



Abstract. *Educational data mining is a developing research trend for exploring hidden patterns and natural associations among a set of student, teacher or school related variables. Discovering profiles of preservice science teachers using data mining methods would give important information about quality of teacher education programs and future science teachers' performance. The aim of this research was to describe characteristics of preservice science teachers and to explore the relations among their motivational beliefs, learning strategy use, and constructivist learning environment perceptions. Participants included 480 preservice science teachers in their final semester of the teacher education program. Data were gathered using Demographic Questionnaire, Motivated Strategies for Learning Questionnaire, Achievement Goal Questionnaire and Constructivist Learning Environment Scale. Findings of clustering analysis revealed gender as a discriminating factor between the obtained two natural groups. Preservice science teachers' characteristics including background characteristics, motivational beliefs, strategy use and constructivist learning environment perceptions were grouped into two clusters, namely males and females. Moreover, the association rules mining analysis revealed strong relations among preservice science teachers' motivational beliefs, learning strategy use, and constructivist learning environment perceptions. This research provided many important findings that can be useful for further decision-making strategies.*

Keywords: *constructivist learning environment, data mining, learning strategies, motivational belief, pre-service science teacher*

**Sevgi Kingir, Bilge Gok,
Ahmet Selman Bozkir**
Hacettepe University, Turkey

EXPLORING RELATIONS AMONG PRE-SERVICE SCIENCE TEACHERS' MOTIVATIONAL BELIEFS, LEARNING STRATEGIES AND CONSTRUCTIVIST LEARNING ENVIRONMENT PERCEPTIONS THROUGH UNSUPERVISED DATA MINING

**Sevgi Kingir,
Bilge Gok,
Ahmet Selman Bozkir**

Introduction

Countries all over the world are struggling to accommodate the demands of everchanging educational needs in today's dynamic world. Scientific and technological breakthroughs alter the way people live, communicate and work, and reshape the economic, social and cultural growth (Schwab, 2016). Many new positions and professions that cannot be foreseen yet will emerge soon (Kumar et al., 2019). The evolving nature of working environment demands highly qualified and skillful personnel who are also able to adapt to change and keep up with the emerging innovations and technologies. The top ten skills that will be demanded in order of priority by employers by 2022 are "analytical thinking and innovation; active learning and learning strategies; creativity, originality and initiative; technology design and programming; critical thinking and analysis; complex problem-solving; leadership and social influence; emotional intelligence; reasoning, problem-solving and ideation; and systems analysis and evaluation" (World Economic Forum, 2018, p. 12). Skills and competencies of future workforce also include working autonomously and self-regulated learning (National Research Council [NRC], 2011).

The proliferation of new technologies not only affects the labor force of countries but also the associated educational system. The skills of future workforce are not easily learned on the job but rather in school settings (NRC, 2011). Quality education and work are the basis of prosperity, dignity, and well-being for individuals and form the backbone of successful economies. Therefore, countries participate in international assessment programs to assess their educational outcomes. Results of the latest international assessments, TIMSS 2015 (Martin et al., 2016) and PISA 2018 (Schleicher, 2019), have indicated that quality of education is neither at the desired level nor keeping up with the needs of a modern economy in many countries including Turkey.

Development of self-regulation is frequently emphasized in the science education literature in order to prepare young people for future career opportunities (Kitsantas et al., 2019). Self-regulation is described as “self-generated thoughts, feelings, and actions that are planned and cyclically adapted to the attainment of personal goals” (Zimmerman, 2000, p. 14). Self-regulation has three cyclical phases encompassing metacognitive, motivational, and behavioral components. The first one, forethought phase, includes analyzing cognitive task, setting goals and choosing a learning strategy for an upcoming task. The second one, performance phase, reflects learning efforts through self-monitoring. The third one, self-reflection, involves evaluating personal effectiveness and monitoring learning performance.

Learners equipped with self-regulatory skills perform tasks strategically, determine goals, choose and use learning strategies, and evaluate their own performance. Learning strategies are grouped based on cognitive, metacognitive and management aspects of learning. Cognitive strategies are categorized as rehearsal (i.e., repeating information from memory), organization (i.e., constructing links among the pieces of information using clustering, outlining etc.), elaboration (i.e., relating new information to already stored knowledge using paraphrasing, summarizing etc.), and critical thinking (i.e., applying acquired knowledge and skills to novel situations) (Pintrich et al., 1991). Cognitive strategies are employed for processing information, while metacognitive strategies are used for controlling and managing cognitive tasks. Metacognitive strategies help individuals in planning, monitoring and regulating their cognition, motivation and behavior (Pintrich, 2002). Management strategies include effort regulation (i.e., ability to manage effort), time and study environment (i.e., management of one's own study time and environment), peer learning (i.e., collaborating with peers), and help seeking (i.e., managing support of others) (Pintrich et al., 1991).

Motivation can be described as “process of instigating and sustaining goal-directed behavior” (Schunk, 2000, p. 300). Self-motivational beliefs act as a driving force for strategy use (Pintrich & De Groot, 1990). Self-efficacy, task value, control of learning beliefs and goal orientation are considered as components of motivational beliefs (Zimmerman, 2000). Self-efficacy is broadly defined as perceptions of people concerning their abilities to attain desired outcomes for specific tasks (Bandura, 1994). Control of learning beliefs is described as people's expectations that positive outcomes are the consequences of their efforts. Studies demonstrated that both control of learning beliefs and self-efficacy are associated with goal orientation, academic achievement and learning strategy use (Kahraman & Sungur, 2013; Sungur, 2007). Individuals perceiving higher self-efficacy and control for their learning are inclined to set challenging goals, participate in science activities and try to use alternative learning strategies in the face of difficulties to accomplish the given task successfully (Kahraman & Sungur, 2013; Sungur, 2007).

Goal orientation and task value are related to students' reasons and purposes for engaging in a learning activity (Eccles & Wigfield, 2002). Task value is defined as “students' perceptions of the course material in terms of interest, importance, and utility” (Pintrich et al., 1991, p. 11). Task value is related to goal orientation, self-efficacy, strategy use and science achievement (Iverach & Fisher, 2008; Kahraman & Sungur, 2013; Pintrich & De Groot, 1990; Sungur, 2007; Sungur & Güngören, 2009). Individuals who attach value to tasks demonstrate higher self-efficacy, academic achievement, and metacognitive strategy use. Meanwhile, goal orientation refers to intentions of an individual's engagement in learning tasks (Schunk, 2000). Elliot and McGregor (2001) categorized achievement goals as mastery approach, mastery avoidance, performance approach and performance avoidance goals. Students pursuing mastery approach goals participate in activities for the sake of enhancing their knowledge and skills, while those adopting mastery avoidance goals abstain from misunderstanding or failure in learning. In a similar vein, individuals having performance approach goals involve in tasks with the aim of demonstrating high ability to other people while those pursuing performance avoidance goals avert from appearing incompetent. Several research studies consistently reported mastery approach goals as associated with positive learning outcomes like task value, self-efficacy, achievement and self-regulation (Elliot & Church, 1997; Iverach & Fisher, 2008; Pintrich & De Groot, 1990; Kahraman & Sungur, 2011; Sungur, 2007; Sungur & Güngören, 2009).

Students' perceptions and experiences about the learning environment significantly affect their learning process (Fraser, 2007). In a learning environment designed with the principles of constructivism, individuals engage in the social negotiation process and construct their own knowledge through the integration of new information with prior learning (Fraser, 2002). Such a constructivist learning environment provides opportunities for personal relevance (i.e., linking learning with daily life experiences), uncertainty (i.e., evolving nature of knowledge in science), critical voice (i.e., questioning the information being presented), shared control (i.e., balanced interaction in which both students and teacher have some control for cognitive task) and student



negotiation (i.e., sharing ideas with other people) (Taylor et al., 1997). Constructivist classroom environment results in positive student outcomes like motivational beliefs, strategy use and academic achievement (Arisoy et al., 2016; Kingir et al., 2013; Yerdelen & Sungur, 2019).

One of the important variables influencing perception of the classroom environment, motivation, achievement, and self-regulatory processes is gender (Pajares, 2002). Indeed, gender equity is commonly investigated in science education literature (Scantlebury, 2012). Gender differences in strategy use were frequently reported in previous literature, in the favor of female students (Pajares, 2002; Yerdelen & Sungur, 2019). For example, Arisoy et al. (2016) revealed that females held higher levels of classroom environment perceptions and motivational beliefs compared to male counterparts. However, self-efficacy was not significantly different across males and females; while it was found significant across gender in a research of Britner and Pajares (2006). Kahraman and Sungur-Vural (2014) could not find any significant gender difference based on elementary school students' task value. A non-significant gender difference was also shown in learning strategy use (Kiran & Sungur, 2012).

The literature mentioned above illustrates that several studies examined the relations among student characteristics, motivation, self-regulation and constructivist learning environment perceptions in different combinations (e.g., Kingir et al., 2013; Sezgintürk & Sungur, 2020; Sungur, 2007). In those previous studies conducted generally with elementary school students, statistical analysis, which forms a hypothesized model and test against data, were used. However, relations among the variables of those studies were not analyzed for embedded significant patterns, which could be useful for educators in decision making to enhance quality of education and accordingly increasing student performance (Ranjan & Malik, 2007). To bridge this gap, this research focused on the analysis of relations among the variables by using unsupervised data mining methods. Aran et al. (2019) defined data mining as a process that reveals meaningful and useful knowledge within the data via methodologies such as structured learning, statistics and machine learning. It also refers to knowledge discovery process within huge datasets, which makes it different from more traditional statistical approaches since it does not rely on predefined hypotheses. Data mining methods do not assume a particular model, rather they automatically extract hidden patterns in data (Dogan & Camurcu, 2008). Moreover, recent years have witnessed the growing body of research regarding application of data mining in educational settings (Aldowah et al., 2019; Kiray et al., 2015).

Data mining (DM) approaches are fundamentally grouped under two categories such as predictive and descriptive (Aran et al., 2019). The predictive DM's goal is to make predictions on new cases by leveraging past examples via supervision of target variables by employing supervised learning techniques. On the other hand, descriptive DM deals with uncovering the hidden patterns embedded in the data without needing any target outcome. In this regard, it can be easily seen that descriptive DM benefits from unsupervised methods such as clustering, auto encoding (Baldi, 2012). Similarly, the approach of unsupervised data mining aims to discover the natural relations in the dataset by requiring no "labels". Thus, without any ground truth labels/outcomes, it enables to extract meaningful and hidden patterns in the data that can further be used for the task of interest. In this regard, two different unsupervised data mining methods namely K-means clustering and generalized rule induction (GRI) were used to reveal useful patterns in this dataset. K-means clustering was employed algorithm to segment the participants based on their attributes in a natural way having no bias. Later, the properties of the obtained clusters were examined to gain more insight into the "big" picture. Moreover, the GRI algorithm was applied to reveal the associative patterns hidden in the views of the participants concerning the items on the applied scale.

The aforementioned studies also indicate that quality of education can be improved by creating constructivist learning environments and developing students' motivational beliefs, self-regulation and in turn conceptual understanding. Teacher knowledge and skills influence meeting higher standards in education. For learners, to determine what they need to know, and devise strategies to acquire the required knowledge is not an easy task. To be successful, teachers should be competent as self-regulated. Today's teachers have to possess adaptive motivational beliefs, perceptions of constructivist learning environment, and effective strategy use to promote high quality learning at schools (Balyer & Özcan, 2014; Yerdelen & Sungur, 2019). The fact that skills are not easily acquired in the field through short term seminars (NRC, 2011) unfolds the importance of preservice teacher education. Besides, Turkish students' relatively low science scores in international assessments are a driving force for examining existing beliefs, skills, and perceptions of preservice science teachers. Therefore, in this research, the aim was to describe characteristics of preservice science teachers and further explore relations among preservice science teachers' motivational beliefs, learning strategies and constructivist learning environment



perceptions in one research context using unsupervised data mining methods. Accordingly, main questions specified in this research were the following:

1. What are the profiles of preservice science teachers with respect to background characteristics, motivational beliefs, learning strategies and constructivist learning environment perceptions?
2. What are the relations among preservice science teachers' background characteristics, motivational beliefs, learning strategies and constructivist learning environment perceptions?

Research Methodology

General Background

A cross-sectional, descriptive research was employed as a non-experimental quantitative research design. Primary objective of descriptive research is to describe characteristics of individuals regarding a topic or event or interest, skill, ability, attitude, etc. on larger samples. It is also used to describe associations that exist among the variables. Additionally, a research is cross-sectional if data collection is carried out at a single point in time (Johnson & Christensen, 2004). Data of the present research were collected from preservice science teachers in May 2018 using self-report instruments. Unsupervised data mining methods were utilized for the analysis of data. The data were partitioned into meaningful segments via clustering analysis. Association rules mining was further employed to discover hidden association patterns embedded in the data.

Participants

Participants included 480 preservice science teachers in their final semester of their teacher education program at a public university in a larger city located in the Middle Region of Turkey. Many times, it is difficult to select either a random or systematic non-random sample. In such cases, convenience sampling can be used. In this type of sampling, a certain group of people are chosen because of their availability and easy access (Fraenkel et al., 2012). Accordingly, participants of this research included preservice science teachers who were selected according to the method of convenience sampling.

When background characteristics of preservice science teachers included in this research were analyzed, it was noticed that 66.50% ($n = 319$) of the participants were female and 19.80% ($n = 161$) were male. 87.50% of the participants ($n = 420$) reported that they wanted to be a teacher after graduation from the university while 12.50% ($n = 60$) of them stated that they did not want to be a teacher after graduation. Parent education level of the participants was quite low: majority of the parents were graduates of a high school or lower level of education. Education level of mothers was lower than that of fathers. The percentage of fathers graduated from a university was 25.2%; while it was 7.7% for the mothers. The percentage of preservice science teachers who reported their appointment possibility as a teacher after graduation between 0-20 was 19.8%, while that of between 21-40, 41-60, 61-80 and 81-100 were 12.7%, 20.6%, 28.8% and 18.1%, respectively.

Data Collection

Data were gathered using Demographic Questionnaire, Motivated Strategies for Learning Questionnaire, Achievement Goal Questionnaire and Constructivist Learning Environment Scale. Permission to gather data was provided from the university before conducting the research. Instruments were administered to the preservice science teachers participating in this research by one of the researchers and lasted about 50 min. Before being presented with the instruments, participants were given information regarding the aim and importance of the research. They were also ensured of their voluntary participation, confidentiality of all information provided, and reminded that they were free to refuse to take part in the research and to leave the research whenever they want for any reason.

The confirmatory factor analysis (CFA) was utilized to find out whether factor structures of scales used in this research were confirmed by the collected data. CFA reveals whether the established models by the data and whether the assumed relations in the theoretical population exist in the dataset obtained as a result of empirical observation (Şimşek, 2007). The structures of the Achievement Goal Questionnaire and Constructivist Learning



Environment Scale were determined by using diagonally weighted least square (DWLS) method, a method of estimation resistant to the violation of normality assumption and in this method variables are categorical. The maximum likelihood ratio (MLR) was used in analyzing the factor structure since Motivated Strategies for Learning Questionnaire had seven categories. Model fit was analyzed using chi-square to degrees of freedom ratio (χ^2/df), 90% confidence interval (CI) of the root mean square error of approximation (RMSEA), comparative fit index (CFI), and standardized root mean square residual (SRMR), based on the criteria shown in Table 1 (Kline, 2005). Detailed information regarding questionnaires and scales used in this research were provided below.

Table 1
Criteria for model fit

Fit indices	Good fit	Acceptable fit
χ^2/df	$0 \leq \chi^2/df \leq 2$	$2 \leq \chi^2/df \leq 5$
RMSEA	$0 \leq RMSEA \leq .05$	$.05 \leq RMSEA \leq .10$
CFI	$.95 \leq CFI \leq 1$	$.90 \leq CFI \leq .95$
SRMR	$0 \leq SRMR \leq .05$	$.05 \leq SRMR \leq .10$

Demographic Questionnaire

There were six items related to preservice science teachers' background characteristics, namely: gender, cumulative grand point average (GPA) score, education level of parents, willingness to be teacher and appointment possibility as a teacher after graduation. The cumulative GPA score was used for academic achievement.

Motivated Strategies for Learning Questionnaire

The Motivated Strategies for Learning Questionnaire (MSLQ), constructed and validated by Pintrich et al. (1991) for use with college students, is a 7-point Likert scale anchored with 1 = not at all true for me and 7 = very true for me. This self-report questionnaire includes two main parts. The first part consists of 31 items assessing students' motivational orientations on six subscales. In this research following subscales of motivation were used: task value, control of learning beliefs, self-efficacy and test anxiety. The second part involves 50 items measuring students' learning strategies on nine subscales: rehearsal, organization, elaboration, critical thinking, metacognitive self-regulation, effort regulation, time and study environment, peer learning and help seeking. Turkish version of this questionnaire was validated by Büyüköztürk et al. (2004). In this research, reliabilities of the sub-dimensions ranged from .58 to .88 for the first part, and from .52 to .79 for the second part of the questionnaire. CFA results showed that the thirteen-dimension model of the instrument generally has a good fit [$\chi^2/df = 2.67$, RMSEA (90% CI) = .05, CFI = .96, SRMR = .05]. Apart from that, the correlation between the factors was ranged between .29 and .80 in the model. Considering the need for correlation between factors to be below .80, it can be stated that the result obtained is adequate.

Achievement Goal Questionnaire

Achievement Goal Questionnaire (AGQ) is a 15-item instrument measured on a 5-point Likert scale ranged between 1 = never and 5 = always (Elliot & McGregor, 2001). This self-report questionnaire has four subscales measuring students' goals for mastery approach, mastery avoidance, performance approach, and performance avoidance. Turkish version of this instrument was validated by Senler and Sungur-Vural (2013) for use with university students. In the present research, reliability of the sub-dimensions ranged between .72 and .83. CFA results revealed that fit indices for four-factor model were acceptable [$\chi^2/df = 4.51$, RMSEA (90% CI) = .07, CFI = .90, SRMR = .06]. In addition to this, the correlation between the factors was found between .41 and .79 in the model.



Constructivist Learning Environment Scale

Constructivist Learning Environment Scale (CLES) is a 20-item 5-point Likert scale anchored with 1 = almost never and 5 = almost always (Johnson & McClure, 2004). This instrument is used to measure constructivist learning environment perceptions on five subscales: personal relevance, uncertainty, critical voice, shared control and student negotiation. Haciomeroglu and Memnun (2013) validated Turkish version of the CLES for use with university students by supporting five-factor structure and obtaining Cronbach alpha coefficients ranging from .67 to .89. In the current research, it was found that the reliability coefficients ranged between .72 and .78 for sub-dimensions of CLES. Additionally, CFA results indicated that the Constructivist Learning Environment Scale fits the five-dimension measurement model [$\chi^2/df = 2.73$, RMSEA (90% CI) = .06, CFI = .95, SRMR = .05]. Besides, in the model the correlation between the factors was found between .32 and .77.

Data Analysis

Questionnaires of 15 preservice science teachers which were incorrectly and incompletely filled in were excluded and the remaining data from 516 participants were put to analysis. First, one-way extreme value analysis was performed on the basis of z scores, and 22 people were excluded from the analysis. Later, 14 participants were removed from the dataset after a multi-directional analysis which was performed according to Mahalanobis distance. As a result, following analyses were conducted with a group of 480 participants.

K-Means Clustering

Clustering is described as a process of segmenting the items in a dataset by considering their properties and the distances with a suitable metric such as Euclidean distance (Likas et al., 2003). The literature involves numerous studies in many fields such as computer vision, robot intelligence, anomaly detection that benefit from clustering methods so far. In essence, clustering methods attempt to group similar items while keeping the dissimilar ones away to form other groups. At this point, clustering methods help to explore the invisible patterns in the big picture when they are decomposed to consistent similar clusters.

Invented by MacQueen (1967), K-means clustering method, segments the dataset samples into k number of clusters by first picking k centroids - candidate cluster centers - in a random fashion. The algorithm then assigns the other examples to the nearest cluster by calculating the natural distance between the sample and candidate centroids. The literature of clustering includes many distance metrics like Euclidean, Manhattan, Minkowski, etc. The selection of the right distance metric is completely left to the user. Next, the distance is computed according to the qualities which the samples have. The K-means algorithm follows an iterative approach having two folds: (1) computing the distance of a sample to each centroid and assign it to the nearest cluster, (2) recomputing the centroids with newly assigned samples. In this way, the centroids are continuously updated by moving the newly assigned cases. This iterative approach lasts until the positions of centroids in high dimension space do not change anymore. This process is also called the convergence of the algorithm and controlled by a cost/objective function indicated as J . In principle, the objective function J computes the error squares in (1).

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x^{(i)} - c_j\|^2 \quad (1)$$

Euclidean distance as formulated with the term $\|x^{(i)} - c_j\|^2$ represents the distance function to calculate the dissimilarity between the sample x_i^j and the j th centroid c_j . The main goal of J is to minimize the total cost to reach a stable state. K-means was chosen as the clustering mechanism due to several reasons: (a) guarantee for convergence, (b) scalability, (c) ease of implementation.

Tan et al. (2006) reported that regardless of the employed method, the optimization functions used in clustering approaches aim to have sets having high intra-class similarities in addition to low inter-class similarities. This property affects the quality of the clustering scheme. As has been noted before, the K-means scheme takes the cluster count k as the hyper-parameter. However, at the initial stage, it is not very straightforward to answer the question of "What should be the optimal value of k ?" Thus, researchers have developed several ways (e.g., Dunn Index, Silhouette coefficient) to determine the optimal value of k since it affects the quality of the analysis.



Determining Optimum Cluster Count via Silhouette Coefficient

Though clustering is a significant technique that partitions data patterns into meaningful segments, it is not directly used for determining the number of clusters (Zhou & Gao, 2014). For instance, K -means is an iterative approach which its outcomes heavily depend on (1) initial centroid selection and (2) the k value that determines the number of clusters. Moreover, k value significantly affects cluster quality and distribution. For this reason, several approaches such as Davies-Boulding Index (Davies & Boulding, 1979), Dunn Index (Dunn, 1974) and Silhouette Coefficient (Rousseeuw, 1987) have been developed in order to determine the optimal cluster number for better validation. In essence, all these methods attempt to obtain well-separated clusters which involve small variances among the members of each cluster such that they exhibit large inter-class variance along with low inter-class variance.

The silhouette coefficient (i.e., score, index), suggested by Rousseeuw (1987), is a cluster validity measure that merges both the measures of cohesion and separation. The term *silhouette value* that ranges between -1 to 1 here refers to a degree that shows whether a sample lies in its correct cluster (cohesion) compared to other clusters (separation). According to Rousseeuw (1987), the silhouette represents "which objects lie well within their cluster, and which ones are merely somewhere in between clusters" (p. 57). Moreover, the score for a sample (i.e., case, object) getting closer to 1 indicates that the sample is better matched to its correct cluster whereas negative values show that the sample is dissimilar to its current cluster. The next step of the algorithm involves visualization of all silhouette scores in a single shot along with the obtained clusters. At this point, having a large portion of high individual scores indicates that the number of clusters was determined correctly. In contrast, obtaining low scores show that the cluster count is not optimal. Consequently, the silhouette coefficient eventually enables determining the optimal cluster count by making a trial-error study. The computation of the silhouette coefficient is done through the formulation given in (2). Given the intra-cluster distance a and mean nearest-cluster distance b , the silhouette score is calculated as follows:

$$\text{Silhouette Score} = \frac{(b-a)}{\max(a,b)} \quad (2)$$

It should be noted that, the term b represents the distance between a sample and the nearest cluster that the sample does not belong to. After having each computed all scores belonging to all samples, silhouette coefficient is computed by taking the mean of all scores. In this research, in order to compute the silhouette coefficient, the scikit-learn (Scikit-learn, 2020) package which is a well-known machine learning Python library was used.

Association Rules Mining and Generalized Rule Induction

Association rules mining (ARM) as another member of descriptive/unsupervised data mining aims to explore hidden association patterns embedded in situations, observations and transactions within a dataset. For instance, customers who purchase a computer and a keyboard are also likely to buy a mouse. As its name suggests, ARM deals with finding those kinds of relations that could be useful for decision-makers. Thus, the method of association rules mining has found widespread usage after its first prototypical employment on market-basket analysis. Similarly, ARM has been employed numerous times in literature covering many studies such as telecommunication, engineering, sales data (Agrawal et al., 1994).

By definition, an association rule is an expression $X \Rightarrow Y$, such that A and B represent sets of items and the rationale behind a rule is how likely X and Y occur together. Toivonen (1996) describes this phenomenon with a solid case by giving the example of $beer \rightarrow chips$ (87%) having the interpretation that the 87% of the customers who bought beer also got chips. Agrawal et al. (1994) have suggested the well-known algorithm so-called *Apriori* for the first time in order to be used for discovering association rules within huge datasets.

Given that $\mathcal{L} = \{l_1, l_2, l_3 \dots l_n\}$ represents the set of items and \mathcal{B} shows the set of transactions (i.e., dataset) where each transaction T becomes an item set satisfying the condition of $T \subseteq \mathcal{L}$ (Agrawal et al., 1994). At this point, $\mathcal{L} = \{l_1, l_2, l_3 \dots l_n\}$ represent the set of attributes meaning that they occur or do not occur. In other words, each attribute l_i is either mapped to 1 or 0 due to its occurrence in each transaction that is called as *TID*. Similarly, a set of items are also called as an item set (i.e., X) which holds the property of $X \subseteq \mathcal{L}$. It is assumed that a transaction T involves an item set X , if $X \subseteq T$ condition is met. In other words, an item set will be a sub-set of \mathcal{L} when it includes zero or more than zero elements. Agrawal et al. (1994) state that an as-



sociation rule can be seen as an implication of the form $X \Rightarrow Y$ that satisfies the condition of (a) $X \subset \mathcal{L}$, (b) $Y \subset \mathcal{L}$ and finally (c) $X \cap Y = \emptyset$. Note that, in this setting, X is called as the *antecedent* whereas Y denotes the *consequent* part of the rule. Here the consequent part of the rule occurs together with the antecedent along with some frequency and posterior probability scores. Formally, the rule extraction scheme works under the control of two important threshold parameters so-called *support* and *confidence*. Following these statements, the support and confidence values of an association rule in a dataset containing N transactions are given in (3) and (4) respectively. Note that, the antecedent part of the rule may contain more than one item while the consequent includes a single item only.

$$\text{Support, } s(X \Rightarrow Y) = \frac{\Psi(X \cup Y)}{N} \quad (3)$$

$$\text{Confidence, } c(X \Rightarrow Y) = \frac{\Psi(X \cup Y)}{\Psi(X)} \quad (4)$$

According to this configuration, the support value of a rule can be inferred as the total co-occurrence frequency of both antecedent and consequent in the dataset. In contrast, the confidence score is computed by dividing the co-occurrence of the antecedent and consequent by the occurrence of the antecedent. While the Apriori algorithm considers both of these scores while thresholding the important rules, Aran et al. (2019) suggested the selection of confidence score rather than support score since the computation of confidence is taking posterior probability into account. However, the calculation of the support score is solely based on the frequency of co-occurrence of antecedent and consequent.

It is important to know that Apriori algorithm takes only discrete variables (i.e., multivariate variables) as input. Therefore, it is a shortcoming of the algorithm when continuous variables come into prominence. In order to overcome, an enhanced version of the Apriori algorithm named Generalized Rule Induction (GRI) that is shipped with SPSS Clementine 12 data mining software was employed. Though it is a variant of the Apriori algorithm, the GRI can handle both continuous and discrete variables. Moreover, it extracts the rules containing continuous/discrete variables as the antecedent while the consequent could be only of discrete attributes. The aim of using GRI algorithm was to discover the hidden relations among the attributes of this dataset.

Description of Data Analysis Process Used in This Research

K-means clustering analysis was run through the data mining application named "Weka". The Weka workbench application (Weka, 2020) is an open-source software and widely employed for various purposes such as teaching, research, and industrial applications. The rationale behind the selection of Weka is that it outputs detailed clustering results and serves ease of use through its graphical user interface. Prior to K-means analysis, the best hyper parameter of k was selected to perform optimal clustering as stated before. Therefore, searching for the best k has been performed via the Silhouette coefficient score. To achieve this procedure, a Python 3.7 script was created employing Scikit learn package. The silhouette coefficient (SC) score was measured for each cluster count ranging from 2 to 5. The results were visualized with "Matplotlib".



Figure 1

Silhouette coefficients for various cluster counts ranging from 2 to 5 (Best viewed in color)

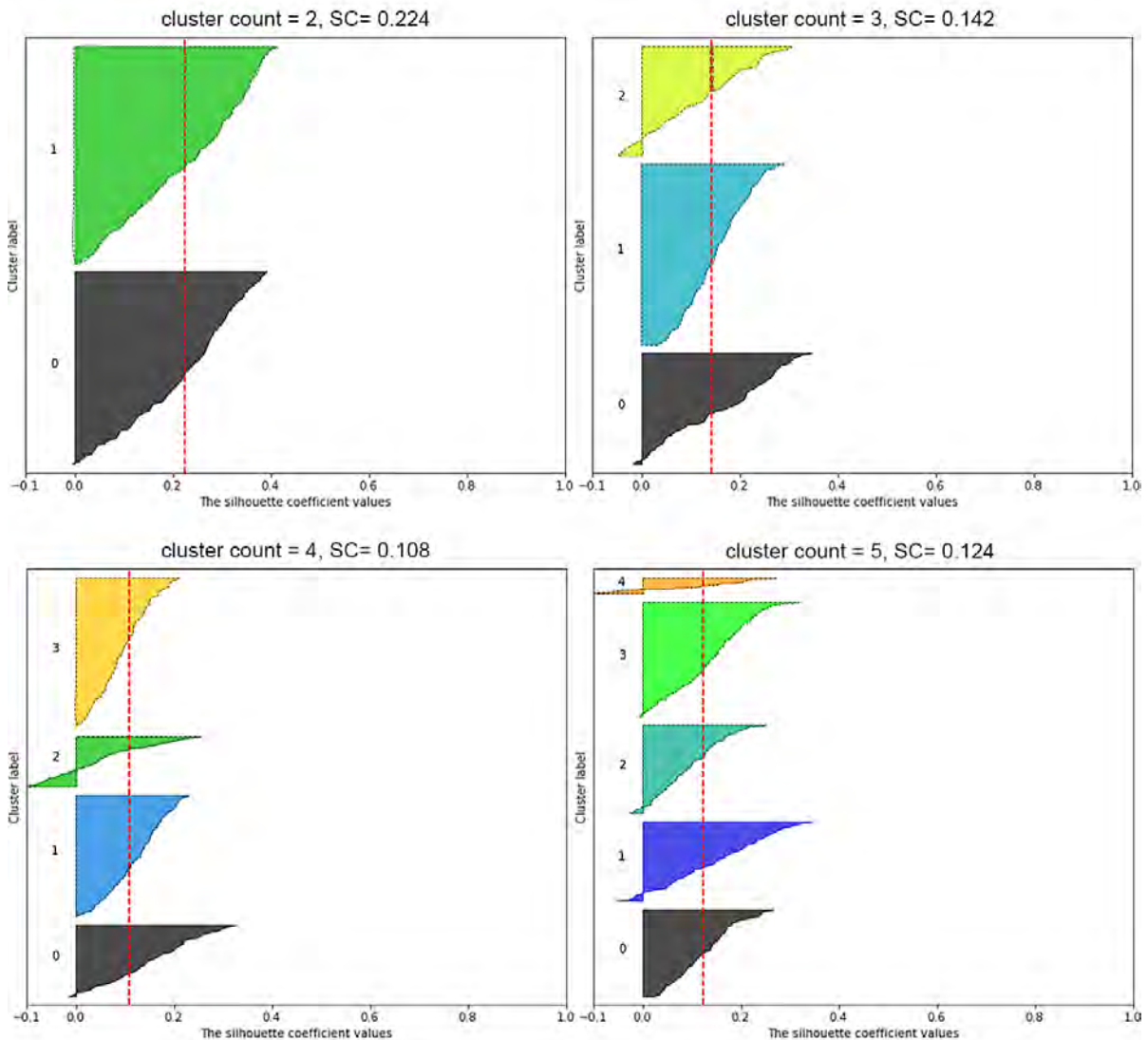


Figure 1 clearly shows that the best cluster assignment could be made by setting the $k = 2$ since the best score of .224 has been computed in this setting. Moreover, the figure reveals that some data points (i.e., subjects) have been assigned to wrong clusters if k is larger than 2. As can be seen from Figure 1, individual silhouette scores of some data points have been detected as less than 0. On the other hand, the configuration with $k = 2$ shows no incorrect or “fuzzy” assignment. After having detected the best k value, the K-means algorithm was run in Weka according to this setting. The findings were then obtained and analyzed.

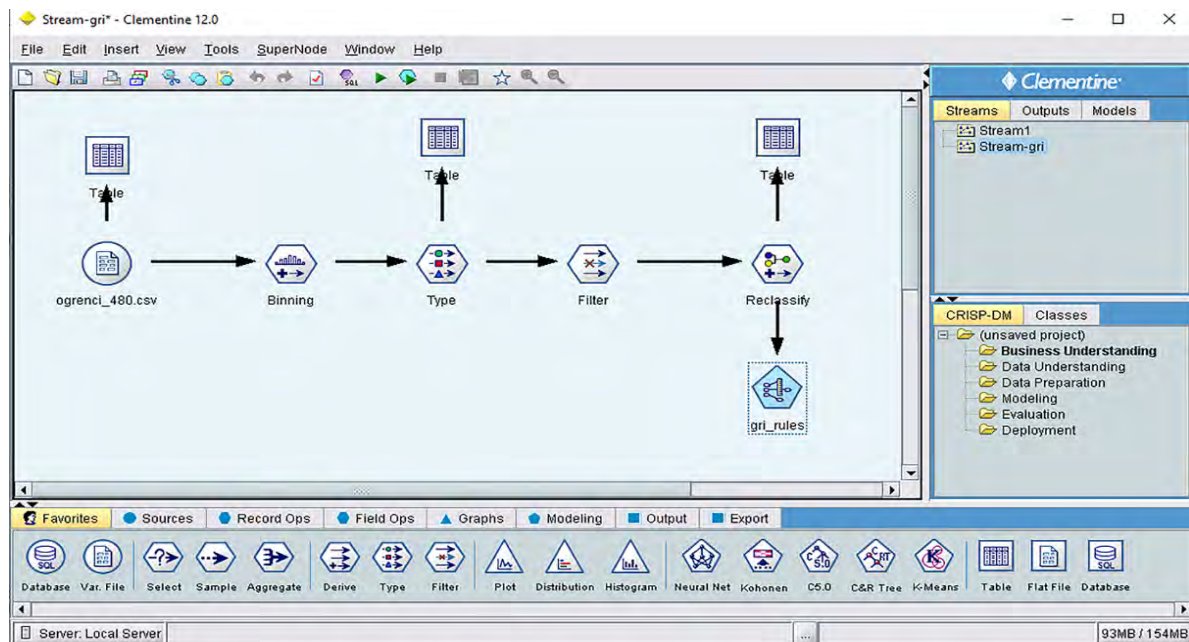
In the next phase, the association rules mining analysis was carried out through the GRI algorithm which is shipped with SPSS Clementine software. SPSS Clementine software package is a licensed visual data mining application developed by SPSS Company. In Clementine software, there exist several task-specific units so-called the *nodes*. Since the aim was to reveal the hidden associations among preservice science teachers' motivational beliefs, learning strategies, and constructivist learning environment perceptions; the related variables were mapped to *both* to ensure that all these variables could occur in both antecedent and consequent part of the extracted rules. Next, the GRI node was picked from the model toolbox and connected to the end of the workflow which is depicted in Figure 2. Note how each task-specific node transforms the data and transmits to its successor. The



minimum support value was chosen to be 30% and minimum confidence as 70% to reveal strong rules. Meanwhile, it took approximately 49 seconds of Clementine to list the results. The obtained GRI rules were later analyzed, and important findings were presented in the next section.

Figure 2

The data preprocessing and modeling workflow of GRI based association rules mining phase



Research Results

Findings of Clustering Analysis

The K-means based clustering analysis revealed two natural groups discriminating preservice science teachers on the variables used in this research. Cluster 1 was composed of just females ($n = 319$), while Cluster 2 was composed of just males ($n = 161$); which meant that gender was a discriminating factor between the clusters. This finding is surprising in that it reveals differences between gender and other variables of interest. The properties of these clusters were depicted in Table 2 along with obtained mean and standard deviations.

There were significant differences between the two clusters (i.e., males and females) in terms of preservice science teachers' GPA scores, father education level, willingness to be teacher, learning strategies and goal orientation. Female preservice science teachers' achievement was higher than that of males ($t = 7.39, p < .01$) with a large effect size (Cohen's $d = .80$). In terms of parent's education level, the two clusters differed with respect to the father education level ($t = 2.62, p < .01$, Cohen's $d = .37$); but did not differ with respect to mother education level. Female preservice science teachers were more willing than males to be a teacher ($t = 3.39, p < .01$, Cohen's $d = .37$). However, there was not any difference between the groups with regard to appointment possibility as a teacher after graduation.



Table 2
Clustering results for preservice science teachers

Characteristics	Cluster 1 M (SD)	Cluster 2 M (SD)	Characteristics	Cluster 1 M (SD)	Cluster 2 M (SD)
Background Charac- teristics			Learning strategies		
GPA	2.77 (.57)	2.24 (.82)	Organization	21.93 (4.26)	19.94 (4.90)
Mother education	1.62 (1.11)	1.45 (1.12)	Elaboration	31.59 (5.46)	29.23 (6.95)
Father education	2.49 (1.21)	2.17 (1.34)	Rehearsal	20.15 (4.32)	18.50 (4.71)
Willingness to be teacher	.91 (.28)	.79 (.40)	Critical thinking	23.47 (5.31)	23.48 (5.69)
Appointment possibility	2.17 (1.31)	2.05 (1.53)	Effort regulation	19.37 (4.64)	17.20 (4.69)
Goal orientation			Metacognitive self- regulation	60.56 (9.71)	56.24 (11.24)
Mastery approach goal	12.50 (2.16)	11.69 (2.60)	Help seeking	17.56 (4.42)	17.03 (4.73)
Mastery avoidance goal	8.95 (2.75)	8.47 (2.97)	Peer learning	12.24 (3.88)	12.42 (3.91)
Performance approach goal	9.87 (3.24)	8.98 (3.50)	Time and study environ- ment	38.25 (7.62)	34.96 (8.44)
Performance avoidance goal	16.71 (5.68)	15.93 (5.95)	<i>Perceptions of CLES</i>		
<i>Motivational beliefs</i>			Personal relevance	13.50 (2.88)	13.20 (3.12)
Task value	31.47 (5.58)	30.30 (6.22)	Uncertainty	13.35 (2.54)	13.03 (2.40)
Control of learning beliefs	20.55 (4.03)	21.31 (4.24)	Critical voice	13.32 (2.89)	13.02 (3.00)
Self-efficacy	41.51 (8.13)	42.02 (8.50)	Shared control	7.52 (3.29)	8.14 (3.40)
Test anxiety	18.31 (6.57)	18.13 (6.51)	Student negotiation	11.79 (2.95)	11.96 (2.96)

Another significant difference detected across two clusters is associated with the learning strategies. Significant differences were found in perceived use of organization ($t = 4.39, p < .01$, Cohen's $d = .44$), elaboration ($t = 3.77, p < .01$, Cohen's $d = .39$), rehearsal ($t = 3.73, p < .01$, Cohen's $d = .37$), effort regulation ($t = 4.80, p < .01$, Cohen's $d = .47$), metacognitive self-regulation ($t = 4.15, p < .01$, Cohen's $d = .42$) and time and study environment ($t = 4.16, p < .01$, Cohen's $d = .42$) strategies in the favor of females, indicating medium effect size. However, perceived use of critical thinking, help seeking, and peer learning strategies were not different across clusters.

Two clusters were also different according to the preservice science teachers' goal orientation. Females held higher approach goals than males; the difference in mastery approach goal orientation was medium in size ($t = 3.42, p < .01$, Cohen's $d = .35$), while the difference in performance approach goal orientation was small in size ($t = 2.71, p < .01$, Cohen's $d = .27$). However, there was not any difference between the two clusters (i.e., males and females) with respect to avoidance goals, other motivational beliefs and constructivist learning environment perceptions.



Findings of Association Rules Mining Analysis

For this dataset, association rules having minimum support of 30% and minimum confidence of 70% were selected as strong rules. Thus, a total of 300 significant association rules were successfully extracted. Note that, thresholding the GRI algorithm via minimum support and confidence parameter causes to decrease the number of observed associations. However, this filters out the insignificant rules and it is also a good point to consider the attributes that cannot be present logically in the antecedent or consequent of the discovered rules. In this dataset, rules including subscales of the same scale in both consequent and antecedent were excluded due to dimensions of a construct share common variance with each other, and thereby it would not be logical to search for the dependency between them. Moreover, some of the discovered rules have identical meanings, are redundant or having random relations. A rule is considered redundant when it adds no information over another rule. For example, a combination of two or more rules does not cover any extra information in the presence of each rule. After eliminating redundant or uninteresting rules, a selection of high-support and high-confidence rules relevant to the research were used in decision making for brevity.

For clarity, findings concerning selected association rules were presented separately by grouping rules with common consequents and then combining those groups based on their relatedness.

Findings Concerning Motivational Beliefs

Table 3 shows a list of rules showing the relations between motivational beliefs and other variables along with the support and the confidence values they have. Motivational beliefs appeared as the consequent in the association rules were mastery approach goal orientation, control of learning beliefs and task value. Mastery approach goal orientation took place in the rules having the strongest consequent due to the highest confidence values (Rule 1-7).

Among the goal orientation variables, the GRI algorithm extracted just mastery approach goal orientation in the rules take place both in antecedent and consequent parts. Other goal orientation variables were not observed in the detected rules. Table 3 shows a list of rules showing the associations between mastery approach goal and other variables along with support and confidence parameters. It is seen that gender was the dominant of antecedent variables, that is 4 of the 7 the rules included gender. Other antecedents were willingness to be teacher, organization, elaboration, personal relevance, uncertainty and GPA.

Table 3 points out that female preservice science teachers held higher mastery approach goals than their male counterparts (Rules 3-5 and 7). Female preservice science teachers, who reported willingness to be teacher, using organization and elaboration strategies and viewing scientific knowledge as evolving, inclined to possess higher levels of mastery approach goal (Rules 3-5 and 7). Also, preservice science teachers who found their science classes linked to daily lives reported higher mastery approach goals. Personal relevance was found as a predictor of mastery approach goal both alone and jointly with the willingness to be teacher (Rules 1 and 2). Furthermore, academic achievement was found directly proportional to the mastery approach goal orientation (Rule 6). That is, preservice science teachers having higher academic achievement demonstrated higher mastery approach goal orientation.

Table 3

Association rules including mastery approach goal orientation (MAG), control of learning beliefs (CLB) and task value (TV) as a consequent

	Antecedent	Consequent	Support %	Confidence%
1	Willingness to be teacher = Yes and Personal relevance = High	MAG = High	33.54	96.27
2	Personal relevance = High	MAG = High	36.25	95.4
3	Gender = Female and Elaboration = High	MAG = High	44.79	90.23
4	Gender = Female and Uncertainty = High	MAG = High	32.08	89.61
5	Gender = Female and Organization = High	MAG = High	48.75	88.03
6	GPA > 2,90	MAG = High	36.88	87.01



	Antecedent	Consequent	Support %	Confidence%
7	Gender = Female and Willingness to be teacher = Yes	MAG = High	60.83	86.3
8	Rehearsal = High and Metacognitive self-regulation = High	CLB = High	32.5	76.92
9	Willingness to be teacher = Yes and Metacognitive self-regulation = High	CLB = High	39.79	74.87
10	Organization = High and Metacognitive self-regulation = High	CLB = High	40.21	74.09
11	Metacognitive self-regulation = High	CLB = High	43.33	74.04
12	Elaboration = High and Metacognitive self-regulation = High	CLB = High	39.17	73.94
13	Willingness to be teacher = Yes and Personal relevance = High	CLB = High	33.54	73.91
14	Organization = High and Rehearsal = High	CLB = High	42.92	73.3
15	Personal relevance = High	CLB = High	36.25	72.99
16	Gender = Female and Metacognitive self-regulation = High	CLB = High	31.87	70.59
17	Shared control = Low and Critical thinking = Medium and Metacognitive self-regulation = Medium	TV = Medium	30	70.83
18	Personal relevance = Medium and Metacognitive self-regulation = Medium	TV = Medium	31.87	70.59
19	Willingness to be teacher = Yes and Organization = High and Critical thinking = High	TV = High	36.04	70.52
20	Elaboration = High and Critical thinking = High	TV = High	37.29	70.39

Association rules in Table 3 indicate that control of learning beliefs was related to gender, willingness to be teacher, perceived constructivist learning environment and strategy use. Majority of the rules included at least one learning strategy as an antecedent (Rules 8-12, 14 and 16). The dominant learning strategy that appeared in the rules was metacognitive self-regulation. Preservice science teachers having higher metacognitive self-regulation both alone and joint with either high levels of organization or rehearsal or elaboration reported higher levels of control of learning beliefs (Rules 8, 10-12). The participants who were willing to be teacher and having higher metacognitive self-regulation or perception of personal relevance held higher control of learning beliefs (Rules 9 and 13). Preservice science teachers perceiving learning task related to everyday life believed that their learning achievement depends on their effort (Rule 15). Moreover, female preservice science teachers who were using metacognitive self-regulatory skills frequently demonstrated higher control over their learning (Rule 16).

Task value was another motivational belief extracted as a consequent in association rules shown in Table 3. Willingness to be teacher, learning strategies and perceptions of constructivist learning environment appeared as factors associated with task value. Each rule included at least one learning strategy as an antecedent. Among the learning strategies, critical thinking seemed to be dominant. Preservice science teachers having a medium level of critical thinking and metacognitive self-regulation and lower perceptions of shared control held task value at a moderate level (Rule 17). Rule 18 also indicated a moderate level of task value if the learning environment was moderately perceived as connected to daily life with a moderate level of metacognitive self-regulation. However, preservice science teachers gave higher importance to learning task if they were willing to be teacher and having higher critical thinking and organization strategies (Rule 19). A combination of high levels of critical thinking and elaboration strategies was also found related to high levels of task value (Rule 20).

Findings Concerning Learning Strategies

A list of rules showing the relations between learning strategies and other variables with the support and the confidence values were given in Table 4. Learning strategies that appeared in the consequent part of the association rules were organization, elaboration, metacognitive self-regulation, critical thinking, and time and study environment.

Table 4 depicts the critical voice as a significant factor associated with organization strategy. A high proportion (75%) of the rules with relatively high reliability contained critical voice. This suggests a strong association



between critical voice and organization strategy. Preservice science teachers expressing critical views in their science courses inclined to use organization strategies. Critical voice predicted organization strategy both alone and joint with mastery approach goal and willingness to be teacher (Rules 1-3). That is, preservice science teachers who were either having higher mastery approach goals or willing to be teacher with higher perceptions of critical voice reported frequent use of organization strategies. In addition, GPA emerged as a factor related with organization (Rule 4). High-achiever preservice science teachers tended to use more organization strategies.

Another learning strategy that appeared in the consequent part of the rules was elaboration. Associated factors with elaboration were GPA, gender, willingness to be teacher, self-efficacy, mastery approach goal, critical voice, uncertainty, student negotiation and personal relevance. Interestingly, out of 8 rules, 7 of them included perception of constructivist learning environment variable. That means, preservice science teachers who perceived their science learning environment as constructivist reported higher levels of elaboration strategies. Critical voice and personal relevance were related to elaboration both alone and joint with the willingness to be teacher (Rules 6, 8, 9 and 11). Moreover, preservice science teachers both having higher mastery approach goals and relating learning task with daily life tended to use elaboration strategies frequently (Rule 7). Female preservice science teachers having high levels of mastery approach goals and moderately sharing their ideas in learning science also reported higher levels of elaboration (Rule 10). Female preservice science teachers viewing scientific knowledge as evolving also held higher elaboration scores (Rule 12). Similar to the organization, higher scores on elaboration were found related to higher scores on academic achievement. Rule 5 indicates that high achieving preservice science teachers who reported both willingness to be teacher and high self-efficacy inclined to use elaboration strategies frequently.

Table 4

Association rules including organization (O), elaboration (E), metacognitive self-regulation (MSR), critical thinking (CT) and time and study environment (TSE) as a consequent

	Antecedent	Consequent	Support %	Confidence %
1	Critical voice = High and Mastery approach goal orientation = High	O = High	32.92	86.08
2	Willingness to be teacher = Yes and Critical voice = High	O = High	33.54	83.85
3	Critical voice = High	O = High	36.25	82.76
4	GPA > 2,95	O = High	34.79	82.63
5	GPA > 2,63 and Willingness to be teacher = Yes and Self-efficacy = High	E = High	37.5	84.44
6	Willingness to be teacher = Yes and Critical voice = High	E = High	33.54	80.75
7	Personal relevance = High and Mastery approach goal orientation = High	E = High	34.58	78.92
8	Critical voice = High	E = High	36.25	78.74
9	Willingness to be teacher = Yes and Personal relevance = High	E = High	33.54	78.26
10	Gender = Female and Student negotiation = Medium and Mastery approach goal orientation = High	E = High	30.83	77.7
11	Personal relevance = High	E = High	36.25	77.59
12	Gender = Female and Uncertainty = High	E = High	32.08	75.32
13	Personal relevance = Medium and Task value = Medium	MSR = Medium	31.25	72.00
14	Task value = High and Control of learning beliefs = High and Mastery approach goal orientation = High	MSR = High	30	70.83
15	Willingness to be teacher = Yes and Task value = High and Self-efficacy = High	CT = High	35.21	71.01
16	Personal relevance = Medium and Critical voice = Medium	TSE = Medium	32.92	70.89
17	Personal relevance = Medium and Uncertainty = Medium	TSE = Medium	32.08	70.13



Unlike organization and elaboration, metacognitive self-regulation, time and study environment management and critical thinking were less frequently obtained in extracted association rules as a consequent. Preservice science teachers having medium level of task value and perception of personal relevance regarding their science learning environment tended to have a medium level of metacognitive self-regulation (Rule 13). If they held high levels of mastery approach goal, task value and control beliefs for learning; then, they had a high level of metacognitive self-regulation (Rule 14). Rules 13 and 14 depicted that task value was the dominant factor associated with metacognitive self-regulation. Task value was also associated with critical thinking. Preservice science teachers who were willing to be teacher and having higher self-efficacy and task value developed high levels of critical thinking skills (Rule 15). Personal relevance, critical voice and uncertainty were found as determinants of time and study environment strategy. Preservice science teachers having a moderate level of perception of personal relevance with either moderate level of perception of critical voice or uncertainty tended to manage time and study environment moderately (Rule 16 and 17).

Discussion

In this research, data mining methods (i.e., clustering and association rules mining) were used in discovering the characteristics of preservice science teachers and hidden relations among those characteristics. Preservice science teachers were classified into two groups as a result of the clustering analysis. One of the natural groups was composed of males, while the other one was composed of females. This finding depicts the importance of gender as a variable that discriminated the characteristics of preservice science teachers.

Gender Difference in Preservice Science Teachers' Characteristics

Results revealed academic achievement as a characteristic of preservice science teachers differing males and females significantly with a large effect size. This finding supports the previous research results in differences in academic achievement across gender frequently favoring female students (Voyer & Voyer, 2014). This research also demonstrated that females were more eager to be teachers than males. This finding might be related to the reasons for preservice science teachers in choosing teaching profession. Balyer and Özcan (2014) revealed that female student teachers chose it for intrinsic reasons, while male counterparts chose it for extrinsic reasons. Regarding background characteristics, the results demonstrated that majority of the preservice science teachers came from lower-educated families. This finding suggests that students coming from highly educated families are less interested in joining the teaching profession (Balyer & Özcan, 2014; Lai et al., 2005). It should be noted that males and females were significantly different with respect to their fathers' level of education but not to mothers' level of education. Besides, preservice science teachers' views about their appointment possibility were found equal across their gender. This finding is expected because opportunities for appointment after graduation are equal for both genders. Teacher appointments in Turkey have been made with a nation-wide exam for a long time.

Gender differences were also obtained in preservice science teachers' mastery and performance approach goal orientations, in favor of females. However, males and females were not significantly different with regard to avoidance goals, task value, self-efficacy and control of learning beliefs. This result is congruent with the existing literature demonstrating mixed results concerning the relation between motivational beliefs and gender. Previous studies indicating significant differences in motivational beliefs were generally in the favor of females. For example, girls held high levels of intrinsic (mastery) goal orientation, task value and control of learning beliefs than boys in the research of Arisoy et al. (2016). In another research conducted by Yerdelen and Sungur (2019), females held higher approach goals than males. Britner and Pajares (2006) reported significant difference in self-efficacy across gender. There were also some studies indicating non-significant associations between motivational beliefs and gender. Females and males were not significantly different regarding task value beliefs (Kahraman & Sungur-Vural, 2014) and self-efficacy in science (Kiran & Sungur, 2012; Sezginürk & Sungur, 2020).

The findings also showed significant gender differences in preservice science teachers' learning strategy use. Female preservice science teachers tended to use rehearsal, elaboration, organization, metacognitive self-regulation, effort regulation and time and study environment management strategies more than their male counterparts. However, preservice science teachers were not significantly different in terms of using critical thinking, help seeking and peer learning strategies. This finding exactly confirms the findings of Bidjerano (2005). Pajares (2002) also reported significant differences across gender in learning strategy use, while others did not (Kiran & Sungur, 2012).



The results concerning the relation of gender with perceived constructivist learning environment demonstrated no significant differences in all subscales of CLES. Females and males perceived their learning environment as constructivists equally. This finding is aligned with the previous literature revealing similar learning environment perceptions across gender (LaRocque, 2008). However, the present research is not parallel with the existing literature which determined differences in perceived constructivist learning environment, across gender, demonstrating an advantage for female students (Arisoy et al., 2016).

Antecedents of Preservice Science Teachers' Motivational Beliefs and Strategy Use

This research further extracted the antecedents of preservice science teachers' motivational orientations and strategy use in the form of association rules. Motivational beliefs obtained as a consequent in association rules mining analysis were mastery approach goal orientation, task value and control of learning beliefs. Learning strategies which appeared as a consequent in the association rules were organization, elaboration, metacognitive self-regulation, critical thinking, and time and study environment. The consequents extracted as a result of association rules mining demonstrate the motivational beliefs frequently held by the preservice science teachers and the learning strategies perceived to be used more often.

The antecedents of mastery approach goal orientation included gender, willingness to be teacher, organization, elaboration, personal relevance, uncertainty and academic achievement. Gender was found as the dominant of the antecedents related to mastery approach goal orientation. This finding is aligned to the clustering analysis which determined a significant difference in mastery approach goal orientation across gender. This result is also aligned to the literature indicating gender difference in mastery approach goals in favor of females (e.g., Yerdelen & Sungur, 2019). Furthermore, GRI results revealed the combination of gender and willingness to be teacher as antecedents of mastery approach goal orientation. This rule is aligned with the findings of the clustering analysis demonstrating the relation between gender and willingness to be teacher. Previous studies also support the combination of gender and willingness to be teacher by demonstrating that female preservice teachers choose profession of teaching for intrinsic reasons, while males choose it for extrinsic reasons (Balyer & Özcan, 2014). Thus, the present research showed that female preservice science teachers who were eager to be a teacher held higher mastery approach goals. This research is also congruent with the literature depicting positive associations between achievement and goal orientation (Sungur & Güngören, 2009). The present research further showed that preservice science teachers who use organization and elaboration strategies engaged in activities to develop their knowledge and skills. Existing studies confirm this finding by demonstrating positive associations between learning strategy use and goal orientation (Sungur & Güngören). This research is also consistent with related literature determined that perceptions of constructivist classroom environment influence the adoption of students' mastery approach goals (Kingir et al., 2013; Yerdelen & Sungur, 2019). Classroom contexts that allow individuals to link subject matter to real-world situations and to view scientific knowledge as evolving would probably support development of mastery approach goals. Iverach and Fisher (2008) also reported personal relevance as a predictor of mastery approach goal orientation.

Common antecedents related to both task value and control of learning beliefs were willingness to be teacher, constructivist learning environment perceptions and strategy use. Gender appeared as a factor associated with control of learning belief rather than task value. However, in the clustering analysis both motivational beliefs were not found different across gender significantly. In association rules mining results concerning control of learning beliefs, gender was included in just one rule jointly with metacognitive self-regulation. Based on the clustering analysis, males and females were significantly different with respect to metacognitive self-regulation. Therefore, the combination of gender and metacognitive self-regulation is an expected antecedent. Besides, a common antecedent, willingness to be teacher, is a motivational orientation; therefore, its relation with other motivational constructs, which are control of beliefs and task value, is expected. Willingness to be teacher is included in the rules in combination with either learning strategies or constructivist learning environment perceptions. That is, preservice science teachers who were eager to be a teacher and reported use of learning strategies were likely to have higher beliefs on control of learning and perceive science tasks as important and useful. The participants who were eager to be a teacher and perceived learning environment as constructivist also held higher control of learning beliefs.

Learning strategies, constructivist learning environment perceptions and their combinations emerged as the factors associated with both task value and control of learning beliefs. This finding supports the related literature demonstrating the associations of learning strategy use and perceptions of classroom environment with



motivational orientations. For instance, in a research conducted by Arisoy et al. (2016), perceived constructivist learning environment was found associated with task value and control of learning beliefs. Interestingly, this research showed that among the strategies, metacognitive self-regulation appeared as a dominant strategy associated with the control of learning beliefs, while critical thinking seemed to be dominant in the rules concerning task value. Since metacognitive self-regulation includes control of cognition, its relation with control of learning beliefs is anticipated (Pintrich et al., 1991). As individuals engage in planning activities and monitor and regulate their learning, they tend to have control beliefs for learning. In a similar vein, the more preservice science teachers transfer their learning into different contexts and make critical analysis to reach decisions, the more they evaluate interest, importance and utility aspects of course material. Additionally, personal relevance was found related to task value and control of learning beliefs. Interestingly, shared control was emerged just related with task value. Preservice science teachers perceiving that they have limited shared control in their classes held moderate levels of task value beliefs.

Furthermore, the current research highlighted association rules concerning use of learning strategies. The results mainly revealed constructivist learning environment perceptions as antecedent of strategy use. Dethlefs (2002) also demonstrated positive associations of personal relevance, shared control and student negotiation with learning strategies. Other antecedents of learning strategy use included motivational beliefs, gender, academic achievement and willingness to be teacher. These antecedents are included in the rules either alone or in combinations among each other. Relevant research demonstrated that motivational beliefs were positively related to learning strategy use (Yumusak et al., 2007). The critical voice appeared as the dominant antecedent of the use of organization strategy. It was found that preservice science teachers who were high-achievers, eager to be teacher and pursuing mastery approach goals were tended to use organization strategies like outlining. Important characteristics of preservice science teachers reporting use of elaboration strategy emerged as their perceptions of constructivist learning environment except shared control. This implied that preservice science teachers who found their learning environment related to everyday life, expressed their views freely, viewed scientific knowledge as evolving, and interacted with their instructor and peers were likely to use elaboration strategy. Other antecedents of elaboration strategy use appearing in combinations among each other were gender, academic achievement, willingness to be teacher, self-efficacy and mastery approach goal orientation. Fadlelmula-Kayan et al. (2015) also demonstrated mastery approach goal orientation as an antecedent of elaboration and organization strategy use.

This research also extracted the antecedents of metacognitive self-regulation, critical thinking and time and study environment with relatively lower reliability than that of organization and elaboration strategies. Preservice science teachers having higher motivational beliefs and constructivist learning environment perceptions appeared to use metacognitive strategies. Sungur (2007) supports this finding by revealing intrinsic goal orientation, task value and control of learning beliefs as predictors of metacognitive strategy use. This finding is also aligned to the previous literature demonstrating positive associations between task value and strategy use (Kahraman & Sungur, 2013; Yumusak et al., 2007), and between mastery approach goal orientation and metacognitive self-regulation (Fadlelmula-Kayan et al., 2015; Kahraman & Sungur, 2011). Additionally, personal relevance perceived in science classes was found related to metacognitive self-regulation in this research. This result is consistent with the existing literature indicating perceived classroom environment as related to higher student self-regulation (Kingir et al., 2013; Yerdelen & Sungur, 2019). Moreover, motivational orientations (i.e., willingness to be teacher, self-efficacy and task value) appeared as antecedent of critical thinking. This finding is congruent with the studies considering motivation as a process activating and maintaining thinking critically to solve problems and make decisions (Valenzuela et al., 2011). Lastly, personal relevance, critical voice and uncertainty were found as the antecedents of time and study environment. This finding implied that preservice science teachers perceiving their classroom as constructivist appeared to put much effort into managing their time and study environment effectively. This claim is generally supported by the idea that learning environment perceptions are associated with strategy use (Sungur & Güngören, 2009).

Conclusions

This research adds to the relevant literature demonstrating both natural grouping of preservice science teachers' characteristics into two clusters based on gender differences and antecedents of motivational beliefs and strategy use. It is recommended that educators should be aware of the existence of a female advantage in preservice science teacher education. In addition, teacher educators and researchers in science education can obtain guideline



from the association rules captured in the current research. They are highly suggested to take into account the results when creating learning environment or conducting research. Preservice science teachers can be engaged in a variety of learner-centered activities including inquiry-based investigations and collaborative group that allow learner autonomy, control and negotiation. In addition, courses in teacher education programs can be enhanced or new courses can be offered to develop preservice science teachers' classroom environment perceptions as constructivist, adaptive motivational beliefs and effective strategy use. Since preservice science teachers are the teachers of the future, their characteristics influence the way they are going to be teaching twenty-first century students. The attainment of educational goals is interrelated with quality of teachers. Therefore, discovering profiles of preservice science teachers who are about to graduate will provide useful information to make a decision regarding the status of preservice science teacher education and how to enhance the quality of teacher education programs and thereby science teachers.

Importance of this research mainly comes from using clustering analysis and association rules mining techniques. Data mining is a promising research area in education, in that it extracts useful information without any hypothesis in advance. Despite its contributions to the relevant literature, this research has some limitations and recommendations for further research. One limitation is about the dependence of data on self-report measures, e.g., preservice science teachers' perceived and actual strategy use might be different. Another limitation is related to the convenience sampling procedure. This research is also limited to preservice science teachers enrolled in a public university. Moreover, this research is interdisciplinary in that data mining techniques coming from statistics and machine learning were used to analyze data collected from educational settings. This research would be motivating for future researchers to broaden the scope of the data mining research in science education area. More research might be conducted with larger samples and different age groups, variables, and academic disciplines.

References

- Agrawal, R., Manilla, H., Srikant, R., Toivonen, H., & Verkamo, A. I. (1996). Fast discovery of association rules. In U. Fayyad, G. Piatesky-Shapiro, P. Smyth, & R. Uthurusamy (Eds.), *Advances in knowledge discovery and data mining* (pp. 307-328). AAAI Press.
- Aldowah, H., Al-Samarraie, H., & Fauzy, W. M. (2019). Educational data mining and learning analytics for 21st century higher education: A review and synthesis. *Telematics and Informatics*, 37, 13-49. <https://doi.org/10.1016/j.tele.2019.01.007>
- Aran, O. C., Bozkir, A. S., Gok, B., & Yagci, E. (2019). Analyzing the views of teachers and prospective teachers on information and communication technology via descriptive data mining. *International Journal of Assessment Tools in Education*, 6(2), 314-329. <https://doi.org/10.21449/ijate.537877>
- Arisoy, N., Cakiroglu, J., Sungur, S., & Telli, S. (2016). Analysing middle school students' perceptions of their science classroom in relation to attitudes and motivation. In R. Taconis, P. den Brok, & A. Pilot (Eds.), *Teachers creating context-based learning environments in science* (pp. 173-190). Sense Publishers.
- Baldi, P. (2012). Autoencoders, unsupervised learning, and deep architectures. *Proceedings of Machine Learning Research*, 27, 37-50. <http://proceedings.mlr.press/v27/baldi12a/baldi12a.pdf>
- Balyer, A., & Özcan, K. (2014). Choosing teaching profession as a career: Students' reasons. *International Education Studies*, 7(5), 104-115. <http://dx.doi.org/10.5539/ies.v7n5p104>
- Bandura, A. (1994). Self-efficacy. In V. S. Ramachaudran (Ed.), *Encyclopedia of human behavior* (pp. 71-81). Academic Press.
- Bidjerano, T. (2005, October). *Gender differences in self-regulated learning*. Paper presented at the Annual Conference of the Northeastern Educational Research Association, Kerhonkson, NY, USA. <https://files.eric.ed.gov/fulltext/ED490777.pdf>
- Britner, S. L., & Pajares, F. (2006). Sources of science self-efficacy beliefs of middle school students. *Journal of Research in Science Teaching*, 43(5), 485-499. <https://doi.org/10.1002/tea.20131>
- Büyükoztürk, Ş., Akgün, Ö. E., Özkahveci, Ö., & Demirel, F. (2004). The validity and reliability study of the Turkish version of the motivated strategies for learning questionnaire. *Educational Sciences: Theory & Practice*, 4(2), 231-239.
- Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2), 224-227.
- Dethlefs, T. M. (2002). Relationship of constructivist learning environment to student attitudes and achievement in high school mathematics and science. *Dissertation Abstracts International*, 63(7), 24-55.
- Dogan, B., & Camurcu, A. Y. (2008). Association rule mining from an intelligent tutor. *Journal of Educational Technology Systems*, 36(4), 433-447. <https://doi.org/10.2190/ET.36.4.f>
- Dunn, J. C. (1974). Well-separated clusters and optimal fuzzy partitions. *Journal of Cybernetics*, 4(1), 95-104. <https://doi.org/10.1080/01969727408546059>
- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53, 109-132. <https://doi.org/10.1146/annurev.psych.53.100901.135153>
- Elliot, A. J., & McGregor, H. A. (2001). A 2x2 achievement goal framework. *Journal of Personality and Social Psychology*, 80, 501-519. <https://doi.org/10.1037/0022-3514.80.3.501>
- Elliot, A. J., & Church, M. A. (1997). A hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology*, 72, 218-232. <https://doi.org/10.1037/0022-3514.72.1.218>



- Fadlelmula-Kayan, F., Cakiroglu, E., & Sungur, S. (2015). Developing a structural model on the relationship among motivational beliefs, self-regulated learning strategies, and achievement in mathematics. *International Journal of Science and Mathematics Education, 13*, 1355-1375. <https://doi.org/10.1007/s10763-013-9499-4>
- Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to design and evaluate research in education* (8th Ed). McGraw-Hill.
- Fraser, B. J. (2002). Learning environments research: Yesterday, today and tomorrow. In S. C. Goh & M. S. Khine (Eds.), *Studies in educational learning environments: An international perspective* (pp. 1-25). World Scientific.
- Fraser, B. J. (2007). Classroom learning environments. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 103-124). Lawrence Erlbaum.
- Haciomeroglu, G., & Memnun, D. S. (2013). Turkish elementary pre-service teachers' perceptions of the constructivist learning environment in mathematics classes. In A. Drujinin, Z. Kostova, I. Sharuho, & E. Atasoy (Eds.), *The science and education at the beginning of the 21st century in Turkey* (pp. 384-394). St. Kliment Ohridski University Press.
- Iverach, M. R., & Fisher, D. L. (2008). An interdisciplinary investigation of high school students' approaches to learning science: The relations amongst achievement goals, constructivist pedagogical dimensions, motivational beliefs and self-regulated learning. In D. Fisher, R. Koul, & S. Wanpen (Eds.), *Science mathematics and technology education: Beyond cultural boundaries* (pp. 233-254). Curtin University of Technology.
- Johnson, B., & Christensen, L. (2004). *Educational research: Quantitative, qualitative and mixed approaches* (2nd Ed). Pearson Education.
- Johnson, B., & McClure, R. (2004). Validity and reliability of a shortened, revised version of the constructivist learning environment survey (CLES). *Learning Environments Research, 7*, 65-80. <https://doi.org/10.1023/B:LERI.0000022279.89075.9f>
- Kahraman, N., & Sungur, S. (2011). The contribution of motivational beliefs to students' metacognitive strategy use. *Education and Science, 36*(160), 3-10.
- Kahraman, N., & Sungur, S. (2013). Antecedents and consequences of middle school students' achievement goals in science. *Asia Pacific Education Researcher, 22*, 45-60. <https://doi.org/10.1007/s40299-012-0024-2>
- Kahraman, N., & Sungur-Vural, S. (2014). The contribution of gender, socio-economic status and socio-cultural influence to Turkish students' task value beliefs in science. *Research in Education, 91*, 30-44. <https://doi.org/10.7227/RIE.91.1.4>
- Kingir, S., Tas, Y., Gok, G., & Sungur-Vural, S. (2014). Relationships among constructivist learning environment perceptions, motivational beliefs, self-regulation and science achievement. *Research in Science & Technological Education, 31*, 205-226. <http://doi.org/10.1080/02635143.2013.825594>
- Kiran, D., & Sungur, S. (2012). Middle school students' science self-efficacy and its sources: Examination of gender difference. *Journal of Science and Educational Technology, 21*, 619-630. <https://doi.org/10.1007/s10956-011-9351-y>
- Kiray, S. A., Gok, B., & Bozkir, A. S. (2015). Identifying the factors affecting science and mathematics achievement using data mining methods. *Journal of Education in Science Environment and Health, 1*(1), 28-48.
- Kitsantas, A., Baylor, A. L., & Hiller, S. E. (2019). Intelligent technologies to optimize performance: Augmenting cognitive capacity and supporting self-regulation of critical thinking skills in decision-making. *Cognitive Systems Research, 58*, 387-397. <https://doi.org/10.1016/j.cogsys.2019.09.003>
- Kumar, K., Zindani, D., & Davim, J. P. (2019). *Industry 4.0: Developments toward the fourth industrial revolution*. Springer.
- Lai, K. C., Chan, K. W., Ko, K. W., & So, K. S. (2005). Teaching as a career: A perspective from Hong Kong senior secondary students. *Journal of Education for Teaching, 31*(3), 153-168. <https://doi.org/10.1080/02607470500168974>
- LaRocque, M. (2008). Assessing perceptions of the environment in elementary classrooms: The link with achievement. *Educational Psychology in Practice, 24*(4), 289-305. <https://doi.org/10.1080/02667360802488732>
- Likas, A., Vlassis, N., & Verbeek J.B. (2003). The global k-means clustering algorithm. *Pattern Recognition, 36*(2), 451-461. [https://doi.org/10.1016/S0031-3203\(02\)00060-2](https://doi.org/10.1016/S0031-3203(02)00060-2)
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In L. M. Le Cam, & J. Neyman (Eds.), *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, Volume 1: Statistics* (pp. 281-297). University of California Press. <https://projecteuclid.org/euclid.bsmsp/1200512992>
- Martin, M. O., Mullis, I. V. S., Foy, P., & Hooper, M. (2016). TIMSS 2015 international results in science. Boston College, TIMSS & PIRLS International Study Center website. <http://timssandpirls.bc.edu/timss2015/international-results/>
- National Research Council. (2011). *Assessing 21st century skills: Summary of a workshop*. The National Academies Press.
- Pajares, F. (2002) Gender and perceived self-efficacy in self-regulated learning. *Theory into Practice, 41*(2), 116-125. https://doi.org/10.1207/s15430421tip4102_8
- Pintrich, P. R. (2002). The role of metacognitive knowledge in learning, teaching and assessing. *Theory into Practice, 41*(4), 219-225. https://doi.org/10.1207/s15430421tip4104_3
- Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*(1), 33-40. <https://doi.org/10.1037/0022-0663.82.1.33>
- Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1991). *A manual for the use of the motivated strategies for learning* (ED338122). ERIC. <https://files.eric.ed.gov/fulltext/ED338122.pdf>
- Ranjana, J., & Malik, K. (2007). Effective educational process: A data-mining approach. *VINE: The Journal of Information and Knowledge Management Systems, 37*(4), 502-515. <https://doi.org/10.1108/03055720710838551>
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20*, 53-65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Scantlebury, K. (2012). Still part of the conversation: Gender issues in science education. In B. J. Fraser, K. G. Tobin, & C. J. McRobbie (Eds.), *Second international handbook of science education* (pp. 499-510). Springer.
- Schleicher, A. (2019). *PISA 2018: Insights and interpretations*. OECD Publishing.
- Schunk, D. H. (2000). *Learning theories: An educational perspective* (3rd ed.). Prentice-Hall.



- Schwab, K. (2016). *The fourth industrial revolution*. World Economic Forum.
- Scikit-learn. (2020). *Scikit-learn machine learning in Python* (Version 0.23). <https://www.scikit-learn.org>
- Senler, B., & Sungur-Vural, S. (2013). Pre-service science teachers' teaching self-efficacy in relation to personality traits and academic self-regulation. *Spanish Journal of Psychology*, 16, 1-20. <https://doi.org/10.1017/sjp.2013.22>
- Sezgintürk, M., & Sungur, S. (2020). A multidimensional investigation of students' science self-efficacy: The role of gender. *Elementary Education Online*, 19(1), 208-218.
- Şimşek, Ö. F. (2007). *Yapısal eşitlik modellemesine giriş: Temel ilkeler ve LISREL uygulamaları* [Introduction to structural equation modeling: Basic principles and LISREL applications]. Ekinoks Yayınları.
- Sungur, S. (2007). Modeling the relationships among students' motivational beliefs, metacognitive strategy use, and effort regulation. *Scandinavian Journal of Educational Research*, 51(3), 315-326. <https://doi.org/10.1080/00313830701356166>
- Sungur, S., & Güngören, S. (2009). The role of classroom environment perceptions in self-regulated learning and science achievement. *Elementary Education Online*, 8(3), 883-900.
- Tan, P.-N., Steinbach, M., & Kumar, V. (2006). *Introduction to data mining*. Addison Wesley.
- Taylor, P. C., Fraser, B. J., & Fisher, D. L. (1997). Monitoring constructivist classroom learning environments. *International Journal of Educational Research*, 27(4), 293-302. [https://doi.org/10.1016/S0883-0355\(97\)90011-2](https://doi.org/10.1016/S0883-0355(97)90011-2)
- Toivonen, H. (1996). Sampling large databases for association rules. In T. M. Vijayaraman, A. P. Buchmann, C. Mohan, & N. L. Sarda (Eds.), *Proceedings of the 22nd International Conference on Very Large Data Bases (VLDB'96)* (pp. 134-145). Morgan Kaufmann Publishers.
- Valenzuela, J., Nieto, A. M., & Saiz, C. (2011). Critical thinking motivational scale: A contribution to the study of relationship between critical thinking and motivation. *Electronic Journal of Research in Educational Psychology*, 9(2), 823-848. <http://doi.org/10.25115/ejrep.v9i24.1475>
- Voyer, D. & Voyer, S.D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin*, 140(4), 1174-1204. <https://doi.org/10.1037/a0036620>
- Weka. (2020). *Weka the Workbench for machine learning* (Version 3.8). <https://www.cs.waikato.ac.nz/ml/weka/>
- World Economic Forum. (2018). *The future of jobs report*. Centre for the New Economy and Society. http://www3.weforum.org/docs/WEF_Future_of_Jobs_2018.pdf
- Yerdelen, S., & Sungur, S. (2019). Multilevel investigation of students' self-regulation processes in learning science: Classroom learning environment and teacher effectiveness. *International Journal of Science and Mathematics Education*, 17, 89-110. <https://doi.org/10.1007/s10763-018-9921-z>
- Yumusak, N., Sungur, S., & Cakiroglu, J. (2007). Turkish high school students' biology achievement in relation to academic self-regulation. *Educational Research and Evaluation*, 13(1), 53-69. <https://doi.org/10.1080/13803610600853749>
- Zhou, H., & Gao, J. (2014). Automatic method for determining cluster number based on silhouette coefficient. *Advanced Materials Research*, 951, 227-230. <https://doi.org/10.4028/www.scientific.net/amr.951.227>
- Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13-39). Academic Press.

Received: June 02, 2020

Accepted: September 26, 2020

Cite as: Kingir, S., Gok, B., & Bozkir, A. S. (2020). Exploring relations among pre-service science teachers' motivational beliefs, learning strategies and constructivist learning environment perceptions through unsupervised data mining. *Journal of Baltic Science Education*, 19(5), 804-823. <https://doi.org/10.33225/jbse/20.19.804>

Sevgi Kingir
(Corresponding author)

PhD, Associate Professor, Hacettepe University, Department of Primary Education, Ankara, Turkey.
E-mail: ksevgi@hacettepe.edu.tr
Website: <https://avesis.hacettepe.edu.tr/ksevgi>
ORCID: <https://orcid.org/0000-0003-1848-4392>

Bilge Gok

PhD, Associate Professor, Hacettepe University, Department of Primary Education, Ankara, Turkey.
E-mail: bilgeb@hacettepe.edu.tr
Website: <https://avesis.hacettepe.edu.tr/bilgeb>
ORCID: <https://orcid.org/0000-0002-1548-164X>

Ahmet Selman Bozkir

PhD, Research Assistant, Hacettepe University, Department of Computer Engineering, Ankara, Turkey.
E-mail: selman@cs.hacettepe.edu.tr
Website: <https://avesis.hacettepe.edu.tr/aselman>
ORCID: <https://orcid.org/0000-0003-4305-7800>

