Evolutionary Clustering of Apprentices' Self-Regulated Learning Behavior in Learning Journals

Paola Mejia-Domenzain, Mirko Marras, Christian Giang, Alberto Cattaneo and Tanja Käser

***Abstract*—Learning journals are increasingly used in vocational education to foster self-regulated learning and reflective learning practices. However, for many apprentices, documenting working experiences is a difficult task. In this paper, we profile apprentices' learning behavior in an online learning journal. Based on a pedagogical framework, we propose a novel multi-step clustering pipeline that integrates different learning dimensions into a combined profile. Specifically, the profiles are described in terms of effort, consistency, regularity, help-seeking behavior, and quality of the written entries. Our results on two populations of chef apprentices (183 apprentices) interacting with an online learning journal (over 121K entries) show that our pipeline captures changes in learning patterns over time and yields interpretable profiles that can be related to academic performance. The obtained profiles can be used as a basis for personalized interventions, with the ultimate goal of improving the apprentices' learning experience.**

***Index Terms*—Vocational education, workplace learning technologies, learner profiles, time series analysis, evolutionary clustering, longitudinal study**

I. Introduction

The digital transformation is changing the way we live and work. The workforce is challenged to adapt to the evolving working environments, roles, and tasks. Consequently, Vocational Education and Training (VET) systems must prepare the future workforce to become lifelong learners able to adapt their competencies to the changing demands. In particular, learners should have the ability to self-regulate their learning process and to reflect on their learning experiences and activities [1].

Learning journals have the potential to foster self-regulated learning (SRL) and reflective learning practices [2], [3] and are consequently increasingly adopted in VET [1]. Typically, apprentices use the learning journals to take notes on the tasks and the skills acquired during workplace experiences [4]. In addition, the journals also help apprentices connect the theoretical knowledge learned in the vocational schools to the practical situations experienced in the workplace [5], [6].

The use of learning journals is, however, not effective *per se* [7], [1]. Independently documenting learning experiences is a challenging task for many learners [8], [9], [10]. Thus, providing *personalized* guidance can improve the learning experience. Profiling learners allows us to identify groups with similar patterns of behaviors, building a basis for targeted interventions. Unfortunately, previous work on learning journals has mainly focused on their effectiveness rather than analyzing and identifying profiles (e.g., [11], [12]).

In contrast, there exists an extensive body of work on profiling learners in a range of digital learning environments and settings (e.g., massive open online courses (MOOCs) [13], intelligent tutoring systems [14], and flipped classroom courses [15]). Nevertheless, applying these methods in the context of online learning journals in VET, has some limitations. Firstly, most of the prior research on profile identification focuses on clustering learning behavior for only one period (e.g., the duration of a MOOC [16], one semester of a course [17]), and only a few works consider the evolution of clusters over time (e.g., learner sessions with an intelligent tutoring system [14]). Secondly, within a period, the use of aggregated features [18] is more common than treating interactions (e.g., within a session) as time series [19], [20]. Lastly, in the context of SRL, [19] and [13] focused on only one dimension of SRL, consistency [19] and regularity [13] respectively. This practice does not consider possible inter-dependencies across learning dimensions [21].

In this paper, we profile apprentices' behavior in an online learning journal throughout their apprenticeship. We propose a novel multi-step clustering pipeline that is based on a pedagogical framework of SRL in formal education and workplace contexts. The pipeline consists of two clustering steps. In the first step, we cluster the apprentices separately for relevant dimensions (focusing on SRL behavior). We transform apprentices' log data into time series, enabling us to retrieve the shape of behavioral patterns over time (e.g., increasing engagement towards the end of the semester). In the second step, we perform another level of clustering based on the cluster labels of the individual dimensions to obtain multi-dimensional learner profiles. Our pipeline has several advantages: 1) in contrast to prior work, we combine time series modeling (e.g., [15], [20], [22], [23]) and evolutionary clustering (e.g., [24], [14]), enabling us to capture complex temporal patterns over a semester and analyze profile evolution throughout the apprenticeship; 2) by integrating single dimensions of behavior into multi-dimensional profiles, we can represent dependencies across dimensions, leading to a more realistic learning model; 3) our pipeline is *transferable*, i.e. it can be applied to other learning environments by adapting the features describing the different learning dimensions.

P. Mejia-Domenzain, C. Giang and T. Käser are with the School of Computer and Communication Sciences, EPFL, Lausanne, Switzerland. E-mail: {paola.mejia, christian.giang, tanja.kaeser}@epfl.ch

C. Giang is also with the Department of Education and Learning, University of Applied Sciences and Arts of Southern Switzerland, Locarno, Switzerland.

M. Marras is with the Department of Mathematics and Computer Science, University of Cagliari, Cagliari, Italy. E-mail: mirko.marras@acm.org

A. Cattaneo is with the Swiss Federal University for Vocational Education and Training, Lugano, Switzerland. E-mail: alberto.cattaneo@suffp.swiss

Manuscript received June 15, 2021; revised June 17, 2022; accepted July 29 2022.

We apply our pipeline to data from two independent populations of chef apprentices from different vocational schools. Both populations interacted with an online learning journal over their three-year apprenticeship. With our analyses, we address four research questions: Can we identify interpretable profiles of apprentices integrating different behaviors, and are these profiles related to academic performance (**RQ1**)? How do these profiles evolve throughout the apprenticeship (**RQ2**)? What type of behavioral patterns (in terms of effort, consistency, regularity, help-seeking behavior, and quality) can we observe during a semester (**RQ3**)? How do the learner profiles compare across vocational schools (**RQ4**)?

Our findings show that we can identify both interpretable patterns (by using theory-based features) and interpretable profiles (by using theory-based dimensions) for specific aspects of learning behavior. We further observe some significant differences in academic performance between the profiles. While apprentices move between profiles throughout their apprenticeship, they tend to move to similar profiles between two consecutive semesters. Finally, only a subset of the obtained profiles is shared between the two populations. The identified profiles contribute to the teachers' and in-company trainers' understanding of the different apprentices' SRL behaviors and build the basis for targeted interventions. Furthermore, our findings confirm the diversity of learning patterns across apprentices and the importance of the context (i.e., the community of practice of the apprentice).

II. Related Work

A. Learning Journals in VET

Learning journals collect learners' (evidence of) work accompanied by reflective comments and so may foster SRL and reflective learning practices [2]. Learning journals are hence increasingly used in VET to support apprentices in connecting theoretical and practical knowledge [5], [6] or act as boundary-crossing objects across VET contexts (i.e., vocational school, company, and inter-company courses) [4], [25]. Prior work has examined the effects of learning journals in professional education (e.g., nursing [11], physiotherapy [12]).

Online learning journals have brought advantages compared to their paper-based counterpart. Not only are they ubiquitously accessible, but also facilitate the integration of visual artifacts (e.g., photos or videos), which can be particularly beneficial in VET [26]. Moreover, they can ease the creation, editing, and storing of text and media entries [27]. From a pedagogical perspective, having visual traces of the experience serves as a trigger to access concrete memory [6]. [28] found that learners engaged more in reflection and generated more entries with online learning journals than paper-based ones.

However, the use of learning journals is not effective *per se*. Their appropriate usage by both apprentices and supervisors is needed to effectively support the learning process [1]. Research in the Swiss VET system has shown that stakeholders often do not share the same conception of the aims and functions of learning journals [7]. Moreover, documenting learning experiences is a hard task for many apprentices [8], [10]. Providing scaffolds, therefore, has the potential to improve the learning process. [29] showed that apprentices' SRL strategies can be improved by scheduling regular feedback meetings with supervisors. [30] found that scaffolding peer feedback improved the quality of the learning documentation.

The digital format of learning journals allows to record apprentices' interactions with journals, derive insights into learning strategies, and support interventions. For example, [1] manually tagged apprentices' reflections and found a positive correlation between (meta-)cognitive learning strategies and academic performance. Unfortunately, the potential of learning analytics still appears unexploited in VET learning journals.

B. Learner Profiling

Extensive research has been done to identify learner profiles in learning environments. We focus on prior work based on the three key aspects of our paper: SRL strategies, time series clustering, and cluster evolution.

A vast part of previous work has focused on clustering students based on SRL strategies. For example, [17] explored effort regulation in university students and found a significant correlation with academic performance. [31] studied students' commitment and consistency in MOOCs. In a blended setting, [20] found that students working consistently had higher academic performance. [23] studied patterns of macro-level processes of planning, engagement, evaluation, and reflection in log data. In MOOCs, [13] quantified regularity and found that regular students outperformed their peers. In [16], student groups were detected based on their help-seeking behavior.

Apart from log data, SRL strategies have been explored via other data sources. For instance, [32] conducted a latent profile analysis with online/blended students to identify SRL profiles from survey answers. Likewise, [33] analyzed learners' strategies using a latent class analysis on the answers to the PISA learning strategy survey. [20] did a latent class analysis to study behavioral engagement in MOOCs. Finally, [15] studied the relationship between detected and self-reported strategies.

Prior work on learner profiles has mostly used aggregated features (e.g., the total number of watched videos in a MOOC [31]). Yet, log data usually represents a time series of events. Hence, recent work on SRL profiling focused on time series. For instance, [15] encoded trace data as action sequences and computed their distance using an optimal matching method. Likewise, [23] used Markov models to represent time series. [22] showed that Dynamic Time Warping (DTW) is more effective than Euclidean distance to compare time series.

Finally, there is limited literature on cluster evolution and how profiles change over time. For example, [20] studied the transition of learning strategies across course weeks. However, transitions were aggregated over the whole course, and it was not possible to see whether some strategies (dis)appeared or were more (less) frequent across weeks. [24] clustered student interactions in a digital learning environment separately at different points in time. [14] proposed an evolutionary clustering approach to obtain temporally consistent clusters.

III. Context

Our work studies data from chef apprentices using an online learning journal designed under the Swiss Dual-T project

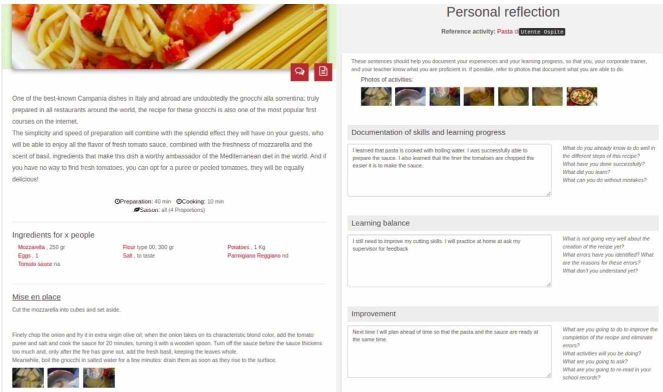


Fig. 1: An example of a recipe (left) and a journal entry (right).

(2015 - 2020) [6]. The Swiss vocational education is organized as a dual system, with apprentices alternating between lessons in vocational schools and their workplaces in companies.

A. Apprenticeship Learning

Apprenticeship learning requires interactive participation in cultural practices and shared learning activities [34]. Thus, their learning process is affected by the learning environment and explained as a legitimate peripheral participation in communities of practice (CoP) [35], [36]. The latter are the social contexts apprentices participate in. For chef apprentices, in-company trainers, chefs, and waiters working at a restaurant are part of their CoP. The participation must be legitimate (i.e., apprentices should have access to the practices and the community), and it is at first peripheral (e.g., chef apprentices may chop vegetables before designing restaurant menus).

In a dual-track VET system (like the Swiss one), vocational apprentices alternate between the vocational school (learning- and technical-oriented) and the workplace (production- and practice-oriented) [37]. A challenge arising from this system is integrating theory and practice [5]. Articulation and reflection are two methods that foster generalization across contexts [34]. The first one involves externalizing thoughts or cognitive processes, while the second one allows us to examine past professional practices. The 'Erfahrraum' model is a pedagogical model that aims to connect theory and practice [6]. It assumes that experiences alone do not lead to knowledge but rather knowledge is constructed through reflection processes. The online learning journal in our study implements this model.

B. Online Learning Journal for Chef Apprentices

The learning environment in our study is an online learning journal platform for chef apprentices (Fig. 1), aimed to support apprentices in linking the theory learned at school with their hands-on workplace experiences. [38] showed that the platform is effective in improving apprentices' learning outcomes (e.g., their declarative knowledge acquisition and meta-cognitive skills development). The platform supports two types of entries: *recipes* and *experiences*. Recipes cover all aspects related to food, while experiences focus on topics like hygiene and work safety. For both types of entries, apprentices can enter a title and a description, upload images, add appropriate tags (about learning topics), and, for recipes, specify the ingredients. Each entry is linked to a learning journal that prompts apprentices to reflect on what went well and identify areas for improvement. The platform also allows apprentices to ask for feedback from their in-company trainers.

C. Participants

We study log data belonging to two populations of chef apprentices, coming from two VET centers located in two different language regions of Switzerland, using an online learning journal for their apprenticeship (6 semesters).

The data set of the first vocational school `VS1` was used to answer **RQ1**-**RQ3**. It contains the log data of 139 apprentices (101,579 entries) from a VET center in the Italian-speaking part of Switzerland. The training was organized biweekly: apprentices went to school for two days every other week. The data set of the second vocational school `VS2` served to compare learner profiles across contexts and answer **RQ4**. It contains the log data of 44 apprentices (20,957 entries) from a VET center in the French-speaking part of Switzerland. Their training was organized weekly: apprentices went to school for one day per week. All participants were informed about the research and had the right to withdraw at any point in time. The study was approved by the responsible institutional review board (HREC number: 0050-2020/05.08.2020).

IV. Method

To study apprentices' behavior over their apprenticeship, we propose a multi-step evolutionary clustering pipeline based on a framework of SRL in formal education and the workplace.

A. Pedagogical Framework

The concept of SRL has been studied extensively in education and psychology over the last three decades. There exist several models/conceptualizations of SRL (see [39] for an overview), divided into two main categories. Models using a 'process-oriented' perspective see SRL as a proactive process organized as a set of (repeating) learning phases; whereas models using an 'aptitude-oriented' perspective characterize SRL by individual differences and identify cognitive, metacognitive, motivational, and emotional aspects of learning. Measurements of SRL as aptitude often vary within individuals over long periods as well as across settings. They are frequently used to predict future behavior (e.g., whether a learner will (not) act on an SRL-related cognition) [40].

A large body of research on SRL targets formal education settings. Given that learners may use a variety of SRL strategies as part of their learning, many of these works (e.g., [17], [19], [20]) examined the relationship between SRL strategies and (academic) performance. In this sense, [41] performed a meta-analysis to investigate the effect of different categories of SRL strategies on academic achievement in an *online education* setting. They used the nine subscales of the Motivated Strategies for Learning Questionnaire (MSQL) [42] as a basis for their meta-analysis and found a significant association to academic achievement for five of these subscales: metacognition (awareness and control of thoughts), time management (ability to plan study time and tasks), effort regulation (persistence in learning), critical thinking (ability to carefully examine learning material), and help-seeking (obtaining assistance from supervisors/instructors).

Much less work focused on studying SRL in the workplace. In contrast to formal education, workplace learning involves interactive participation in cultural practices and shared learning activities (see Section III-A). In workplace settings, SRL is highly social and structured by work tasks [43]. Other research [44], [45] emphasizes the importance of knowledge artifacts created in the workplace for SRL. [46] have found that the workplace learning context is a predictor of SRL. Finally, [47] explored the effects of technological scaffolding of SRL on workers in European organizations.

Our use case (online learning journals of apprentices) spans formal education (vocational school) and workplace learning, although the use of the online learning journal is controlled through the school (i.e. the teachers instruct the apprentices to document their recipes and workplace experiences). Our framework, therefore, combines elements from SRL concepts in formal education and workplace learning. Furthermore, we assume an 'aptitude-oriented' perspective of SRL and acknowledge that behavior is influenced by the environment and the context (in particular in the case of workplace learning [46], [48]). Therefore, we hypothesize that there will be differences in the SRL behavior across apprentices from different contexts (CoP or schools). Given that we access log data only, based on the findings of [42], we use apprentices' learning behaviors in the system as approximations of SRL processes. Following [41], we represent apprentices' SRL as a composition of effort regulation (*Effort*), time management (*Regularity*, *Consistency*), and help-seeking (*Help-Seeking Behavior*). We study time management in short-term [13] (*Regularity*) and long-term [49] (*Consistency*). The nature of our use case and log data does not allow us to measure metacognition or critical thinking. We capture the influence of the workplace by modeling interactions between apprentices and in-company trainers into the help-seeking dimension. Finally, we integrate a *Quality* dimension into our representation of behavior, serving as a proxy of student learning over the semester: prior work has emphasized the importance of knowledge artifacts in workplace learning [44], [45] and found correlations between the quality of journal entries [3], [1] and academic performance.

B. Multi-Step Evolutionary Clustering Pipeline

Our multi-step evolutionary clustering pipeline (exemplified in Fig. 2) integrates different dimensions and consists of three main steps that are taken for each semester.

In the first step, we transform the log data into features (indicators) that aim to approximate different dimensions of learning. The features are computed as time series (Section IV-B1). Thus, per feature and apprentice, the output from the first step is a time series (vector) of length equal to the number of (bi)weeks in the semester. The goal of the second step is to study and cluster each dimension individually (Section IV-B2). Therefore, we compute the similarity matrix between apprentices separately for each feature ($N \times N$, where N is the number of apprentices). Then, to integrate the features from the same dimension, we add their similarity matrices and obtain the dimension similarity matrix ($N \times N$). Next, per dimension, we account for the apprentices' behavior from previous semesters by smoothing the dimension similarity matrix ($N \times N$). Following, to identify the behavioral patterns, we perform Spectral Clustering. We use as input the smoothed similarity matrix and obtain a vector of labels (indicating the cluster an apprentice belongs to), which we interpret using domain knowledge. In the third step, we integrate the information from the SRL dimensions into a multi-dimensional learner profile (Section IV-B3). We perform a second clustering step via K-modes using the five vectors of labels obtained in the second step (one vector for each dimension) as input. The obtained profiles from this second clustering step are interpretable per se as they are composed of theory-based dimensions. We describe each step of the pipeline below. Technical details and code are provided in on our repository[1].

1) Dimensions of Self-Regulated Learning: As described in Section IV-A, we represent learners' SRL behavior using five dimensions: *Effort*, *Quality*, *Consistency*, *Help-Seeking Behavior*, and *Regularity*. Practically, from the log data, we extracted features serving as indicators for these SRL dimensions.

Table I shows the dimensions of behavior and their corresponding features. The Regularity features are scalars, while the other features are time series. Prior work has shown that learners shift their learning strategies across domains and even within a course in response to contextual factors [50]. Modeling features as time series instead of aggregated values allows us to account for this temporal perspective. Hence, we computed the features per (bi)week to build a time series of length equal to the number of (bi)weeks in the semester.

The ***Effort*** dimension monitors the intensity of the apprentices' commitment to manage tasks and challenges in their learning. [17] used the amount of time spent on an online course in higher education as an indicator of effort and demonstrated its relation to academic performance, while [51] used the number of events to measure the strength of students' engagement. Based on these prior studies, we characterize learners' effort by calculating the total time spent on the platform as well as the total number of writing events.

In contrast to Effort, the ***Consistency*** dimension focuses on the relative shape rather than the absolute magnitude of events. It measures how learners' effort varies throughout the semester and estimates their intra-course time management skills. These skills are important in learning journals, where the accumulation of material should be made over some time, not 'in one go' [3]. Regular journal entries can update the cognitive structure and promote the absorption and connection of new knowledge into the updated cognitive structure. [19] studied the consistency of study habits and examined whether student behavior is constant throughout the semester or visible only at the beginning or at the end of the course; the data was processed as time series containing the number of activities per unit of time. Following [19], we computed the relative

[1] https://github.com/epfl-ml4ed/evolutionary-srl-clustering/

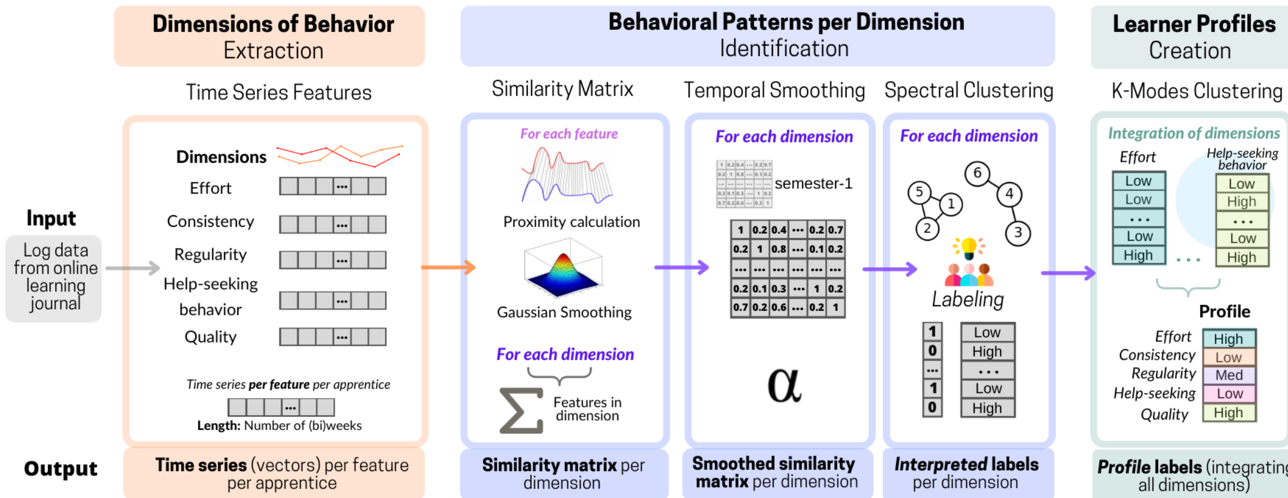


Fig. 2: **Proposed Pipeline**. In the first step, we take as input the log data aiming to extract meaningful indicators of behavior. In the second step, we study each dimension of behavior individually. We take as input the time series features to cluster apprentices for each dimension (Behavioral Patterns). In the third step, to obtain multi-dimensional profiles, we do another level of clustering using the cluster labels from the second step (Learner Profiles).

Dimensions	Feature Name	Feature Description
Effort	Platform usage Number of writing events	How many minutes did the apprentice spend on the platform? How much did the apprentice use the platform to write/edit content?
Consistency	Average session duration Relative platform usage Relative number of writing events	On average, how many minutes did an apprentice session in the platform last? How many minutes did the apprentice *relatively* spend on the platform? How much did the apprentice *relatively* use the platform to write/edit content?
Regularity	Peak of (bi)weekday Periodicity of day hours Periodicity of (bi)week hours	Does the apprentice tend to work more on certain days than others? Does the apprentice tend to work more on certain hours of a day than others? Does the apprentice tend to work more on certain hours and days than others?
Help-Seeking Behavior	Feedback request ratio Feedback response ratio	Out of all the recipes and experiences, how many of them had a feedback request? Out of all the feedback requests, how many of them received an answer?
Quality	Average reflection length Image ratio Ingredients ratio Tags ratio	On average, how many characters did the apprentice write in their reflections? Out of all recipes and experiences, how many of them have at least one image? Out of all recipes, how many of them have at least one ingredient? Out of all recipes and experiences, how many of them have at least one tag?

TABLE I: Extracted features for each behavioral dimension considered in our study.

platform usage and the relative number of writing events by normalizing the respective time series per student to observe the shapes over time. Therefore, Effort measures the intensity, and Consistency the shape of the same indicators. Moreover, [52] used the average session duration and the standard deviation to measure the consistency of student activity in MOOCs. Similarly, we calculated the average session duration per unit of time and built a time series that captures the variations.

The ***Regularity*** dimension assesses apprentices' intra-week and intra-day time management patterns (i.e., capturing whether an apprentice is regularly engaged on specific weekdays or day times). Prior work has analyzed this dimension in MOOCs [13] showing its relation with student performance. Building on this study, we included three features of apprentice regularity from [13]: periodicity of day hour (being more active on certain hours of the day), periodicity of weekday (being more active on certain days of the week), and periodicity of week hour (being more active on certain hours **and** weekdays).

The ***Help-Seeking Behavior*** dimension measures learners' ability to ask for support when needed. In prior work, [16] explored this dimension in MOOCs and found that engagement in forums resulted in higher performance. In our context, instead of asking questions in a forum, apprentices can ask individual questions and request feedback on their recipes and experiences from their in-company trainers. We describe help-seeking behavior as the feedback request ratio (the number of entries with a feedback request divided by the total number of entries) and the feedback response ratio (the number of feedback requests with a response). This dimension captures the interaction between the apprentices and their CoP [36].

The ***Quality*** dimension aims to capture the completeness of writing entries (recipes and experiences). Prior work has shown the importance of learning documentation completeness [3], [1]; thus, we model the entries' quality as the ratio of recipes with at least one ingredient/tag related to the syllabus. In learning journals for baker apprentices, [1] also found a significant positive correlation between learners' performance and the number of pictures in recipes; we hence use the image ratio as a quality proxy. Lastly, [53] found that the number of words was a good predictor of writing quality; so our last quality measure is the average length of the reflection entries.

2) Behavioral Patterns: The second step of the pipeline aims at identifying distinct behavioral patterns per dimension and semester. We first compute the pairwise similarity matrix per semester and dimension, then smoothen it based on the previous semester, and finally perform a spectral clustering.

Similarity Matrix. For each feature, we compute the pairwise distance between the apprentices to build the distance matrix of size $N \times N$ (where N is the number of apprentices). To calculate such distances, we use the Euclidean distance for the regularity features (real values ≥ 0). For the other features (time series), we use the DTW distance (e.g., [22]). DTW is a distance measure that searches for the optimal alignment between two given time series. Compared to Euclidean distance, DTW is more flexible and tends to correctly identify similar patterns (such as peaks) despite small variations (shifts) in time. Then, we sum the distance matrices of all the features in the same dimension. Following, we compute the similarity matrix S by applying a Gaussian kernel. Thus, the outputs from this step are five similarity matrices (one per dimension) per semester that are used as input for the following step.

Note that, for each feature in the dimension, we obtain the standard deviation (σ) of the Gaussian kernel and the window size w that constrains the DTW degree of flexibility via a grid search in the range $0.5 \leq \sigma \leq 1.5$ with steps of 0.1 for sigma and range $0 \leq w \leq 4$ with increments of 1 for the window size. We choose the optimal values to maximize the *Silhouette* score of the *Spectral Clustering* (see next subsection).

Temporal Smoothing. The apprentices as well as their working contexts and learning (e.g. from peripheral to more complex tasks) will evolve over apprenticeship and, therefore, we can also expect an evolution of their behavior [46], [48]. One possible approach to represent apprentices' evolving behavior is to feed the similarity matrices into a standard clustering method, yielding a separate cluster solution for each dimension per semester. However, this type of approach does not make use of the temporal information available and can be very sensitive to noise, leading to temporally inconsistent clusters. Therefore, we use the evolutionary clustering approach proposed by [14] and smoothen the current similarity matrix using the similarity matrix of the previous time frame. We assume that the similarity matrix S is the sum of the unknown true similarity matrix Ψ and random noise N. Instead of clustering directly on S, the *true* similarity matrix Ψ is estimated for every semester t. The premise is that Ψ is free from noise and clustering it instead of S will lead to clusters of higher quality. The true similarity matrix is computed as follows:

$$\Psi_t = \alpha_t \, \hat{\Psi}_{t-1} + (1 - \alpha_t) \, S_t, \tag{1}$$

where α_t controls the amount of smoothing applied to Ψ. The optimal smoothing factor α_t depends on the amount of new information S_t contains compared to the similarity matrix of the previous semester and the estimated noise in S_t: if large differences between S_t and S_{t-1} (the similarity matrix of the previous semester) are observed, α_t will be low (to be able to capture novel behaviors). If S_t is very noisy, α_t will be large.

Spectral Clustering. Finally, we apply *Spectral Clustering* [54] to cluster the smoothed similarity matrices Ψ per semester and dimension. In contrast to K-Means, Spectral Clustering is not limited to clusters that form convex sets and is particularly good at identifying outliers [55]. The idea behind spectral clustering is that points in a data set can be represented as nodes of a graph and the (weighted) edges connecting the nodes denote the similarity between the points. The clustering task is therefore turned into a graph partitioning problem where the similarity matrix (Ψ) is the weighted adjacency matrix. Then, k-Means Clustering with $k = K$ is applied to the first K eigenvectors of the Graph Laplacian (see [55] for its definition). As output, we obtain the index (number) of the cluster each apprentice belongs to. Using domain knowledge, we interpret and describe them with meaningful labels. The interpretation describes the dimension in terms of magnitude (*Low*, *Medium*, *High*), shape (*Decreasing*, *Increasing*, *Normal*), and peaks (*Low Peaks*, *High Peaks*, *Alternate Peaks*). The magnitude labels are used when the magnitude of the dimension is more or less consistent over the semester. The shape labels describe how the magnitude changes over the semester: high magnitude at the beginning of the semester (*Decreasing*), high magnitude at the end of the semester (*Increasing*), or magnitude following a normal distribution (*Normal*). Finally, the peak labels are relevant only for the regularity dimension. *High Peaks* denotes clusters with a strong preference for working on specific days of the (bi)week/hour of the day, *Low Peaks* indicates a slight preference, and *Alt. Peaks* (Alternate Peaks) represents apprentices who prefer specific days, but not hours of the day.

Model Selection. To find the optimal number of clusters, we chose the *Silhouette* score [56] over other heuristics because it is easy to interpret. It ranges from $-1, ...1$, with higher values indicating that a cluster member is close to its own cluster and far away from the other clusters (high separability). We compute the optimal number of clusters for each dimension and semester via a grid search for $k = 2, ..., 10$ clusters.

3) Learner Profiles: Learning is a process involving elements that follow different sets of logic and work together in a complex interaction [21]. In this step, we integrate the different dimensions of behavior into multi-dimensional profiles, to obtain a complete picture of apprentice behavior and insights into dependencies between dimensions. To obtain the profiles per semester, we take as input the five annotated cluster labels per dimension and semester. Then, we use K-Modes to cluster the annotated labels (the input) and output the multi-dimensional profile each apprentice belongs to (e.g., [19]). K-Modes extends K-Means to cluster categorical data. The former uses the mode (most frequent element) instead of the mean to compute the cluster centroids. We again use the *Silhouette* score to determine the optimal number of clusters.

V. Experimental Results

To answer research questions **RQ1**-**RQ3**, we applied our pipeline to the data set of the first vocational school (`VS1`). Our results show that we can obtain interpretable apprentice profiles related to academic performance. We then applied our pipeline to profile apprentices from a second vocational school (VS2) to answer **RQ4**. While the proposed pipeline yields interpretable profiles also for VS2, only a subset of the obtained profiles are shared between the two populations.

A. Profile Exploration

We profiled the apprentices in VS1 using biweeks as the time unit (features per apprentice per biweek), yielding clusters over the six semesters of apprenticeship. Biweekly time units were chosen to adhere to the apprenticeship format in VS1.

Apprentice Profiles. We hypothesized that our pipeline could identify different profiles of apprentices (**H1**). Our hypothesis is based on the assumption of an 'aptitude-oriented' conceptualization of SRL, that describes SRL based on individual differences (see [39]), and on the findings of [48], who showed that the quality and the intensity of participation (learning) under modern apprenticeships varies widely. We further assumed that these individual differences are manifested in apprentices' learning strategies (see [42]). Based on the findings of previous work on the relationship between SRL strategies and academic achievement (e.g., [17], [20], [23], [19]), we also hypothesized that there would be significant differences in academic performance between profiles.

In a first analysis, we, therefore, examined the resulting apprentices' profiles from the multi-step clustering for the six semesters of the apprenticeship. Table II shows the resulting profiles per semester (rows) and dimension (columns). While the results of all semesters are aggregated for conciseness reasons, we ran the pipeline separately for each semester. The profile descriptions/interpretations were obtained using the cluster centroids for each profile (see Section IV-B2). Given that we used K-Modes, the centroid is the most frequent label per dimension. For example, 95% of the apprentices in profile *B* in semester 1 had *Low Peak*s in Regularity whereas 5% had *High Peaks*. Thus, the centroid for Regularity is *Low Peaks*. As another example, the centroids for profile *B* in semester 1 are *Low* (Effort), *Low* (Quality), *Increasing* (Consistency), *Low* (Help-Seeking Behavior), and *Low Peaks* (Regularity).

At a first glance, using the cluster centroids as cluster representatives to interpret the different behaviors is straightforward [19]. However, while profile *B* exhibits a clear pattern per dimension in semester 1, for some combinations of profile-dimension-and-semester, we do not observe a clear majority. For example, for profile *C* in semester 1, the Help-Seeking Behavior dimension is labeled as *Low*. Nonetheless, for this semester, profile *C* has 57% of the apprentices in cluster *Low* and 43% of the apprentices in cluster *High*; therefore, the provided interpretation is misleading. To address this limitation and to make meaningful interpretations, we provide a confidence estimate for each profile, dimension, and semester combination in Table II. We indicate four different confidence levels depending on the percentage of apprentices in the majority class: *** ($\geq$ 90%), ** ($\geq$ 80%), * ($\geq$ 75%), + ($\geq$ 65%). For example, more than 90% of the apprentices in profile *A* and semester 1 have *High* Effort (in this case, *High* Effort is the majority class), thus, in Table II there are three stars (***) for profile *A*, semester 1 and dimension Effort. We consider an interpretation as valid only if at least $2/3$ of the apprentices belong to the majority class. If, for a certain profile-semester-and-dimension, the majority of the apprentices' labels account for more than 65% of the apprentices in that profile, that interpretation is this valid. Otherwise, there is a white space on the table. For instance, for profile *C* and dimension Help-Seeking Behavior, the interpretation for semester 1 is invalid.

Out of the 13 distinct profiles, four are present in more than one semester. Profile *B* has the highest frequency, occurring in four out of six semesters; profiles *A, C* and *F* are found in two semesters; and the rest of the profiles appear only once. It is interesting to note that the first semester has three frequent profiles *A, B*, and *C* and the last semester, in contrast, has three unique profiles *K, L* and *M*. Moreover, there are two to three categories per dimension, yielding 956 possible combinations. In semester 3, 5, and 6, Effort and Quality have three categories, thus the probability of both being *Low*, *High* or *Medium* is one third. Most of the profiles in these semesters (6 out of 10) have matching Effort and Quality. In addition, some profiles are similar to each other, e.g., the pair of profiles *A* and *C* and the group of profiles *B, F*, and *H* only differ in Consistency; profiles *E* and *K* vary in Help-Seeking Behavior.

Academic Performance. We then checked if there were significant differences between the profiles in terms of apprentices' semester grades. Grades range from 1 to 6, with 6 being the best grade and 4 indicating a passing grade. Based on a significant Levene's test (F(12, 821) = 2.22 p = 0.009) indicating unequal variances, we used the non-parametric Kruskal-Wallis test to assess whether there were significant differences between profiles (χ^2(12) = 65.97, p = 1.8e-09). We then performed a pairwise comparison between clusters using the Wilcoxon Rank Sum test, correcting for multiple comparisons via a Benjamini-Hochberg (BH) procedure. The results of the pairwise comparisons are displayed in Table III).

Interestingly, profiles *B* and *C* have most statistical differences with other profiles. The apprentices from profile *B* have significantly lower grades than those from profiles *A*, *C*, *D*, *E*, *F*, *G*, and *I*. Conversely, the apprentices from profile *C* have significantly higher grades than those from profiles *B*, *H*, *J*, *K*, *L*, and *M*. Profiles *B* and *C* have contrasting characteristics for Effort, Quality and Consistency: profile *B* has *Low* Quality, *Low* Effort, and *Increasing* Consistency, while profile *C* has *High* Quality, *High* Effort and *Decreasing* Consistency.

Furthermore, while profile *F* shows a significantly higher academic performance than profile *B*, their main difference lies in the Consistency dimension: profile *B* has an *Increasing* Consistency pattern whereas profile *F* shows a *Normal* pattern. An *Increasing* Consistency means that apprentices worked more towards the end of the semester (see Fig. 6), which might be due to some form of *procrastination*. However, it is important to treat the different dimensions in combination to reason about academic performance. For example, in the case of profiles C and K, the *Normal* Consistency pattern of profile K does not lead to significantly better academic performance (profile C outperforms profile K).

In summary, this analysis confirms hypothesis **(H1)** by

Profile	Effort	Quality	Consistency	Help Seeking	Regularity
A	High	High	Increasing	Low	Low Peaks
Sem. 1	***	*	*		***
Sem. 4	***	**		**	+
B	Low	Low	Increasing	Low	Low Peaks
Sem. 1	***	***	**	***	***
Sem. 2	+		+	+	***
Sem. 4	**	**	***	***	***
Sem. 6	*	***	***	***	+
C	High	High	Decreasing	Low	Low Peaks
Sem. 1	***	***	***		
Sem. 3	***		***	**	
D	Low	High	Decreasing	Low	High Peaks
Sem. 2		**	***		
E	High	High	Normal	Increasing	Low Peaks
Sem. 2	**	***	+		***
F	Low	Low	Normal	Low	Low Peaks
Sem. 3	***		***	***	***
Sem. 5	***	***		**	**
G	Low	Medium	Normal	Low	Low Peaks
Sem. 3	***		***	***	***
H	Low	Low	Decreasing	Low	Low Peaks
Sem. 4	**	**	**	***	**
I	High Increasing	Medium	Normal	Low	Alt. Peaks
Sem. 5		+		***	
J	High Increasing	High	Normal	High	Low Peaks
Sem. 5		***		**	*
K	High	High	Normal	Low	Low Peaks
Sem. 6			**	**	*
L	High	High	Increasing	High	Low Peaks
Sem. 6		**	***	***	+
M	High Increasing	Medium	Increasing	Low	Low Peaks
Sem. 6		**	**	***	*

*** $\geq$ 90%, ** $\geq$ 80%, * $\geq$ 75%, + $\geq$ 65%

TABLE II: Interpretation of apprentice profiles. Overall, we observe 13 distinct profiles. The italic text denotes the semesters a profile was present in (*Sem. 1 - Sem. 6*) and indicates the reliability of the interpretation for each dimension based on the percentage of apprentices of the profile conforming to the interpretation label.

showing distinct profiles. Using their cluster centroids, we interpreted the profiles and found meaningful differences in their composition and academic performance. Moreover, academic performance is influenced by the combination of dimensions rather than a single dimension **(RQ1)**.

Profile Evolution. In a second analysis, we studied how different profiles form, merge and divide over time. As newcomers to a CoP, apprentices advance their skills and become more responsible [34]; considering this, we hypothesized that apprentices would advance their SRL skills over time (reflected in moving to profiles with stronger SRL skills). This is in line with the 'aptitude-oriented' perspective where aptitude often varies among individuals over long periods (H2).

	B	C	D	E	F	G	H	I	J	K	L	M
A	***											
B		***	**	**	*	***		***				
C							**		**	**	*	**
H								*				
I										*		*

* $\leq$ 0.05, ** $\leq$ 0.01, *** $\leq$ 0.001

TABLE III: Lower triangular matrix of significant differences in academic performance between profiles.

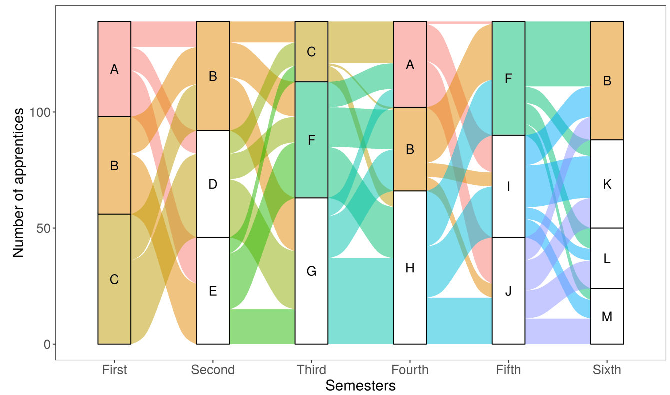


Fig. 3: Profiles over semesters for `VS1`. Each letter denotes a profile of Table II. The boxes of the profiles occurring in more than one semester are colored (e.g., profiles *A*, *B*, *C*, and *F*).

Fig. 3 shows the evolution of the profiles previously discussed over the six semesters of the apprenticeship. Certain profiles (e.g., profile *D* in semester 2) are formed from other profiles splitting a semester before. Other profiles, such as profile *B* in semester 1 and 2, appear in consecutive semesters. A group of apprentices remains in the same profile.

In the transition between semesters 1 and 2, we see examples of profiles dissolving and forming. Profile *B* is split into two with 40% of the apprentices remaining in cluster *B* during the semester 2. A larger part of the apprentices (60%) move to profile *E*: they spend more time on the platform, provide documentation of higher quality, and work more consistently during the semester. This change is of educational relevance because it exemplifies how some apprentices learn and improve their SRL behavior during the apprenticeship (H2). Conversely, in this same transition, profile *C* with *High* Effort dissolves into profile *B* and *D*; one-third of the apprentices in the *High* Effort group move to the *Low* Effort profile in semester 2.

Between semesters 3 and 4, there are some examples of more stable flows of apprentices. For example, 60% of the apprentices in profile *G* move to profile *H*; both profiles are alike in the meaningful dimensions except for Consistency. Another example is the flow from profile *C* to *A*; the majority (70%) of the apprentices in profile *C* move to profile *A*. Both profiles have similar characteristics: *High* Effort and *Low* Help-Seeking. Finally, 60% of the apprentices from profile *F* move to profiles with *Low* Effort and Quality (profiles *B*, *H*).

In summary, this exploration only partially supports our hypothesis **(H2)**. While some apprentices indeed improve their SRL behavior and move to better profiles, we also observe

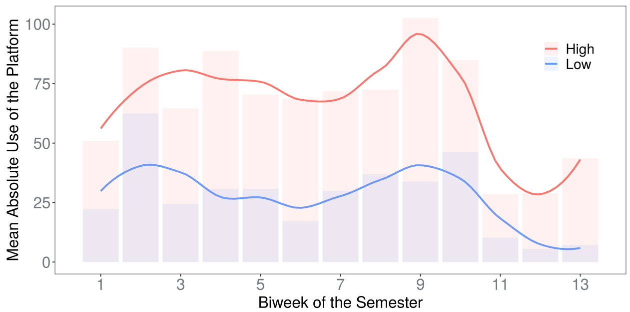


Fig. 4: **Effort** for semester 1.

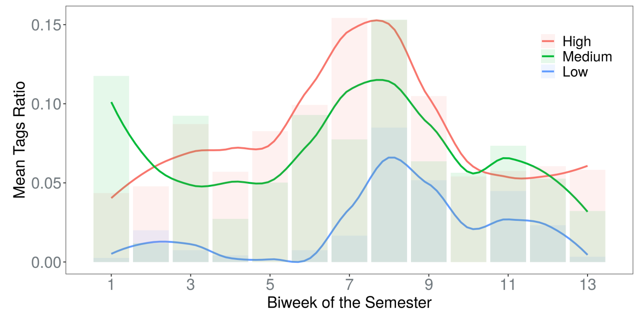


Fig. 5: **Quality** for semester 5.

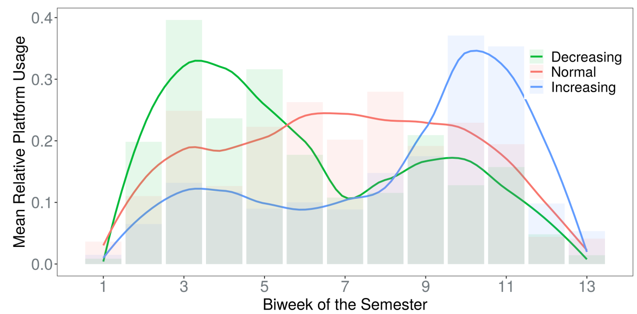


Fig. 6: **Consistency** for semester 2.

apprentices with the opposite behavior and, as shown with the transition from semester 3 to 4, apprentices tend to move to profiles similar to those they were in before **(RQ2)**.

Distinct Behavior Patterns. In a third analysis, we investigated apprentices' behavioral patterns per dimension and semester and developed hypotheses for the expected behavior separately for each dimension. Based on [17], [31], we expected some learners to show higher engagement (Effort) than others and learner engagement to vary over the semester (**H3.1**). For Quality, we hypothesized that it would mainly be dominated by magnitude (*High* vs. *Low*) [1] (**H3.2**). For Consistency, prior findings suggest that some apprentices would exhibit high consistency [19], while others would increase [20] or decrease engagement over time [20], [49] (**H3.3**). For Help-Seeking Behavior, we hypothesized that we would observe patterns of higher and lower activity [16] as well as changes over time, e.g., apprentices stopping to ask for help or exhibiting an increased number of feedback requests over time (**H3.4**). Finally, based on [13], we expected that Regularity would be dominated by magnitude (*High* vs. *Low*) (**H3.5**).

In the following, we discuss each dimension for a selected semester. Note that not all the dimensions are relevant for all the semesters. For each dimension, we have therefore selected a semester where this dimension can be considered relevant for all profiles (i.e., the interpretation labels were significant ($\geq 65\%$) for each profile in that semester). In Figs. 4, 5, 6, and 7, the x-axis denotes the biweeks of the semester and the y-axis the explored feature of the respective dimension.

In terms of ***Effort***, Fig. 4 shows the platform usage in terms of the number of writing events for semester 1. We obtain two clusters of similar shape, with one cluster (*Low*) spending considerably less time on the platform than the other cluster (*High*). In semester 1, profiles *A* and *C* have a *High* Effort pattern, whereas profile *B* has a *Low* Effort pattern. We obtain similar patterns for the other features in this dimension.

As an example for the ***Quality*** dimension, Fig. 5 shows the average tags ratio per group (*High*, *Medium*, and *Low*) and biweek. Analogously to Effort, the difference in the three patterns stems from the magnitude rather than the shape: apprentices in cluster *High* add on average more tags per event than the apprentices from the other groups. Again, we obtain similar patterns for the other features in this dimension.

For ***Consistency***, Fig. 6 shows the mean relative platform use per biweek for semester 2. Compared to Effort and Quality, Consistency patterns differ in shape. Two clusters worked more at the beginning (*Decreasing*) and end (*Increasing*) of the semester, while a third one worked consistently (*Normal*). The other two features in this dimension exhibit similar patterns.

Regarding ***Help-Seeking Behavior***, we obtain two different clusters in semester 3, : *High* and *Low*. Fig. 7 shows the mean feedback request ratio per biweek for the two groups. The feedback response ratio was generally low for both groups, but relatively higher for the group with *High* feedback requests.

Finally, Fig. 8 illustrates features Peak on the Biweek (a) and Periodicity of Day Hour (b) for ***Regularity*** in semester 5. We observe three different groups: *High Peaks*, *Low Peaks*, and *Alternate Peaks*. The *High Peaks* group represents apprentices that work mostly on specific days and hours, while apprentices in the *Low Peaks* cluster tend to work on the platform on different days and hours. Cluster *Alternate Peaks* contains apprentices that work mainly on the same days per biweek, but do not have a preferred day hour. Fig. 9 shows the average platform use per day of the biweek for two example apprentices, one apprentice belonging to cluster *Low Peaks* and the other one belonging to cluster *High Peaks*. The days marked in red are the days apprentices go to school. The apprentice of the *High Peaks* cluster has a strong preference for working on school days. The apprentice from the *Low Peaks* cluster tends to work every day of the biweek.

In summary, we observe distinct behavioral patterns within each dimension. Contrary to our hypothesis **(H3.1)**, absolute student engagement does not vary over the semester. It seems that highly engaged learners spend more time on the platform in general. For Quality, we observed three levels of magnitude (*High*, *Medium*, *Low*), confirming our hypothesis **(H3.2)**. For Consistency, our hypothesis **(H3.3)** is also confirmed as we

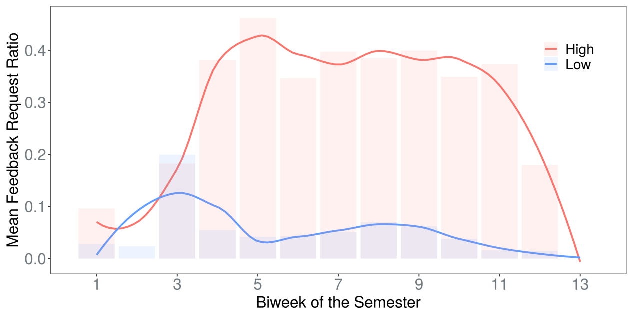
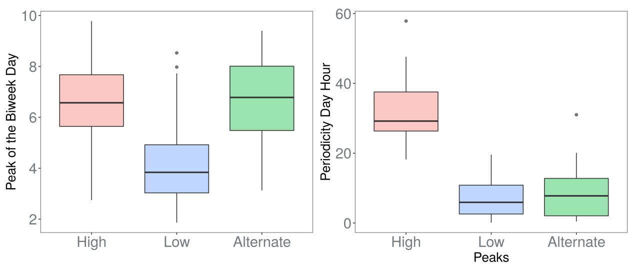


Fig. 7: **Help-Seeking Behavior** for semester 3.



(a) Peak of the Biweek Day (b) Periodicity Day Hour

Fig. 8: **Regularity** for semester 5.

observe *Normal*, *Increasing*, and *Decreasing* patterns. Regarding Help-Seeking Behavior **(H3.4)**, our hypothesis is partially confirmed: we observe low and high activity as well as *Increasing* patterns (see profile *E*), but no *Decreasing* patterns. Similarly, our hypothesis **(H3.5)** can be partially confirmed for Regularity. We observe apprentices with high (*High Peaks*) and low (*Low Peaks*) Regularity. However, a third group prefers specific days, but not hours (*Alternate Peaks*) **(RQ3).**

B. Profile Comparison

In our second experiment, we compared profiles across different contexts. The learning process is influenced by the environment and context [35], [36]. The latter can be characterized as expansive (e.g., having time off the job for college attendance and reflection) or restrictive (e.g., being all-on-job), and these characteristics have an impact on the learning environment. Despite following the same training plan, the two considered schools are located in two different regions, have different teachers, and a dissimilar periodicity. Given that the school training is organized weekly instead of biweekly (like in `VS1`); the apprentices in `VS2` have more constant contact with the teachers and supervisors; hence, we hypothesized that there might be less need to develop SRL strategies. This could result in more stable clusters than those obtained for `VS1` (H4).

To test this hypothesis, we applied our pipeline to data from the second vocational school `VS2` (using weeks instead of biweeks as a basis for computing the time series). Table IV shows the obtained profiles for `VS2`. We identified 12 distinct profiles, with five of them (*B, C, F, H, K*) present also in `VS1`. Fig. 10 shows the evolving apprentice profiles over time.

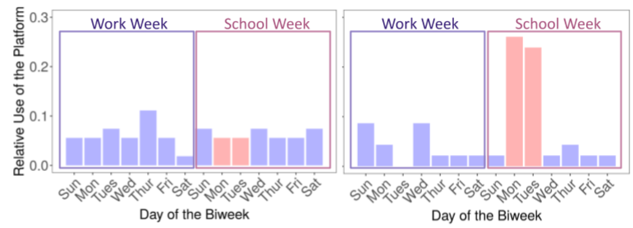


Fig. 9: Examples of intra-biweek Regularity for semester 5.

Profile	Effort	Quality	Consistency	Help Seeking	Regularity
B	Low	Low	Increasing	Low	Low Peaks
Sem. 1	**	***	**	***	+
Sem. 3	***	**	***	***	***
Sem. 6	***	***	*	***	*
C	High	High	Decreasing	Low	Low Peaks
Sem. 4	***	***		***	***
F	Low	Low	Normal	Low	Low Peaks
Sem. 4	***	***	+	***	***
Sem. 5	***	***		***	**
H	Low	Low	Decreasing	Low	Low Peaks
Sem. 6	***	***	***	***	***
K	High Increasing	Medium	Normal	Low	Alt. Peaks
Sem. 1	***	+	+	***	+
P	Low	High	Decreasing	Low	Low Peaks
Sem. 1	**	**	***	***	**
Sem. 4	***		***	***	
Q	Low	Low	Decreasing	Decreasing	Low Peaks
Sem. 2	**	***		***	+
R	Low	High	Normal	Decreasing	Low Peaks
Sem. 2	+	*	*	**	***
S	High	Low	Normal	Decreasing	High Peaks
Sem. 2	**	***	***	***	*
T	High	Low	Decreasing	Low	High Peaks
Sem. 3	+	**	***	***	*
Sem. 5	***	*	*	**	+
W	High	Low	Normal	Low	Low Peaks
Sem. 5	***	+	***	***	***
X	High	Low	Decreasing	Low	Low Peaks
Sem. 6	+	**	***	***	

*** $\geq 90\%$, ** $\geq 80\%$, * $\geq 75\%$, + $\geq 65\%$

TABLE IV: Interpretation of apprentice profiles based on the dimensions for the `VS2`. Overall, we observe 12 distinct profiles, 5 shared with the `VS1` (profiles *B, C, F, H* and *K*).

As observed for `VS1`, some profiles are repeated in several semesters. Profiles *B, F, P*, and *T* appear more than once. A key difference between the two populations is that the profiles in `VS2` seem more stable, in particular between semesters 2 and 4. In the transition between semesters 3 and 4, profile *S* dissolves almost completely to form profile *T*; both profiles have *High* Effort, *Low* Quality, and *High Peaks* in Regularity. From semester 4 to semester 5, profile *P* is formed out of profile *T*. Profiles *P* and *T* both have a *Decreasing* Consistency,

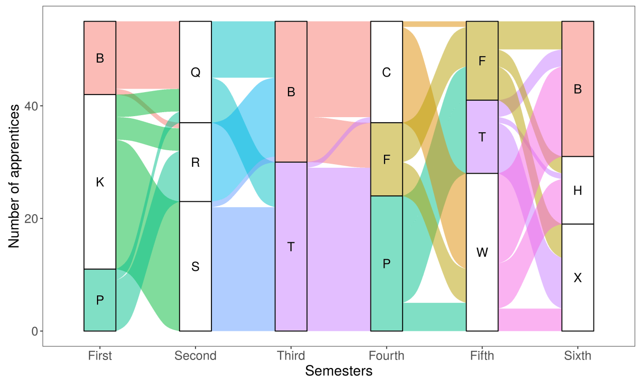


Fig. 10: Apprentice profiles over semesters for `VS2`. Each letter denotes an apprentice profile described in Table IV. The boxes of the profiles occurring in more than one semester (e.g., profiles *B*, *P*, *F*, and *T*) are colored.

Low Help-Seeking Behavior, and *Low Peaks* in Regularity. Differently from `VS1`, `VS2` shows more variability in Effort and Quality. For `VS1`, most profiles with *High* Effort also have *High* Quality, while six of the new seven profiles of `VS2` (i.e., not present in `VS1`) have mismatches on the levels in these dimensions (i.e., *Low* Effort, but *High* Quality and vice-versa).

Furthermore, compared to `VS1`, there is less movement of apprentices across profiles. For example, between semester 2 and 3 most apprentices from Profile *S* move to Profile *T*, and from semester 3 to 4, most apprentices in Profile *T* move to Profile *P*. Profiles *S* and *T* have a very similar characterization and a possible explanation is that the apprentices did not change their SRL strategies from the end of their first year to the beginning of the second year. Interestingly, Profile *P* has *Low* Effort compared to the *High* Effort exhibited by the apprentices in the previous semesters.

In summary, our second experiment supports our hypothesis **(H4)**. Compared to `VS1`, we observed fewer movements between profiles across semesters. We hypothesize that apprentices have more stable SRL patterns as a result of the weekly interaction with teachers and supervisors but future work is needed to explore this in-depth.

VI. Discussion

In this paper, we aimed at identifying apprentices' profiles as a basis for adaptive guidance. We were interested in answering the following research questions: Can we identify interpretable profiles of apprentices integrating different behaviors, and are these profiles related to academic performance (**RQ1**)? How do these profiles evolve throughout the apprenticeship (**RQ2**)? What type of behavioral patterns (in terms of effort, quality, consistency, help-seeking behavior, and regularity) can we observe during a semester (**RQ3**)? How do the learner profiles compare across vocational schools (**RQ4**)?

A. Lessons Learned

Our findings show that it is possible to identify interpretable multi-dimensional profiles (**RQ1**). However, not all dimensions are meaningful to the same extent. For example, effort appears to be constant in several profiles (e.g., *A, C, E, K, L* have a *High* effort and *B, D, F, G, H* have a *Low* effort). Profiles with the same effort magnitude differ based on other dimensions (e.g., Consistency). This is in line with [18], where three groups showed the same effort but a different consistency. Interestingly, a profile with a *Low* or *Decreasing* pattern in all dimensions (profile *H*) was also found among university students [31], but included dropping students.

Compared to prior work profiling learners based on individual SRL aspects (e.g., [19], [17], [16]), we integrated five dimensions into a combined profile. Our study therefore better reflects the diverse nature of learning. These differences could be attributed to internal factors (e.g., individual differences [39]) or external factors (e.g., differences in the CoP [34] or of the learning environment [48]). Though school segments, teachers, structure, and design of the apprenticeship were the same for all apprentices, the *participative memory* and tradition of apprenticeship of the CoP (people and relationships) varied. The behavioral differences may be the result of the work and inclusion culture of the CoP [34], [48]. We also hypothesize that the VET profiles are more diverse than the profiles from pure school-based settings given that, in VET, each learner is exposed to a different CoP. In addition, we would expect school-based profiles to show higher regulation as these learners have more frequent interactions with the teachers. Future work is required to test these hypotheses.

Our analyses also confirmed that there were some significant differences in academic performance between the profiles (**RQ1**). These results are coherent with findings from prior work (e.g.,[19], [13], [20], [16]) showing that achievement was significantly higher for students with high SRL skills (focusing on a single dimension). Our results are also in line with [17] and [1], who demonstrated that apprentices in high performing profiles often exhibit either *High* Effort, *High* Quality, or both. However, our results also show that it is important to take into account the dependencies between SRL dimensions and analyze them in combination rather than focusing on an isolated dimension. While prior work [20], [19] for example found that students who worked consistently exhibited a higher academic performance, our findings demonstrate that working consistently is not enough: profile *C* (Increasing Consistency) for example shows a significantly better performance than profile K (Normal Consistency). Moreover, differently from [16], Help-Seeking Behavior did not appear to have much weight in our profiles. And while regularity was demonstrated to be a predictor of academic performance in MOOCs [13], in our case Regularity was often overruled by other dimensions.

We then showed that the profiles considerably change and evolve throughout the apprenticeship (**RQ2**). Similarly to [14], we found that the number of clusters and cluster size varies over time. One reason why the clusters in [14] are more stable than ours might be that they did not study the flow of learners across clusters, but mainly the evolution of the number and the size of the clusters. They also dealt with cluster evolution over short sessions in an intelligent tutoring system, whereas our evolution covers a much longer time frame (three years) where personal development and changes in the learning environment influence the SRL behavior more [34]. The influence of the time frame on the cluster stability was also observed in MOOCs [20] in terms of transition between clusters in subsequent course weeks. In addition, we observed patterns of growth where apprentices move from to profiles with stronger SRL behaviors and vice-versa. These findings spark new research opportunities to study why some of the apprentices do not increase their SRL skills and how to build interventions to assist them in developing stronger skills.

In a third analysis, we investigated the apprentices' distinct behavioral patterns separately for each dimension in a single semester (**RQ3**). The shapes of the *Increasing* and *Decreasing* patterns found for the Consistency dimension are very similar to the ones found by [19]. At a first glance, an *Increasing* Consistency pattern could be a sign that the apprentices are procrastinating [49]. However, [19] hypothesized that the increasing pattern could be a consequence of active delay (i.e. learners deliberately delaying tasks because they prefer to work under pressure). The latter interpretation might explain why there are no significant differences in the academic performance between the following pairs of profiles differing only in the Consistency dimension: *B* (*Increasing*) and *H* (*Decreasing*) , and *A* (*Increasing*) and *C* (*Increasing*).

Moreover, our results are in line with [17] and [1], showing that patterns in the Effort and Quality dimensions are mainly driven by magnitude (*High*, *Medium*, *Low*). [17] found a significant correlation between effort and self-efficacy (students' confidence about their learning); it is hence possible that the learners that exhibit *Low* patterns of Effort are also underestimating their capacities (low self-efficacy).

Coherent with [16], we also found patterns of high and low help-seeking activity. However, in contrast, to [16], we also observed behavioral changes over time (*Increasing* Help-Seeking Behavior). From an SRL perspective, help-seeking is a social dimension that involves the in-company trainers. An *Increasing* Help-Seeking behavior could indicate that the apprentice moved to the zone of proximal development (ZPD) [34] and thus required guidance and benefit from feedback.

[13] found that in the case of MOOCs, students can be divided into high (working on specific weekdays and hours) and low regulators (not showing a preference for specific days and hours). While we also identified these patterns (*High Peaks*, *Low Peaks*), in our case a third pattern emerged, describing apprentices who work on specific days but do not have a preference for the hour (*Alternate Peaks*). [13] argued that regularity is influenced by both internal and external factors, for example, learners with *High* Regularity might also be more motivated. However, the authors also showed that external factors like employment also impact the regularity patterns. It is therefore likely that the schedule and dynamics of the apprentices' CoP influenced their working patterns.

We also showed the learner profiles emerged from an independent population of a different vocational school (**RQ4**). We found that apprentices' behavior is influenced by the learning environment, as only a subset of the obtained profiles appeared for both populations. This is in line with prior research, which showed 1) the influence of the learning context and CoP on apprentices' learning process [35], [36], 2) the influence of the learning environment (i.e. expansive versus restrictive) on apprentice learning [48], and 3) the significant association between workplace context and SRL behavior [46].

Despite the different environments, the two populations share common profiles (*B, C, F, H, K*). Notably, two of the profiles appearing for both data sets showed significant differences from all other profiles in terms of academic performance (on `VS1`). Profile *C* describes apprentices with optimal SRL behavior, while apprentices in profile *B* exhibit a suboptimal behavior in every dimension. Moreover, we found less movement or growth between the profiles, suggesting that the weekly structure and supervision can support students in maintaining their skills and a biweekly structure might force apprentices to develop the skills on their own.

B. Limitations

While our results are promising, some limitations should be considered. First, learning is a complex phenomenon, and our analysis was restricted to log data from learning journals. As a consequence we can only measure and study logged online behavior, excluding other learning aspects (e.g., time spent in school, conversations with supervisors). From a usability point of view, one of the strengths of this study is that the data was collected from an in-use real-world system not particularly designed to study SRL; as a consequence, our pipeline can be extended to other systems without the need to run controlled experiments or collect extra information. Nevertheless, from a research point of view, only having quantitative data on learners' behavior, limits us from understanding why they did it. We used indicators from previous work that aim to approximate certain behaviors and measure abstract constructs (e.g., effort and consistency); however, a key challenge is whether our measurements (features) accurately capture the considered constructs. For example, if a learner works intensively and consistently in an offline editor and then pastes the content at the end of the semester, the logged data will not register the time spent and frequency. In future work, we will triangulate the data from teacher and self-reported questionnaires to better evaluate our measurements. In addition, teachers' qualitative input could be used to generate context-specific hypotheses.

Second, the absence of pre-and post-tests is a challenge for result interpretation. For example, apprentices with *High* Effort generally had good grades, but that does not necessarily mean that spending more time on the platform will increase the academic performance of another apprentice.

Third, identifying interpretable multi-dimensional learner profiles may come at the expense of added complexity. However, other methods that are computationally less expensive and still valid like latent class analysis (e.g., [33], [20]) are not able to take the temporal aspect into account or to represent the dependencies between different learning dimensions.

Fourth, our analysis is influenced by the underlying numerical assumptions of the clustering methods chosen (Spectral Clustering and K-Modes). Both are hard clustering methods, requiring to assign every apprentice to exactly one cluster. Soft clustering methods (e.g., Gaussian Mixture Models and Latent Class Analysis), which instead output the probability of belonging to each cluster, can be explored. Further research is needed to integrate such approaches into our pipeline and evaluate them in comparison to the advantages of the current methods (identification of outliers and non-convex clusters).

Lastly, SRL depends on the context of learning. Transferring the learner profiles from the apprenticeship context to a different context could lead to misinformed decisions and the exclusion of learners from potential beneficial interventions. Alternatively, our flexible pipeline should be implemented in the new context.

C. Implications and Recommendations

Our analysis describes how apprentices use the platform through different lenses. With this information, teachers, in-company trainers, and program designers can reflect on whether these are the desired patterns and, if not, what can be done to intervene and improve the apprentices' learning experience. For instance, Profile *B* exhibited significantly worse academic performance than other profiles. Thus, if we identify the apprentices that would be in Profile *B* early in the semester, we could intervene to improve their learning experience and performance. More generally, the findings can suggest possible platform modifications tailored to individual learner profiles. Another implication is that the learners have different behavioral patterns and profiles. Future work on online learning journals and future interventions must acknowledge this diversity. For example, learners in profiles with *Low* Regularity could receive personalized reminders on specific days of the week to encourage them to work more regularly. Possible support for learners with *Low* Consistency would be to add a dashboard where they can visualize their consistency patterns and badges can be awarded for the desired patterns. Reflection prompts could be personalized to encourage learners that struggle with the quality of their reflective entries to reflect deeper and peer examples or auto-generated feedback could be shown to apprentices to get inspired or increase the feedback response rate for learners in profiles that struggle with help-seeking. A dashboard could allow teachers to monitor such patterns. Moreover, future work must study apprentices together with their CoP, and interventions should recognize its critical role.

To conclude, this work contributes to the ongoing research of reusable analytics. To study SRL behavior, we proposed a new generalizable pipeline applied across contexts and settings, contributing to the generality of theories and evaluating transfer of SRL patterns. Our work showcases the potential learning analytics has in VET, serving as a starting point for data-driven learning journal explorations and data-driven support to teachers for designing interventions in VET.

References

[1] B. A. Schwendimann, G. Kappeler, L. Mauroux, and J.-L. Gurtner, "What makes an online learning journal powerful for VET? Distinguishing productive usage patterns and effective learning strategies," *Empir. Res. Vocat. Educ. Train.*, vol. 10, no. 1, pp. 1–20, Jul. 2018, doi: 10.1186/s40461-018-0070-y.

[2] M. Nückles, S. Hübner, and A. Renkl, "Enhancing self-regulated learning by writing learning protocols," *Learn. Instruct.*, vol. 19, no. 3, pp. 259–271, Jun. 2009, doi: 10.1016/j.learninstruc.2008.05.002.

[3] J. A. Moon, *Learning Journals: A Handbook For Reflective Practice and Professional Development*. Routledge, 2006, doi: 10.4324/9780203969212.

[4] V. Caruso, A. Cattaneo, and J. Gurtner, "Exploring the potential of learning documentation as a boundary object in the Swiss vocational education and training system," *Zeitschrift für Berufs- und Wirtschaftspädagogik*, vol. 29, pp. 213–232, 2020.

[5] L. Mauroux, J. D. Zufferey, E. Rodondi, A. Cattaneo, E. Motta, and J.-L. Gurtner, "Writing Reflective Learning Journals: Promoting the Use of Learning Strategies and Supporting the Development of Professional Skills," in *Writing for Professional Development*, G. Ortoleva, M. Bétrancourt, and S. Billett., Eds. Brill, 2016, pp. 107–128, doi: 10.1163/9789004264830_007.

[6] B. A. Schwendimann, A. A. Cattaneo, J. D. Zufferey, J.-L. Gurtner, M. Bétrancourt, and P. Dillenbourg, "The 'Erfahrraum': a pedagogical model for designing educational technologies in dual vocational systems," *J. Vocat. Educ. Train.*, vol. 67, no. 3, pp. 367–396, Jul. 2015, doi: 10.1080/13636820.2015.1061041.

[7] V. Caruso, A. Cattaneo, and J.-L. Gurtner, "Learning documentations in VET systems: An analysis of current Swiss practices," *Vocat. Learn.*, vol. 9, no. 2, pp. 227–256, Jan. 2016, doi: 10.1007/s12186-016-9149-4.

[8] J. A. de Stavenga Jong, R. F. A. Wierstra, and J. Hermanussen, "An exploration of the relationship between academic and experiential learning approaches in vocational education," *Br. J. Educ. Psychol.*, vol. 76, no. 1, pp. 155–169, Mar. 2006, doi: 10.1348/000709905X42932.

[9] T. S. O'Connell and J. E. Dyment, *Theory into practice: Unlocking the power and the potential of reflective journals*. IAP, 2013.

[10] A. Taylor and S. Freeman, "'Made in the trade': Youth attitudes toward apprenticeship certification," *J. Vocat. Educ. Train.*, vol. 63, no. 3, pp. 345–362, Sep. 2011, doi: 10.1080/13636820.2011.570455.

[11] S. Epp, "The value of reflective journaling in undergraduate nursing education: A literature review," *Int. J. Nurs. Stud.*, vol. 45, no. 9, pp. 1379–1388, Mar. 2008, doi: 10.1016/j.ijnurstu.2008.01.006.

[12] J. Wessel and H. Larin, "Change in reflections of physiotherapy students over time in clinical placements," *Learn. Health Soc. Care*, vol. 5, no. 3, pp. 119–132, Jul. 2006, doi: 10.1111/j.1473-6861.2006.00124.x.

[13] M. S. Boroujeni, K. Sharma, Ł. Kidziński, L. Lucignano, and P. Dillenbourg, "How to Quantify Student's Regularity?"," in *Proc. 11th. Eur. Conf. Technology Enhanced Learning*, Sept. 13-16, 2016, pp. 277–291, doi: 10.1007/978-3-319-45153-$4_2$1.

[14] S. Klingler, T. Käser, B. Solenthaler, and M. Gross, "Temporally Coherent Clustering of Student Data," in *Proc. 9th Int. Conf. Educ. Data Mining*, Jun./Jul. 29-2, 2016, pp. 102–109.

[15] D. Gasevic, J. Jovanovic, A. Pardo, and S. Dawson, "Detecting learning strategies with analytics: Links with self-reported measures and academic performance," *J. Learn. Anal.*, vol. 4, no. 2, p. 113–128, Jul. 2017, doi: 10.18608/jla.2017.42.10.

[16] L. Corrin, P. G. de Barba, and A. Bakharia, "Using Learning Analytics to Explore Help-Seeking Learner Profiles in MOOCs," in *Proc. 7th Int. Conf. Learning Analytics and Knowledge*, Mar. 13-17, 2017, p. 424–428, doi: 10.1145/3027385.3027448.

[17] M.-H. Cho and D. Shen, "Self-regulation in online learning," *Distance Education*, vol. 34, no. 3, pp. 290–301, Oct. 2013, doi: 10.1080/01587919.2013.835770.

[18] S. Mojarad, A. Essa, S. Mojarad, and R. S. Baker, "Data-Driven Learner Profiling Based on Clustering Student Behaviors: Learning Consistency, Pace and Effort," in *Proc. 18th Int. Conf. Intelligent Tutoring Systems*, Jun. 11-15, 2018, pp. 130–139, doi: 10.1007/978-3-319-91464-0_13.

[19] V. Sher, M. Hatala, and D. Gašević, "Analyzing the Consistency in Within-Activity Learning Patterns in Blended Learning," in *Proc. 10th Int. Conf. Learning Analytics & Knowledge*, Mar. 23-27, 2020, p. 1–10.

[20] A. Barthakur, V. Kovanovic, S. Joksimovic, G. Siemens, M. Richey, and S. Dawson, "Assessing program-level learning strategies in MOOCs," *Comput. Hum. Behav.*, vol. 117, p. 106674, Apr. 2021, doi: 10.1016/j.chb.2020.106674.

[21] K. Illeris, "Towards a contemporary and comprehensive theory of learning," *Int. J. Lifelong Educ.*, vol. 22, no. 4, pp. 396–406, 2010, doi: 10.1080/02601370304837.

[22] S. Shen and M. Chi, "Clustering Student Sequential Trajectories Using Dynamic Time Warping," in *Proc. 10th Int. Conf. Educational Data Mining*, Jun. 25-28, 2017, pp. 266–271.

[23] J. Saint, A. Whitelock-Wainwright, D. Gašević, and A. Pardo, "Trace-srl: A framework for analysis of microlevel processes of self-regulated learning from trace data," *IEEE Trans. Learn. Technol.*, vol. 13, no. 4, pp. 861–877, Oct 2020, doi: 10.1109/TLT.2020.3027496.

[24] J. Kinnebrew, D. Mack, and G. Biswas, "Mining temporally-interesting learning behavior patterns," in *Proc. 6th Int. Conf. Educational Data Mining*, Jul. 6-9, 2013, pp. 252–255.

[25] C. F. Jahreie and S. R. Ludvigsen, “Portfolios as boundary object: Learning and change in teacher education,” *Res. Pract. Technol. Enhanc. Learn.*, vol. 2, no. 03, pp. 299–318, 2007.

[26] C. Aprea and A. A. P. Cattaneo, *Designing Technology-Enhanced Learning Environments in Vocational Education and Training*. John Wiley Sons, Ltd, 2019, ch. 19, pp. 373–393, doi: 10.1002/9781119098713.ch19.

[27] J. Dehler Zufferey, L. Mauroux, F. Jimenez, and J. Gurtner, “Learning journals in VET: Testing of a paper-based vs. a mobile and online tool,” in *Congress on Research in Vocational Education and Training in Switzerland, Zollikofen*, 2011.

[28] A. Gleaves, C. Walker, and J. Grey, “Using digital and paper diaries for learning and assessment purposes in higher education: a comparative study of feasibility and reliability,” *Assess. Eval. High. Educ.*, vol. 32, no. 6, pp. 631–643, 2007, doi: 10.1080/02602930601117035.

[29] W. Kicken, S. Brand-Gruwel, J. J. Van Merriënboer, and W. Slot, “The effects of portfolio-based advice on the development of self-directed learning skills in secondary vocational education,” *Educ. Technol. Res. Dev.*, vol. 57, no. 4, p. 439, 2009.

[30] E. Boldrini and A. Cattaneo, “Scaffolding collaborative reflective writing in a vet curriculum,” *Vocat. Learn.*, vol. 7, no. 2, pp. 145–165, 2014, doi: 10.1007/s12186-014-9110-3.

[31] M. Khalil and M. Ebner, “Clustering patterns of engagement in Massive Open Online Courses (MOOCs): the use of learning analytics to reveal student categories,” *J. Comput. High. Educ.*, vol. 29, no. 1, pp. 114–132, 2017, doi: 10.1007/s12528-016-9126-9.

[32] J. Broadbent and M. Fuller-Tyszkiewicz, “Profiles in self-regulated learning and their correlates for online and blended learning students,” *Educ. Technol. Res. Dev.*, vol. 66, no. 6, pp. 1435–1455, 2018, doi: 10.1007/s11423-018-9595-9.

[33] S.-W. Lin and W.-C. Tai, “Latent class analysis of students’ mathematics learning strategies and the relationship between learning strategy and mathematical literacy.” *Univers. J. Educ. Res.*, vol. 3, no. 6, pp. 390–395, 2015.

[34] J. Eberle, “Apprenticeship learning,” in *Int. handbook of the learning sciences*. Routledge, 2018, pp. 44–53, doi: 10.4324/9781315617572.

[35] J. Lave and E. Wenger, *Situated learning: Legitimate peripheral participation*. Cambridge University Press, 1991, doi: 10.1017/CBO9780511815355.

[36] E. Wenger, *Communities of practice: Learning, meaning, and identity*. Cambridge University Press, 1999, doi: 10.1017/CBO9780511803932.

[37] K. Illeris, *The fundamentals of workplace learning: Understanding how people learn in working life*. Routledge, 2010, doi: 10.4324/9780203836521.

[38] A. A. Cattaneo and E. Motta, “‘I Reflect, Therefore I Am… a Good Professional’. On the Relationship between Reflection-on-Action, Reflection-in-Action and Professional Performance in Vocational Education,” *Vocat. Learn.*, pp. 1–20, 2020, doi: 10.1007/s12186-020-09259-9.

[39] E. Panadero, “A Review of Self-regulated Learning: Six Models and Four Directions for Research,” *Front. Psychol.*, vol. 8, p. 422, 2017, doi: 10.3389/fpsyg.2017.00422.

[40] P. H. Winne and N. E. Perry, “Chapter 16 - measuring self-regulated learning,” in *Handbook of Self-Regulation*. San Diego: Academic Press, 2000, pp. 531–566, doi: 1016/B978-012109890-2/50045-7.

[41] J. Broadbent and W. Poon, “Self-regulated learning strategies academic achievement in online higher education learning environments: A systematic review,” *Internet High. Educ.*, vol. 27, pp. 1–13, 2015, doi: 10.1016/j.iheduc.2015.04.007.

[42] P. R. Pintrich and E. V. de Groot, “Motivational and self-regulated learning components of classroom academic performance,” *J. Educ. Psychol.*, vol. 82, no. 1, pp. 33–40, Mar. 1990, doi: 10.1037/0022-0663.82.1.33.

[43] A. Margaryan, C. Milligan, A. Littlejohn, D. Hendrix, and S. Graeb-Koenneker, “Self-regulated learning and knowledge sharing in the workplace,” in *Proc. 14th Int. Conf. Organizational Learning, Knowledge and Capabilities*, 2009, pp. 1–15.

[44] S. Paavola, L. Lipponen, and K. Hakkarainen, “Models of Innovative Knowledge Communities and Three Metaphors of Learning,” *Rev. Educ. Res.*, vol. 74, no. 4, pp. 557–576, 2004, doi: 10.3102/00346543074004557.

[45] A. Littlejohn, C. Milligan, and A. Margaryan, “Charting collective knowledge: supporting self-regulated learning in the workplace,” *J. Workplace. Learn.*, vol. 24, no. 3, pp. 226–238, 2012.

[46] C. Milligan, R. P. Fontana, A. Littlejohn, and A. Margaryan, “Self-regulated learning behaviour in the finance industry,” *J. Workplace. Learn.*, vol. 27, no. 5, pp. 387–402, Jul. 2015, doi: 10.1108/JWL-02-2014-0011.

[47] M. Siadaty, D. Gaević, and M. Hatala, “Measuring the impact of technological scaffolding interventions on micro-level processes of self-regulated workplace learning,” *Comput. Hum. Behav.*, vol. 59, pp. 469–482, Nov. 2016, doi: 10.1016/j.chb.2016.02.025.

[48] A. Fuller and L. Unwin, “Learning as apprentices in the contemporary uk workplace: creating and managing expansive and restrictive participation,” *J. Educ. Work.*, vol. 16, no. 4, pp. 407–426, Jun. 2003, doi: 10.1080/1363908032000093012.

[49] D. Hooshyar, M. Pedaste, and Y. Yang, “Mining educational data to predict students’ performance through procrastination behavior,” *Entropy*, vol. 22, no. 1, Dec. 2019, doi: 10.3390/e22010012.

[50] J. D. Vermunt and V. Donche, “A learning patterns perspective on student learning in higher education: state of the art and moving forward,” *Educ. Psychol. Rev.*, vol. 29, no. 2, pp. 269–299, Mar. 2017, doi: 10.1007/s10648-017-9414-6.

[51] H. Wan, K. Liu, Q. Yu, and X. Gao, “Pedagogical intervention practices: Improving learning engagement based on early prediction,” *IEEE Transactions on Learning Technologies*, vol. 12, no. 2, pp. 278–289, 2019, doi: 10.1109/TLT.2019.2911284.

[52] F. Chen and Y. Cui, “Utilizing student time series behaviour in learning management systems for early prediction of course performance.” *J.of Learning Analytics*, vol. 7, no. 2, pp. 1–17, 2020, doi: 10.18608/jla.2020.72.1.

[53] S. A. Crossley, J. L. Weston, S. T. M. Sullivan, and D. S. McNamara, “The development of writing proficiency as a function of grade level: A linguistic analysis,” *Written Communication*, vol. 28, no. 3, pp. 282–311, 2011, doi: 10.1177/0741088311410188.

[54] A. Y. Ng, M. I. Jordan, Y. Weiss *et al.*, “On spectral clustering: Analysis and an algorithm,” *Adv. Neural Inf. Process. Syst.*, vol. 2, pp. 849–856, Dec. 3-8, 2001, doi: 10.5555/2980539.298064.

[55] U. Von Luxburg, “A tutorial on spectral clustering,” *Stat. Comput.*, vol. 17, no. 4, pp. 395–416, 2007, doi: 10.1007/s11222-007-9033-z.

[56] P. J. Rousseeuw, “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis,” *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, 1987, doi: 10.1016/0377-0427(87)90125-7.

Paola Mejia-Domenzain is a Ph.D. Student at EPFL, Switzerland. She received the M.S. degree in Data Science from ITAM, Mexico in 2020. Her research focuses on learning analytics, student profiling, and vocational education.

Mirko Marras is an Assistant Professor at University of Cagliari, Italy. He received the Ph.D. degree in Computer Science from the same University in 2020. His research focuses on machine learning, personalization, and online education.

Christian Giang is a Postdoctoral Researcher at EPFL and the University of Applied Sciences and Arts of Southern Switzerland, Switzerland. He received the Ph.D. degree in Robotics from EPFL in 2020. His research focuses on educational technologies, personalized learning, vocational education.

Alberto Cattaneo is Professor at the Swiss Federal University for Vocational Education and Training, Switzerland. He received the Ph.D. degree in Social, Developmental and Organizational Psychology from the University of Bologna, Italy, in 2005. His research focuses on instructional design, multimedia learning, and teachers’ professional development, especially when dealing with technology integration.

Tanja Käser is an Assistant Professor at EPFL, Switzerland. She received the Ph.D. degree in Computer Science from ETH Zurich, Switzerland, in 2016. Her research focuses on educational data mining, personalized learning, and vocational education.