters 1 (A) and 2 (AYAn), indicating effective strategy use as intended by the OLM design. However, by the end of the course, the dominant s-cluster for sequence 9 is 9-4, which almost exclusively consists of m-cluster 4 (S), indicating consistent and deliberate answer copying. Strikingly, the same "copy or guess" strategy was only adopted by 2 students in sequence 1 in s-cluster 1-5.

Students also changed their learning strategies within a single OLM sequence. For example in sequence 6, a shift in strategy is observed between modules 6 and 7 for all s-clusters. While the dominant strategy for the first 5 OLMs consists of m-clusters 1 (A) and 2 (AYAn), there was an obvious increase in m-clusters 4 (S), 5(AYSn), and 6 (SYSn) on the last 4 modules. Both m-clusters 5 and 6 are characterized by finishing on multiple short attempts following a study event, which could indicate that those students had difficulty learning from the instructional materials and therefore resorted to guessing on subsequent attempts. This shift in strategy could be explained in part by a change in content difficulty, as OLMs

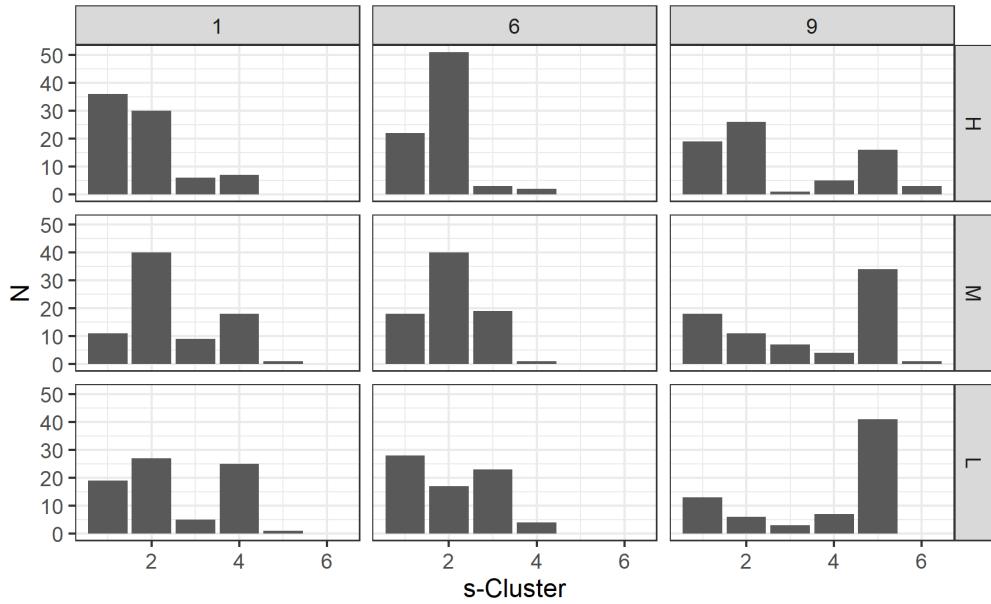


Figure 6: S-cluster membership for each sequence and quiz score tertile cohort.

1-6 focused on introducing individual concepts, whereas M7-M10 focused on the synthesis of multiple concepts and mathematical modeling.

Moreover, a smaller fraction of students displayed unique learning strategies towards the end of the semester when the content became even harder. In s-cluster 9-3 for example, students consistently adopted m-clusters 6 (SYSn) and 7 (SYAn). A similar trend was also observed for the last four modules in s-clusters 6-2 and 6-3. Both m-clusters are characterized by a short first attempt, followed by a study event, which could be interpreted as students strategically skipping the required first attempt, likely due to a lack of self-confidence in their ability to solve the problem. Right after the first guessing attempt, those students followed their plan to study the instructional materials, with the outcome being a mix of success (7) and failure (6).

The correlation between s-cluster membership and course assessment performance provides further evidence for the interpretation of s-clusters. The three s-clusters that are more frequently observed among the top or middle tertile, 1-1, 6-2, and 9-3, all consist of predominantly m-clusters 1 (A) and 2 (AYAn). On the other hand, the two s-clusters, 6-3 and 9-5, involving a significant fraction of m-clusters 4 (S), 5 (AYSn), and 6 (SYSn), are significantly more frequently observed among the bottom tertile. The one exception being 1-4, which is predominantly m-cluster 2, yet contains significantly more low performers. Consistent with existing research on incoming knowledge [19], students who do not have prior knowledge of the content in the first week of the course generally performed lower on assessments.

The current study makes several contributions both in terms of analysis methodology for trace data and new insights into students' SRL behavior in an online learning environment. Regarding analysis methodology, we overcame the challenges imposed by a more

limited and restrictive online system with three main innovations. First, key information contained within the duration of events is extracted using mixture modeling. Second, we used an SRL theoretical framework to inform the clustering algorithm of pedagogically important features in trace data and showed that it outperformed more conventional data-based sequence-mining techniques. Finally, we used sequence pattern analysis in the final stage of clustering to reveal information on students' change of learning strategies over time.

Regarding the insights gained on students' SRL behavior, the most important takeaway from the current analysis is that the majority of students are continuously self-regulating and adjusting their strategies from one module or one sequence to the next. We did not find a group of students whose behavior consistently indicates a lack of self-regulation. Rather, even when a student employs counterproductive strategies such as answer copying or guessing, it is more likely the result of a deliberate choice in response to challenges such as increased content difficulty or fatigue towards the end of the semester. For instructors, this observation reminds us that rather than blaming students for lack of academic integrity or self-discipline, it is essential to re-examine existing instructional design. It could be beneficial to pick up the pace early on when most students are actively engaged, and space out the more challenging content towards the end to allow for more opportunity to overcome challenges. Moreover, it seems that the prime time to implement interventions directed at improving students' SRL skills is the middle or latter half of the semester, rather than at the beginning.

4.1 Caveats and Future Directions

From a technical standpoint, the current m-clusters reflected planning and decision-making more than learning outcomes, as the

same m-cluster includes traces that pass on either 2 or 5 attempts. This could be caused by the choice of feature weights for the Gower coefficients, but also by the restriction of having a maximum of 10 clusters. In addition to conducting a more systematic search of the clustering parameter space, future studies could explore whether there can actually be more than 10 meaningful clusters, or that the trace data can be clustered in more than one way. In addition, future studies could employ process mining algorithms to discover whether there exist less frequent processes in each m-cluster. An exploratory analysis employing heuristic miner algorithm [15] on 80% of traces created causal nets that are highly similar to the current process maps, but with some non-trivial differences in less frequent paths and states.

From a pedagogical standpoint, future studies will need to investigate students' different strategies interacting with the instructional materials and practice problems, which in the current study is simplified into a single binary variable. Moreover, important insight into students' learning behavior can be gained by tracing how students change strategies from one sequence to the next over the semester, especially when combined with self-report data such as surveys on students' SRL strategies and goals and orientations.

Finally, a highly valuable future direction would be to combine or integrate the current multi-level clustering analysis scheme with other schemes such as Trace-SRL, to study students' SRL behavior in an open online learning environment consisting of multiple restrictive sub-systems.

ACKNOWLEDGMENTS

This work is supported by NSF Award No. DUE-1845436. We would like to thank UCF Center for Distributed Learning for creating the UCF Open project and the Obojobo platform, in particular Dr. Francisca Yonekura, Ian Turgeon, Allison Banzon, and Zachary Berry.

REFERENCES

- [1] Yoav Benjamini and Daniel Yekutieli. 2001. The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics* 29, 4 (2001), 1165 – 1188. <https://doi.org/10.1214/aos/1013699998>
- [2] B Bloom. 1968. Learning for Mastery. Instruction and Curriculum. Regional Education Laboratory for the Carolinas and Virginia, Topical Papers and Reprints, Number 1. *Evaluation comment* 1 (1968), 12.
- [3] Zhongzhou Chen, Mengyu Xu, Geoffrey Garrido, and Matthew W. Guthrie. 2020. Relationship between students' online learning behavior and course performance: What contextual information matters? *Physical Review Physics Education Research* 16, 1 (jun 2020), 010138. <https://doi.org/10.1103/PhysRevPhysEducRes.16.010138>
- [4] K Anders Ericsson, Ralf T Krampe, and Clemens Tesch-Römer. 1993. The role of deliberate practice in the acquisition of expert performance. *Psychological Review* 100, 3 (1993), 363–406. <https://doi.org/10.1037/0033-295X.100.3.363>
- [5] Yizhou Fan, John Saint, Shaveen Singh, Jelena Jovanovic, and Dragan Gašević. 2021. A learning analytic approach to unveiling self-regulatory processes in learning tactics. *ACM International Conference Proceeding Series* (2021), 184–195. <https://doi.org/10.1145/3448139.3448211>
- [6] Zachary Felker and Zhongzhou Chen. 2020. The impact of extra credit incentives on students' work habits when completing online homework assignments. In *2020 Physics Education Research Conference Proceedings*. American Association of Physics Teachers (AAPT), Virtual Conference, 143–148. <https://doi.org/10.1119/perc.2020.pr.felker>
- [7] Alexis Gabadinho, Gilbert Ritschard, Nicolas S. Müller, and Matthias Studer. 2011. Analyzing and Visualizing State Sequences in R with TraMineR. *Journal of Statistical Software* 40, 4 (2011), 1–37. <https://doi.org/10.18637/jss.v040.i04>
- [8] Geoffrey Garrido, Matthew W. Guthrie, and Zhongzhou Chen. 2020. How are students' online learning behavior related to their course outcomes in an introductory physics course?. In *2019 Physics Education Research Conference Proceedings*, Ying Cao, Steven Wolf, and Michael B. Bennett (Eds.). American Association of Physics Teachers, Provo, UT. <https://doi.org/10.1119/perc.2019.pr.Garrido>
- [9] Matthew W. Guthrie, Tom Zhang, and Zhongzhou Chen. 2020. A tale of two guessing strategies: interpreting the time students spend solving problems through online log data. In *Physics Education Research Conference Proceedings*. American Association of Physics Teachers (AAPT), Virtual Conference, 185–190. <https://doi.org/10.1119/perc.2020.pr.guthrie>
- [10] Gert Janssenswillen. 2020. *bupaR: Business Process Analysis in R*. <https://CRAN.R-project.org/package=bupaR> R package version 0.4.4.
- [11] Gert Janssenswillen. 2020. *processmapR: Construct Process Maps Using Event Data*. <https://CRAN.R-project.org/package=processmapR> R package version 0.3.4.
- [12] Jelena Jovanović, Dragan Gašević, Shane Dawson, Abelardo Pardo, and Negin Mirriahi. 2017. Learning analytics to unveil learning strategies in a flipped classroom. *Internet and Higher Education* 33 (2017), 74–85. <https://doi.org/10.1016/j.iheduc.2017.02.001>
- [13] Martin Maechler, Peter Rousseeuw, Anja Struyf, Mia Hubert, and Kurt Hornik. 2021. *cluster: Cluster Analysis Basics and Extensions*. <https://CRAN.R-project.org/package=cluster> R package version 2.1.2 – For new features, see the 'Changelog' file (in the package source).
- [14] Jorge Maldonado-Mahauad, Mar Pérez-Sanagustín, René F. Kizilcec, Nicolás Morales, and Jorge Muñoz-Cama. 2018. Mining theory-based patterns from Big data: Identifying self-regulated learning strategies in Massive Open Online Courses. *Computers in Human Behavior* 80 (2018), 179–196. <https://doi.org/10.1016/j.chb.2017.11.011>
- [15] Felix Mannhardt. 2020. *heuristicsmineR: Discovery of Process Models with the Heuristics Miner*. <https://CRAN.R-project.org/package=heuristicsmineR> R package version 0.2.4.
- [16] Marcos Oliveira Prates, Celso Rômulo Barbosa Cabral, and Víctor Hugo Lachos. 2013. mixmsmn: Fitting Finite Mixture of Scale Mixture of Skew-Normal Distributions. *Journal of Statistical Software* 54, 12 (2013), 1–20. <https://www.jstatsoft.org/v54/i12/>
- [17] Peter J. Rousseeuw. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* 20 (1987), 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- [18] J. Saint, A. Whitelock-Wainwright, D. Gašević, and A. Pardo. 2020. Trace-SRL: A Framework for Analysis of Microlevel Processes of Self-Regulated Learning From Trace Data. *IEEE Transactions on Learning Technologies* 13, 4 (Oct 2020), 861–877. <https://doi.org/10.1109/TLT.2020.3027496>
- [19] Shima Salehi, Eric Burkholder, G. Peter Lepage, Steven Pollock, and Carl Wieman. 2019. Demographic gaps or preparation gaps?: The large impact of incoming preparation on performance of students in introductory physics. *Phys. Rev. Phys. Educ. Res.* 15 (Jul 2019), 020114. Issue 2. <https://doi.org/10.1103/PhysRevPhysEducRes.15.020114>
- [20] Daniel L. Schwartz and John D. Bransford. 2005. Efficiency and Innovation in Transfer. In *Transfer of Learning from a Modern Multidisciplinary Perspective (Current Perspectives on Cognition, Learning and Instruction)*. Jose P. Mestre (Ed.). IAP - Informaiton Age Publishing Inc., Charlotte, 1–51.
- [21] Kyle M. Whitcomb, Matthew W. Guthrie, Chandrakha Singh, and Zhongzhou Chen. 2021. Improving accuracy in measuring the impact of online instruction on students' ability to transfer physics problem-solving skills. *Physical Review Physics Education Research* 17, 1 (mar 2021), 010112. <https://doi.org/10.1103/PhysRevPhysEducRes.17.010112>
- [22] Philip H. Winne. 2018. Cognition and Metacognition within Self-Regulated Learning. In *Handbook of self-regulation of learning and performance* (2nd ed.) (2nd ed.), Patricia A. Alexander, Dale H. Schunk, and Jeffrey A. Greene (Eds.). Number 10449. Routledge, New York, Chapter Cognition, 36–48. <https://doi.org/10.4324/9781315697048.ch3>
- [23] Tom Zhang, Michella Taub, and Zhongzhou Chen. 2021. Measuring the Impact of COVID-19 Induced Campus Closure on Student Self-Regulated Learning in Physics Online Learning Modules. In *LAK21: 11th International Learning Analytics and Knowledge Conference* (Irvine, CA, USA) (LAK21). Association for Computing Machinery, New York, NY, USA, 110–120. <https://doi.org/10.1145/3448139.3448150>
- [24] Barry J. Zimmerman. 2000. Attaining self-regulation: A social cognitive perspective. In *Handbook of self-regulation*. Academic Press, San Diego, 13–39. <https://doi.org/10.1016/B978-012109890-2/50031-7>
- [25] Barry J. Zimmerman. 2013. From Cognitive Modeling to Self-Regulation: A Social Cognitive Career Path. *Educational Psychologist* 48, 3 (2013), 135–147. <https://doi.org/10.1080/00461520.2013.794676>
- [26] Barry J. Zimmerman and Dale H. Schunk. 2011. Self-regulated learning and performance: An introduction and an overview. In *Handbook of self-regulation of learning and performance*. Routledge/Taylor & Francis Group, New York, 1–12.