

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Jääskelä, Päivikki; Heilala, Ville; Kärkkäinen, Tommi; Häkkinen, Päivi

Title: Student agency analytics : learning analytics as a tool for analysing student agency in higher education

Year: 2021

Version: Accepted version (Final draft)

Copyright: © 2020 Taylor & Francis

Rights: In Copyright

Rights url: <http://rightsstatements.org/page/InC/1.0/?language=en>

Please cite the original version:

Jääskelä, P., Heilala, V., Kärkkäinen, T., & Häkkinen, P. (2021). Student agency analytics : learning analytics as a tool for analysing student agency in higher education. *Behaviour and Information Technology*, 40(8), 790-808. <https://doi.org/10.1080/0144929X.2020.1725130>

Student Agency Analytics: Learning Analytics as a Tool for Analyzing Student Agency in Higher Education

Päivikki Jääskelä^{a,*}, Ville Heilala^b, Tommi Kärkkäinen^b, Päivi Häkkinen^a

^a*Finnish Institute for Educational Research, University of Jyväskylä, Finland*

^b*Faculty of Information Technology, University of Jyväskylä, Finland*

Abstract

This paper presents a novel approach and a method of learning analytics to study student agency in higher education. Agency is a concept that holistically depicts important constituents of intentional, purposeful, and meaningful learning. Within workplace learning research, agency is seen at the core of expertise. However, in the higher education field, agency is an empirically less studied phenomenon with also lacking coherent conceptual base. Furthermore, tools for students and teachers need to be developed to support learners in their agency construction. We study student agency as a multidimensional phenomenon centering on student-experienced resources of their agency. We call the analytics process developed here *student agency analytics*, referring to the application of learning analytics methods for data on student agency collected using a validated instrument (Jääskelä et al., 2017a). The data are analyzed with unsupervised and supervised methods. The whole analytics process will be automated using microservice architecture. We provide empirical characterizations of student-perceived agency resources by applying the analytics process in two university courses. Finally, we discuss the possibilities of using agency analytics in supporting students to recognize their resources for agentic learning and consider contributions of agency analytics to improve academic advising and teachers' pedagogical knowledge.

Keywords: student agency, learning analytics, robust statistics

1. Introduction

The growing capacity of current technologies has made it possible to collect evidence of learning progress in different learning environments. As a result, a new emergent field, learning analytics (LA), has been gaining interest in the last decade (Bond et al., 2018). The purpose of learning analytics is to collect and analyze educational data by creating models and patterns to understand and improve learning and arrangements within learning environments (Conole et al., 2011; Ferguson, 2012; Siemens, 2013). Learning analytics has roots in applied disciplines of machine learning, intelligent tutoring systems, and data mining (Rosé, 2018). According to Chatti et al. (2012), different learning analyt-

ics methods include statistics, information visualization, data mining, and social network analysis.

Moreover, as Saarela (2017) illustrates, the data mining methods used in learning analytics consist of clustering and relationship (association rule) mining in the unsupervised case and classification and prediction (linear and nonlinear regression) methods with supervised data. Zhang et al. (2018) describe the current stage of methods development in learning analytics (after 2015) as a phase of theoretical reconstruction, which is toward systematic analytics.

Learning analytics has been used for providing feedback on students' progress, for predicting their future performance, and for supporting instructors to tailor education based on the needs of the students (Redecker and Johannessen, 2013; Siemens and Baker, 2012; Pardo and Siemens, 2014; Waheed et al., 2018). However, there is still little evidence of the effect of learning analytics on learning outcomes or on the support of learning and teaching in higher education (Viberg et al., 2018; Ferguson and Clow, 2017; Zhang et al., 2018).

*Corresponding author

Email addresses: paivikki.jaaskela@jyu.fi (Päivikki Jääskelä), ville.s.heilala@jyu.fi (Ville Heilala), tommi.karkkainen@jyu.fi (Tommi Kärkkäinen), paivi.m.hakkinen@jyu.fi (Päivi Häkkinen)

To improve learning practice using learning analytics, Viberg et al. (2018) suggest to critically consider the choice of data and purpose of its use while taking into account the discussion in learning sciences as well as the teacher's pedagogical knowledge. Also, the design of LA for improving learning and teaching should emphasize the role of educational theory, e.g., the theoretical knowledge of learning (Wise and Shaffer, 2015) and student agency (Wise, 2014). In line with this suggestion, our leading idea is to focus on an understanding of human experiences and behaviour in learning situations by utilizing the recent conceptual and methodological development in the field and to ground computational facets of learning analytics in this knowledge.

Haggis (2009) criticizes the narrow perspectives on studying learning and calls for grasping the complexity and dynamic interaction related to learning situations in higher education. We see the possibilities with the concept of agency in providing a holistic perspective to understand the constituents of intentional, purposeful, and meaningful learning. The importance of agency in the learning process and institutional strategies to increase agency to enhance academic outcomes was already noted in Thomas (1980). It is through agency that students are seen to attend to their knowledge construction (Scardamalia, 2002), engage in authentic tasks that demand advanced collaborative practices (Damşa et al., 2010), contribute to development of each other, and exert influence on their own educational trajectories (Klemenčič, 2017). Student agency is set as a longstanding educational aim at policy level (OECD 2018), but in educational practice of higher education, however, prerequisites for, development of, and support for agency have received little explicit attention.

Student agency has been empirically scarcely studied in higher education, and the research in the field have focused on small sample of using qualitative methods (e.g., Lipponen and Kumpulainen, 2011; Damşa et al., 2010). A limitation of prior studies on student agency is also that they do not draw from a coherent or holistic conceptual base but rather focus on only some aspects of agency (e.g. epistemic agency, i.e. cognitive responsibility in knowledge construction, Scardamalia, 2002) and centering on individual factors, such as self-efficacy (e.g., Van Dinther et al., 2011). There is a lack of knowledge concerning students' experiences and resources for agency across different fields. This scarcity of studies on student agency in higher education context is surprising taking into account that recent educational research on professional agency (e.g., Eteläpelto et al., 2013; Goller and Paloniemi, 2017) has broadly analyzed the concept and argued for the central role of

agency in experts' work. To support students in their agency construction toward expertise during higher education, research-based tools—that take into account the multidimensional nature of the concept (Jääskelä et al., 2017a)—for analyzing agency experiences and informing students and teachers about them in the course context are needed.

In this study, we connect the conceptual and methodological development on student agency to learning analytics. Linking student agency and learning analytics is not completely new. Prinsloo and Slade (2016) have examined the ways to increase student agency and empower students as active participants in learning analytics instead of being just quantified data objects. However, our setting of linking LA and student agency is different from that of Prinsloo and Slade (2016), focusing on the phenomenon of agency itself—as students' assessments of their own agency resources under the arrangements of an individual course in higher education. We utilize Jääskelä et al.'s (2017a) conceptualization of student agency in the higher education context, which adds to the literature on agency (e.g., Van Dinther et al., 2011; Scardamalia, 2002) by extending the focus beyond unitary dimensions. We use the validated multidimensional Agency of University Students (AUS) Scale questionnaire, similarly to the learning style inventory questionnaire used in Benson et al. (2018) and Jena (2018), to collect data and study students' agency experiences.

The AUS offers a novel methodological contributions by examining individual, relational and participatory resources of agency in the course context. It utilizes a person-/subject-centred approach emphasized in recent literature (e.g., Eteläpelto et al., 2013; Su, 2011) and grounds on the understanding that agency is intrinsically intertwined with learning as an affective experience, cognition, and action in the courses and learning relations (e.g., Su, 2011). We then apply learning analytics methods to acquire knowledge of student-perceived resources of agency in the course context. The overall process can be referred to as student agency-based learning analytics, or *student agency analytics* in short. Therefore, this article makes the following contributions:

1. Introduce the concept of student agency and a quantitative scale developed based on the conceptualization.
2. Describe robust educational data mining methods for student agency data analysis.
3. Depict a service-based architecture that supports

the provisioning of student agency analytics as a service.

4. Examine the applicability of the proposed agency analytics process at the course level. In this respect, we pose the following research questions:

RQ1: What kind of characterizations of student agency can be found using agency analytics at the course level?

RQ2: How different student agency characterizations can inform pedagogical practices at the course level?

2. Theoretical background

Agency is used as a concept in different disciplines, and for this reason the definitions of agency possess various emphases depending on the disciplines' ontological and epistemological bases. For example, in social science agency is understood as individuals' capability to engage in intentional, self-defined, meaningful, and autonomous action in circumstances constrained by power relations and structural, contextual factors (e.g., Archer and Archer, 2003; Foucault, 1975; Giddens, 1984). In social-cognitive psychology, agency is typically linked to individuals' self-processes, intentionality, and self-reflection (e.g., Bandura, 2001), motivational beliefs such as utility value (e.g., Eccles, 2005; Ryan and Deci, 2000), and efficacy and competence beliefs (Malmberg and Hagger, 2009; Schunk and Zimmerman, 2012).

More precisely, within the framework of social-cognitive psychology, Bandura (1986, 2001) sees agency as the mediating factor from thoughts to action intertwined with individuals' intentionality and self-processes, such as motivation and self-efficacy. As Seifert (2004, p. 145) puts it: "Students who feel confident, have a sense of agency and perceive meaning in their academic work will pursue learning goals." Despite the emphasis on individual agency in his definition, Bandura (1986) perceives human agency as being inherently interactional: individuals' construct beliefs of their capabilities through social interaction and experiences in the context.

In educational sciences, the roots for the discussion of agency can be found in the era of enlightenment, when agency was understood as autonomous action through education (Biesta and Tedder, 2007). The idea of student agency is embedded in the constructivist and socio-cultural conceptions of learning (Martin, 2004; Packer and Goicoechea, 2000). In terms of agency, the former emphasizes learners' active role in their construction of

knowledge structures and the manifestation of agency, such as the ability and capability to set goals and to make choices and act on those choices during learning (Zimmerman and Pons, 1986; Martin, 2004). The latter underscores, from the agentic learner perspective, one's participation in social practices and involvement in the social construction of knowledge. Accordingly, learning is not seen as merely epistemic questions of the knowledge structures; it also involves identity construction as a member of the community and the adoption of the practices peculiar to this community (Greeno, 1997; Lave and Wenger, 1991).

During the last decade, the explicit discussion of agency emerged especially within studies on the workplace and lifelong learning (Billett et al., 2006). Agency was generally understood as the *power to act*, manifesting itself as affecting matters, making decisions and choices, and taking stances on work (Vähäsantanen, 2015). A subject-centered sociocultural view of agency (Eteläpelto et al., 2013) brought attention to the *interdependence of individual learners and the socio-cultural context* and the *existence of agency at the individual/subject level*. Also, it stressed a need for acquiring knowledge of subjects' interpretations, meanings, and purposes for actions to understand agency in the dynamic learning situations of the workplace. Studies in this field support the conception that agency plays an important role in expert work demanding creativity, collaboration, and the transformation of work practices (e.g., Hökkä et al., 2017); in constructing meaningful careers (Eteläpelto et al., 2013); and in coping with changes in (work) life and constructing abilities of lifelong learning Su (2011).

In educational context, it is a common belief over various subjects that effective pedagogical practices are linked with increased student agency and deep learning (Ruohotie-Lyhty and Moate, 2015). These types of prior studies centre on the manifestations of agency, such as action with the learning tasks or nature of knowledge construction. For instance, taking agency into account when designing instructions and guidance for a course can aid student learning and strengthen their engagement in challenging learning tasks (Lindgren and McDaniel, 2012). For increased agency in learning, the instructional setting should activate students to ask the so-called educationally productive questions, which support the building of knowledge structures (Scardamalia and Bereiter, 1991). Also, students' possibilities for participatory learning (Starkey, 2017) as well as for contributing to their educational settings (Bransford et al., 2006) have been presented as ways of increasing agency. Two qualitative stud-

ies reported the forms of student agency in the contexts of collaborative knowledge creation (Damşa et al., 2010) and collective inquiry learning courses (Lipponen and Kumpulainen, 2011). In these studies, agency manifested itself in action and discourses as to varying degrees knowledge-related (epistemic agency), and process-related/relational agency, with reflecting on the performance of the tasks. Damşa et al. (2010) concluded that agency/action including shared epistemic, intentional and intersubjective characters form the capacity among the students that enables them to successfully carry out task. As for, Lipponen and Kumpulainen (2011) noticed that pre-service teachers' agency can be transformative and cultivate them to upcomping profession through the reciprocity and dialogue between the teacher and students, and giving students space and opportunities to take initiatives and influence the course (e.g., Lipponen and Kumpulainen, 2011). Previous studies have also acknowledged that students may experience the same pedagogical practices differently and do not always exercise their agency for purposeful learning and in growth-oriented ways (e.g., Harris et al., 2018). Agency is, for example, resourced or constrained by factors in the sociocultural context, such as power relations, experiences and evaluations of trust and equality among the participants (e.g., Hökkä et al., 2017; Eteläpelto and Lahti, 2008; Juutilainen et al., 2018), and of a sense of being capable in performing the tasks (e.g., Seifert, 2004; Ayllón et al. 2019). Ayllón et al. (2019) presented evidence that teachers' involvement in supporting students and especially their self-efficacy were strongly and positively related to achievement. Students got higher marks when they perceived their teachers as dependable and available to offer resources, and when they felt capable themselves of organizing and implementing the courses of action necessary to acquire knowledge. These findings concerning the link between students' self-efficacy beliefs and performance are supported by Bandura (1982), who sees the perceived self-efficacy an important component of agency. Thus, to understand this complex dynamics in learning situations, agency as student experiences and as perceived resources and affordances in context need to be studied.

Based on the previous literature, Jääskelä et al. (2017a) constructed a multidimensional view to study student agency in the higher education context and conceptualized agency as a student's experience of access to/having (and using of) personal, relational (i.e., interactional), and context-specific participatory resources to engage in intentional and meaningful action and learning. *Personal* resources include students' perceived self-efficacy (e.g., students' sense of having self-

confidence as learner) and competence beliefs (e.g., sense that understand and having competence needed for learning contents in the course). *Relational resources* encompass, in particular, power relations between the teacher and students, manifesting as students' experiences of trust and emotional support from the teacher as well as experiences of being treated as equals with other students in the course. *Participatory resources* refer to set of factors that enables active and engaged participation, particularly students' self-assessed interest and opportunities for peer support as well as opportunities to make choices, influence, and actively contribute to learning situations in the course. When self-assessing agency, one may experience e.g. a strong sense of agency regard to participation or influencing but not perceive oneself as competent or empowered afforded by the relations in the context.

Jääskelä et al. (2017a) see agency as being dynamic, contextually situated, and relationally constructed in nature (c.f., Emirbayer and Mische, 1998; Eteläpelto et al., 2013). Their conceptualization of agency is in line with the notions (by Klemenčič, 2017) that i) agency is shaped in a particular context of action; ii) the experiences of agency can vary in different situations; and iii) different temporalities affect students' sense of what can and should be accomplished in a given situation (by acting accordingly). When studying agency as individual experiences, Jääskelä et al. (2017a) present analyses focusing on the students' experienced opportunities (e.g., for ownership and influence) and their self-assessed capabilities as learners (which are constructed in interaction through the beliefs, c.f., Bandura (1986))—rather than agentic action (see Klemenčič, 2017; O'Meara et al., 2014). Ideally, these foci of the study force to take attention on the prerequisites and affordances for practicing and constructing agency experienced by the students in the courses' learning situations.

3. Materials and methods

3.1. Research design

The research design is based on the holistic conceptualization of the student agency in higher education as presented in the previous section. The research process presented in this paper is organized according to the general aims of the research as listed in the introduction: conceptualization of the student agency in higher education (Section 2), quantification of student agency analytics (Sections 3.3 and 3.4), provisioning of the analytics processes as a service (Section 4), and, finally, study applicability of the proposed agency analytics process at the course level in Section 5.

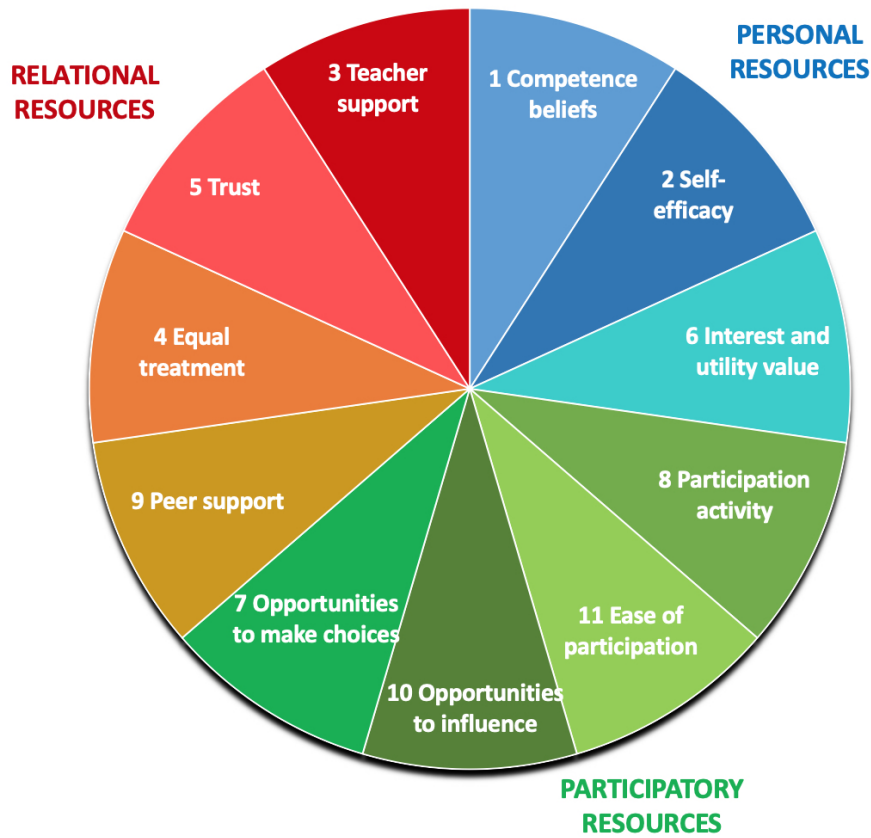


Figure 1: The dimensions of the Agency of University Student (AUS) Scale (Jääskelä et al., 2017b).

3.2. Participants and data collection procedures

All the participants studied in the courses, whose teachers participated in the university level cross-disciplinary teaching network and voluntarily allowed to implement the questionnaire in their courses. The online questionnaire responses were collected at the end part of the courses before final grades or completing the course.

We use two different datasets. The first dataset, later referred as *the reference dataset*, is used to develop the learning analytics workflow in Section 3.4. The reference dataset, which was also used in AUS Scale validation (see Section 3.3), consisted of 270 students' responses to AUS Scale in a Finnish university (167 women; 102 men; missing data for one participant). The participants represented various disciplines and their mean age was 22.66 years ($SD = 4.63$, range 18–55).

The second dataset, later referred as *the empirical dataset*, is used to examine the applicability of the pre-

sented agency analytics process at the course level. The empirical dataset consisted of 208 students' responses to AUS Scale from two faculties (information technology ($n = 130$) and teacher education ($n = 78$)) in the same university as where the reference dataset was collected. The participants' mean age in the information technology was 25.11 years ($SD = 6.09$, range 19–55), and in the teacher education 20.77 years ($SD = 1.93$, range 18–28). The participants were chosen because the courses represented two different scientific fields and two different forms of instruction. However, the common features were that both courses represented basic studies of their respective study programs as well as scientific fields with an applied professional focus.

As described earlier, respondents in the empirical dataset represented two different university courses. In the first course, students in the Faculty of Information Technology studied in the first computer programming course (CS1 equivalent). The course top-

ics included basic principles of structured programming, algorithms, and data types and structures for simple problem-solving. The course consisted of lectures, programming labs, self-study, assignments, and a final exam. At the end of the course, students also designed and created a small program using C# programming language. Study success of individual students was assessed in grades from 1 to 5 (highest) given by the teacher. The course is a fundamental part of the bachelor-level studies. Thus, extensive support was provided for students by teachers, teaching assistants, and peers.

Students in the second course in the empirical dataset studied in the Department of Teacher Education. The students took part in basic studies in education in the primary school teacher education program. Primary school teacher training aims to train educational experts with a strong communal and exploratory approach to learning, teaching, and education. During the first two years, a large part of the studies is done in groups of 10 to 15 students facilitated by one lecturer. The groups are formed at the beginning of the studies. Each group has its own specific theme (e.g., multidisciplinary learning and teaching, educational technology, multilingualism), which offers a specific perspective to study the contents of the curriculum. One study group was especially concentrating on student agency, which was realized as teacher's pedagogical emphasis on agency (e.g., making effort to establish trust between teacher and students) and as having course content about agency. In general, the students were required to commit to the group and participate actively in thematic group discussions.

3.3. Measures

Based on their conceptualization work, Jääskelä et al. (2017a) developed the AUS Scale and examined/validated the factor structure of the scale with confirmatory factor analyses (CFA) (Jääskelä et al., 2017b, 2019 submitted). The analyses resulted in the 11 factor model with an acceptable model fit: ($\chi^2(1529, n = 270) = 2527.96, p < .001; CFI = 0.86; TLI = 0.85; RMSEA = 0.05; SRMR = 0.07$). The final scale consists of 58 items at the course level and capture three main domains of agency resources, and their respective 11 dimensions (Figure 1):

A. Personal resources

1. Competence beliefs
2. Self-efficacy

B. Relational resources

3. Equal treatment
4. Teacher support
5. Trust

C. Participatory resources

6. Participation activity
7. Ease of participation
8. Opportunities to influence
9. Opportunities to make choices
10. Interest and utility value
11. Peer support

Each dimension of student agency contains three to seven items rated using a five-point Likert scale (1 = fully disagree; 2 = partly disagree; 3 = neither agree nor disagree; 4 = partly agree; and 5 = fully agree). Examples of the items tapping each resource area include: "Thus far I have understood the presented course contents well" (Personal resources–Competence beliefs), "I believe I will succeed in the more challenging tasks in the course" (Personal resources–Self-efficacy), "I feel that I have had an equal position with the other students in this course" (Relational resources–Equal treatment), "I feel that I can trust the course teacher" (Relational resources–Trust), "It has been possible for me to express my thoughts and views without being afraid of ridicule" (Participatory resources–Ease of participation), and "I feel that I had an opportunity to choose course contents that interested me" (Participatory resources–Opportunities to make choices). Abbreviated items of the AUS Scale have been presented in Appendix A.

To describe the agency analytics method, we use the reference dataset as described in Section 3.2. The first step in the analytics process is to invert the scale of reverse items (Jääskelä et al., 2017a) from [1, 5] into [5, 1] using linear scaling. As described in section 2, we then compute the values of the 11 student agency factors. The basic computation of factors uses standardization and linear scaling with the factor pattern matrix. However, to improve the understandability between the original Likert scale items and the computed factors, we propose applying a rescaled factor pattern matrix as follows: The original matrix is multiplied by the inverse of the diagonal matrix, which is obtained by applying the basic factor pattern matrix to the unit vector of the number of items. In doing this and omitting the z-scoring of factors we enforce the range of computed factors from 1 to 5, similarly to the raw data. In practice this just

changes the scale of factors and does not affect comparisons or further processings of the factor values.

To prevent the underestimation of the factors, the missing values in the raw data are filled using the nearest neighbor (NN) imputation (Chen and Shao, 2000) with, similarly to the robust statistics, minimal assumptions on the actual distribution of data. The distribution of the reference dataset is illustrated in Figure 2, and the distribution of the rescaled factors is depicted in Figure 3. To conclude, the multiplication by the scaled factor pattern matrix together with the NN imputation is the basic transformation from the original questionnaire scale into the factor space.

3.4. Learning analytics methods

Next we describe the purpose and methods for the main phases of the agency analytics process. The methods are described by using the reference dataset. Currently the processing takes place off-line, after the AUS data collection; immediate on-line feedback of agency is part of future research. The volume of the processed data is typically small, composed of tens or hundreds of observations on number of the scale items. Hence, the scalability of the processing methods is not a primary concern, but their reliability and proven capabilities with educational datasets are taken as prerequisites for analysis methods selection.

We use here a special set of learning analytics and educational data mining methods (Kärkkäinen and Heikkola, 2004; Kärkkäinen and Äyrämö, 2005; Saarela and Kärkkäinen, 2015; Hämäläinen et al., 2017; Saarela and Kärkkäinen, 2017; Saarela et al., 2017; Niemelä et al., 2018), whose basic constructs are based on robust statistics (Huber, 1981; Hettmansperger and McKean, 1998; Kärkkäinen and Heikkola, 2004). The main reason underlying the choice of robust, non-parametric methods is the typically small amount of data on the Likert-scale, which prevents the use of classical, second-order statistical methods relying on assumptions of Gaussian error distribution of the statistical estimates (Huber, 1981; Hettmansperger and McKean, 1998; Kärkkäinen and Heikkola, 2004).

3.4.1. Unsupervised factor profiles using robust clustering

The purpose of the basic agency analytics processing is to provide information on the agency for *i*) individual students, also in comparison to peers in the same course, and *ii*) course teacher(s), about the student agency profiles in the course. We describe the analytics methods for these two unsupervised purposes next.

As argued in (Saarela and Kärkkäinen, 2015), the natural error distribution for a discrete set of integer data in the Likert scale [1, 5] is the uniform distribution. When such data are linearly transformed as a result of the multiplication with a scaled factor pattern matrix with 3–7 dominant factor loadings, we cannot assume that the error distribution would be transformed as the Gaussian distribution. Hence, the statistical methods for the unsupervised processing of the agency factor data must be based on nonparametric, robust methods (Huber, 1981; Hettmansperger and McKean, 1998; Kärkkäinen and Heikkola, 2004), which allow deviations from normality assumptions while still producing reliable and well-defined estimators.

The most central estimate in statistics is the so-called location estimate, which depicts the general behavior of data. Instead of the data mean, the two basic location estimates in robust statistics are the median and spatial median (Kärkkäinen and Heikkola, 2004). The median, a middle value of the ordered coordinate-wise sample—unique only for an odd number of points (Kärkkäinen and Heikkola, 2004)—, is inherently univariate and discrete, having thus very low sensitivity for the 11 agency factors. On the contrary, the spatial median is truly a multidimensional location estimate and varies continuously in the value range, similarly to the mean. Moreover, the spatial median has many attractive statistical properties: it is rotationally invariant, and its breakdown point is 0.5; i.e., it can handle up to 50% of the contaminated data, which makes it very appealing for datasets with imbalanced distributions and outliers, possibly in the form of missing values. For such cases, the available data strategy together with the successive-overrelaxation solution method determine an efficient and reliable approach to estimate data location (Kärkkäinen and Äyrämö, 2005; Äyrämö, 2006).

The spatial median for the reference dataset with 58 missing values (0.4%) was computed and rescaled into the factor space. This is illustrated in Figure 3. This overall factor profile is referred to as the general agency profile (GAP) of a course, which can be used by a student in comparison to her/his own profile, and by a teacher, concerning the general student agency profile of the course.

Our next task, again proceeding with the reference data and robust procedures, is to consider what kind of different student agency profiles would be visible in the course under analysis (see Saarela and Kärkkäinen, 2015; Gavriushenko et al., 2017). The role of these profiles is to summarize the basic forms of student agency in the course for the teacher. Both the form and the number (K) of different student profiles in the factor

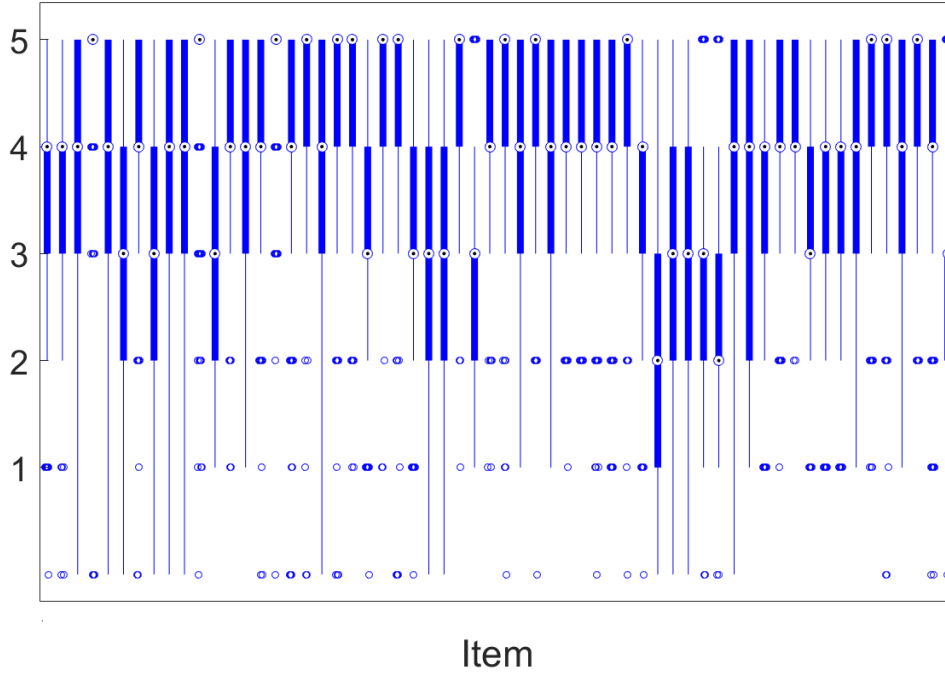


Figure 2: Boxplot of the reference dataset (missing value as zero).

representation should be determined. For this purpose, we use the robust k-SpatialMedians++ algorithm as described and theoretically analyzed (local convergence guaranteed) in Hämäläinen et al. (2017). To estimate the number of clusters K , the best cluster validation indices (CVIs) from Jauhiainen and Kärkkäinen (2017) and Hämäläinen et al. (2017) were applied, with the simplified formulae as defined in Niemelä et al. (2018). For clustering, the factor data were min-max scaled into $[-1, 1]$, and 1,000 repetitions were used similarly to Hämäläinen et al. (2017).

The clusters were computed and compared for the values $K = 2 - 10$ using CVIs because this result needs to be disseminated to the teacher(s), and, hence, a small number of profiles is preferred. The results are illustrated in Figure 4. All cluster indices suggested 2–4 clusters, which are also seen as the knee points (Thorndike, 1953) in Figure 4 (left). The Pakhira-Bandyopadhyay-Maulik (PBM) cluster validation index, which was also found most useful in Tuhkala et al. (2018), suggested four clusters (Figure 4 (right)) which was fixed as the number of different agency profiles communicated to the teacher.

The visual information of different student agency

profiles, compared to GAP, is illustrated in Figure 5. The four profiles are first ordered in ascending order based on the total mass (i.e., sum of values). These profiles and their deviations from the whole student agency profile are then visualized. With the reference agency data, the sizes and portions (in percentages) of the four clusters in Figure 5 were as follows: P1(38/14%), P2(78/29%), P3(98/36%), and P4(56/21%).

The low number of student agency profiles in Figure 5 allows visual interpretation of the differences between the different factors in the profiles. However, as suggested in Cord et al. (2006) and generalized to the population level in Saarela et al. (2017), the feature separability ranking of the robust clustering result can be estimated using the H statistics of the nonparametric Kruskal–Wallis test (Kruskal and Wallis, 1952). Moreover, one can use the pairwise Mann–Whitney U test as the post hoc test to estimate the separability of the factors between any two profiles.

With the four profiles of the reference agency data, the ranking of student agency factors by means of how strongly they separate the profiles is the following (rounded value of H statistics in parentheses):

10 - Opportunities to influence (223)

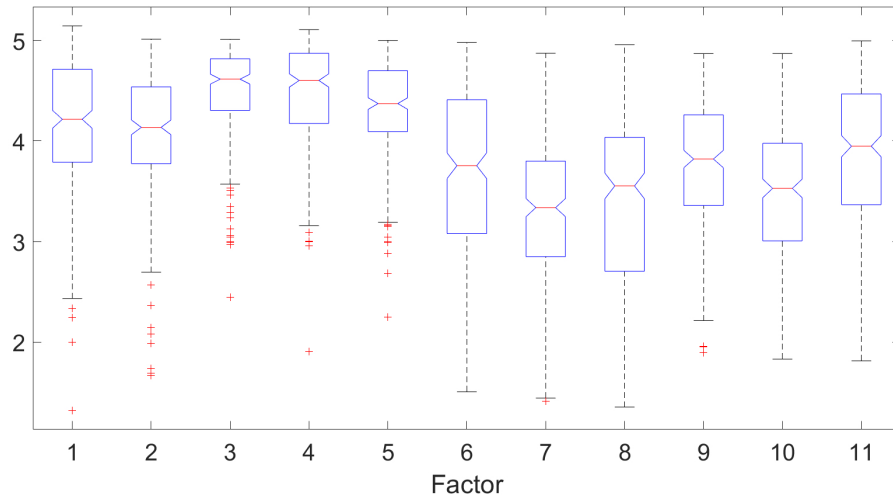


Figure 3: Boxplot of the agency factors.

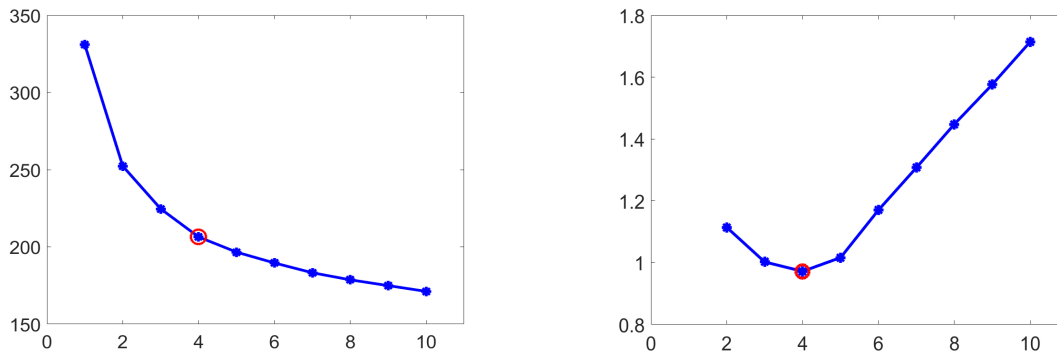


Figure 4: Clustering error (left) and the PBM CVI (right) for $K = 2 - 10$. Minimum of PBM ($K = 4$) marked.

- 8 - Participation activity (204)
- 5 - Trust (197)
- 9 - Peer support (195)
- 11 - Ease of participation (191)
- 6 - Interest and utility value (171)
- 3 - Teacher support (155)
- 4 - Equal treatment (134)
- 7 - Opportunities to make choices (119)
- 2 - Self-efficacy (96)
- 1 - Competence beliefs (71)

The participants here represent a versatile set of Finnish university students with strict entrance criteria. Therefore, the personal agency resources which are generally in a high level (see Figure 3) provide the smallest separation between the four student profiles.

The profile view and the factor deviation analysis provide information on those resources and factors that can

be affected by pedagogical arrangements. For example, in the reference data the influence opportunities separated the student profiles three times stronger compared to the competence beliefs. Hence, mixed perceptions on influence opportunities together with a generally lower GAP value and high separability of the participation activity suggest improvements toward this direction in course arrangements.

In summary, the student agency profile analysis showed that the general level of student agency, GAP, was high in the reference dataset. There was a group of students ($n = 56, 21\%$) who evaluated their agency even higher, close to the maximum level 5. But also a group of students ($n = 38, 14\%$) with a clearly weaker level of agency was identified. The two middle groups had a profile close to GAP, but the second largest group of students ($n = 78, 29\%$) had strictly smaller than normal

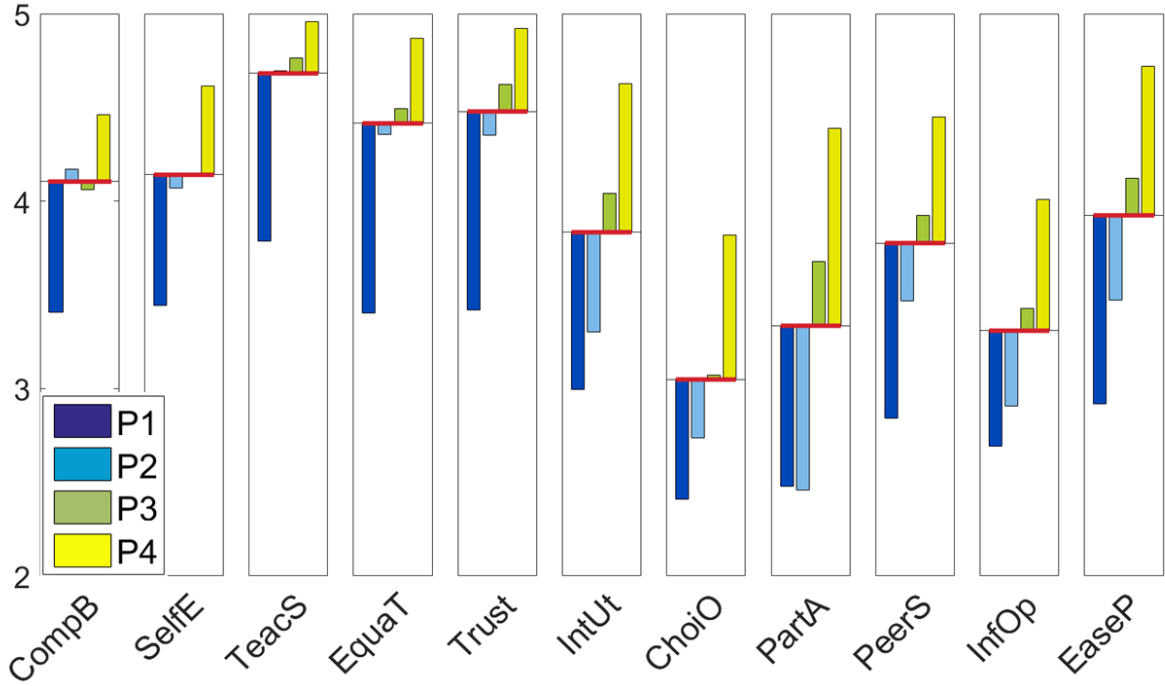


Figure 5: Deviations of the four agency profiles from the GAP in the reference dataset.

agency on “8–Participation activity”, “6–Interest and utility value”, “10–Influence opportunities”, and “11–Ease of participation”. These factors together with the “9–Peer support” were the most separating factors between the four student agency clusters. Competence and Self-efficacy (representing the personal resources of agency) were found least significant.

3.4.2. Supervised linkage of agency factors with course grades

From the assessment point of view, it might be interesting to investigate the possible effect of student agency resources on course grades. As an explorative measure, we utilize unsupervised analysis in order to examine which factors of student agency might be the most important in explaining the course grades. A supervised analysis can be progressed if we can link data on course grades to the student agency factors. In the case of course grades (Saarela and Kärkkäinen, 2015), the latent ingredient in the supervised analysis is the way the course is being evaluated by the teacher, i.e., whether, e.g., student activity is part of the grading or not. Information on course grades from the courses where the reference dataset was collected was not available. Therefore, we only briefly depict the methods for the supervised processing next and give real analytics

results in Section 5.1.

From a machine learning perspective, the most useful method is the estimation of the feature importance of a predictive model from agency factors to course outcomes (John et al., 1994; Guyon and Elisseeff, 2003; Liu and Motoda, 2012). The model can be of restricted form and flexibility, such as in the statistical regression analysis (Hastie et al., 2009), or universal, being able to approximate any nonlinear, deterministic behavior, such as the MultiLayered Perceptron (MLP) or Radial-Basis Function Network (Hornik et al., 1989; Park and Sandberg, 1991). For a discrete or discretized performance output (Dougherty et al., 1995), the first natural way to link unsupervised and supervised information is to cross-tabulate the four student agency profiles with the outcomes (e.g., course grades) and use the χ^2 test (Everitt, 1992).

The statistical regression analysis (Hastie et al., 2009) testing the effects of individual variables can be used for ranking the agency factors, and, if some of the factors have no statistical significance, to remove them from further supervised modeling. The significant factors can then be used to construct a universal MLP model (Saarela and Kärkkäinen, 2015; Kärkkäinen, 2015). This model can be built from factor values to outcomes or by using the residual of the linear model as the target

of the nonlinear regression. In the latter case, the factor significance is obtained as the combination of both processing phases.

Without going into the details, which are documented in the references given, the basic components of the method read as follows: We train the one-hidden-layer feedforward neural network with a sigmoidal activation function for the min-max scaled input-output data (Kärkkäinen, 2002). The size of the hidden layer m and the weight decay parameter β (see Kärkkäinen, 2002) are grid-searched using the 10-fold cross-validation error with the Dob-SCV folding strategy (Moreno-Torres et al., 2012; Kärkkäinen, 2014; Kärkkäinen, 2015). The mean absolute value of the analytic sensitivity (MAS) is then used to estimate the factor sensitivity. *Differently from the earlier work* (Saarela and Kärkkäinen, 2015; Kärkkäinen, 2015), where a new MLP was trained after fixing the metaparameters m and β , we here propose to compute the final MAS values for ranking the factors as the mean over the foldwise MAS values. In this way, we do not need additional training of the MLP model, and the MAS values directly correspond to the 10 different MLP models providing the smallest cross-validation error.

4. Student agency analytics as a service

In this section, we describe the process for automating the student agency analytics. To utilize analytics in real learning and teaching settings, one needs to address two essential requirements: 1) the analytics must be implementable into existing learning environments or management systems, and 2) the process must align with the General Data Protection Regulation (GDPR) (Regulation [EU] 2016/679, 2016). Thus, we decided to separate the data processing into its own service using a microservice architecture. Also, we make use of the controller-processor dichotomy and pseudonymization in order to comply with the GDPR. The purpose is to hand over the full control of personal data to the instance representing the users (i.e., educational institution). We call this approach Student Agency Analytics as a Service (SA⁴S).

4.1. The agency analysis process as a whole

The process starts by collecting AUS data from students taking part in a higher education course using the validated questionnaire (Jääskelä et al., 2017a). In the sequence diagram in Figure 6 the starting point for the questionnaire is the course learning environment in a learning management system (LMS). However, the

starting point can be whatever system is used in the educational institution. The functionality of the analytics inside the LMS is implemented as a built-in feature, a plugin, or a module to guarantee ease of use.

After the student completes and submits the questionnaire, the LMS extracts the numerical questionnaire values. The LMS then transforms the values into a predefined form, for example in JSON data format. Before passing the data to the processor, the LMS pseudonymizes the data by assigning unique identifiers. The linking information used to re-identify the student, and the context is saved under the control of the educational institution.

The connection between the LMS and the agency analytics service provider uses a well-defined interface, for example, Representational State Transfer (REST) over a secure TLS connection. The analytics service receives the data from the students in the same course, and when enough data are collected, the analysis is executed as depicted in section 3.4. After analysis, the service sends the analysis results including identifiers back to the LMS. The data are re-identified using the linking information and visualized. The student receives a personal agency factor analysis in relation to the whole course factors. The course teacher gets an aggregated overview containing the four agency profiles.

As argued in section 1, taking ethical considerations into account is essential in LA. Our overall process of analyzing student agency is an effort to address some of the challenges presented by Ferguson et al. (2016). The purpose of student agency analytics is to use the collected data to benefit learners. It aims to provide accurate, timely, and understandable results to the end users. The purpose of separating processor and controller in addition to the use of pseudonymization is to comply with the law and clarify the ownership of the data.

4.2. Using microservices architecture

In microservices architecture, applications are composed of several independent software components collaborating with each other (Lewis and Fowler, 2014). According to Namiot and Sneps-Sneppé (2014, p. 24), a microservice is a “lightweight and independent service that performs single functions and collaborates with other similar services using a well-defined interface.” Newman (2015) describes the key benefits of using the microservices architecture, which are technology heterogeneity, composability, and replaceability. Different microservices working together can be implemented using different technologies. They can also be used in multiple different ways or even replaced completely. By using microservices in analyzing student agency, we can

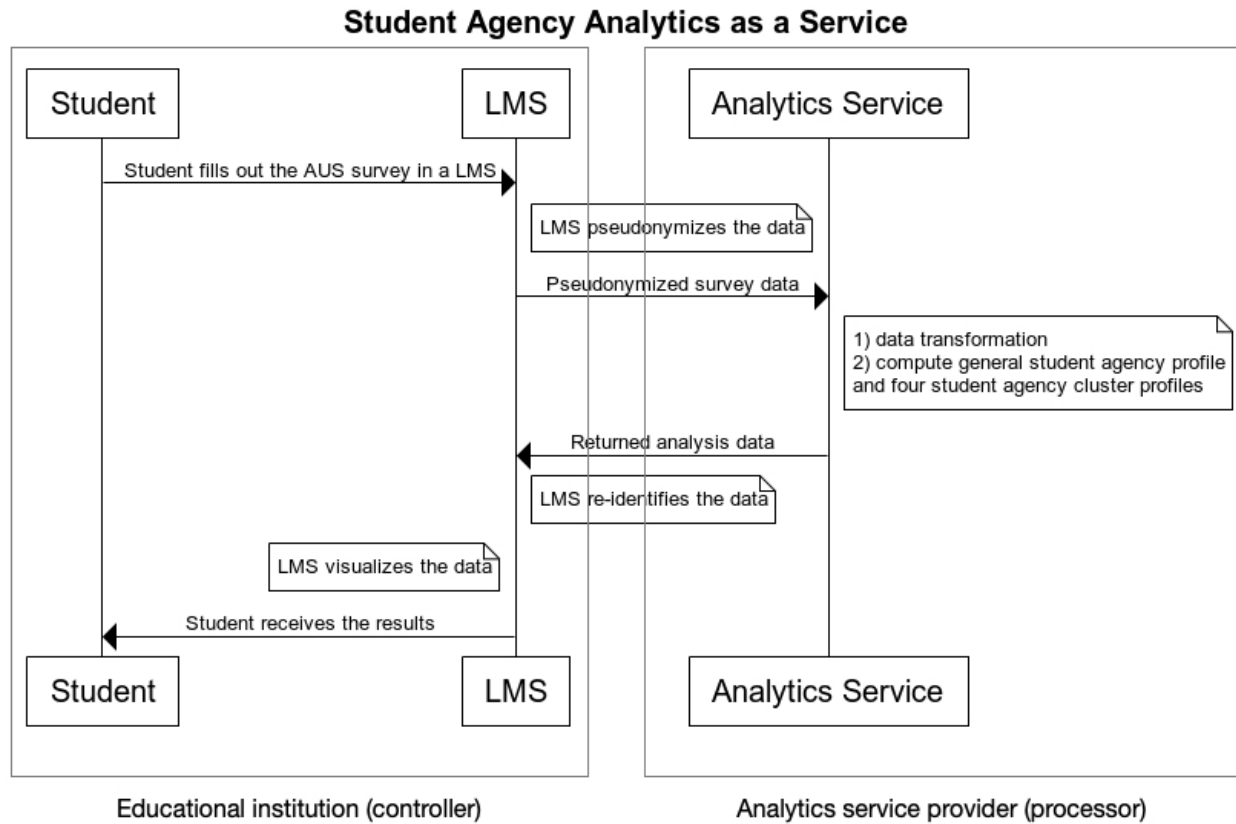


Figure 6: A sequence diagram presenting the flow of data between processor and controller.

make our analysis component more interoperable and reusable as it can be used as a service in different systems. We also maintain control of the component and analysis model while releasing the control of personal data.

4.3. Processing pseudonymized data

The GDPR (Regulation [EU] 2016/679, 2016) defines two entities who take part in the handling of personal data. Article 4(7) of the GDPR defines the *controller*, which “means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data.” The same article defines the *processor*, which “means a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller.”

Article 4(5) of the GDPR also introduces a concept of *pseudonymization*. Pseudonymization is a specific type of de-identification, which “both removes the association with a data subject and adds an association between a particular set of characteristics relating to the

data subject and one or more pseudonyms” (ISO, 2017, p. 5). Ferguson et al. (2016) mention anonymizing and de-identifying individuals as one of the many important challenges in LA. According to Recital 28 of the GDPR, the purpose of pseudonymization is to help controllers and processors fulfill the data-protection obligations and reduce the risks to the data subjects. As stated in Article 25 of the GDPR, pseudonymization is one but not the only way of implementing appropriate technical and organizational measures to meet the requirements of privacy by design and by default. Also, it is worth noting that Recital 26 of the GDPR states that pseudonymized data are still personal data if the person can be identified by using additional information. Another important concept, *data minimization*, is also worth mentioning as it in addition to pseudonymization helps controllers and processors to comply with the regulation. The principle of data minimization in Article 5(1c) of the GDPR states that only necessary data should be collected.

When handling the student agency data, our aim is to use pseudonymization and follow the data minimization principle to collect only necessary data. The AUS

Scale data consist of numerical Likert scale values ranging from 0 to 5. As such, it is impossible to identify a person based on only these numerical data. To allow the agency analytics results to be linked to the identifiable right person after analysis, two unique identifiers are attached to the AUS Scale data. One identifier is used to identify the person, and the other identifier is used to determine the course or other context where the AUS survey has been executed. The data controller (i.e., educational institution) has the linking information, which is used to re-identify the person and attribute the analysis results to the right student in the proper context based on the unique identifiers. Only a minimal amount of data is handled, and data are pseudonymous from the data processor point of view.

5. Results

5.1. Basic course on computer programming

The answers were clustered into four profiles, as described in the section 3.4.1. Figure 7 illustrates the GAP of the course and the deviating profiles from the GAP for the four groups of students. Based on the non-parametric Kruskal–Wallis test, the three most-separating agency factors between the student profiles were trust, self-efficacy, competence beliefs, and ease of participation.

The agency profiles and their deviations from the GAP are presented in Figure 7. The students in the profile P1 assessed their agency resources lower than other students in all 11 dimensions of agency. On the contrary, the students in P4 assessed their agency higher than assessed in the GAP level in most of the dimensions of agency, especially related to individual (competence beliefs, self-efficacy) and relational (teacher support, equal treatment, trust) resources of student agency. The students in the P3 profile assessed their individual resources of agency as lower than assessed in the GAP level. However, their participatory resources of agency appear slightly higher than the GAP level. This is clearly seen in the dimensions of participatory activity, peer support, opportunities to influence, and ease of participation. This might be due to the extensive support provided for students during the course.

Students were asked permission to combine their agency profiles with their course grades for research purposes. A total of 71 % of the respondents (92 out of 130) gave permission. A chi-square test of independence was performed to examine the relation between student agency profile and course grade. The relation between these variables was significant, $\chi^2(12, n =$

Table 1: The number of course grades (1–5) in the four agency profiles (P1–P4) in the course on computer programming, $\chi^2(12, n = 92) = 27.9, p < .01$.

	Course grades				
	1	2	3	4	5
P1	6	7	5	3	1
P2	4	1	5	2	10
P3	2	3	4	11	16
P4	4	1	2	1	4

92) = 27.9, $p < .01$. Table 1 shows that there are higher grades (4 and/or 5) in the P3 profile, which was characterized by higher participatory resources of agency compared to the GAP.

Because of small number of instances in an individual cell in Table 1, we next merged low- and high-grade values to create a binary variable related to the course performance. More precisely, the lower grade was linked to original grades 1–3 and the higher grade encoded original grades 4 and 5. Table 2 presents the contingency table and chi-square test of independence between the binarized grades and the 4 agency profiles. The relation between aforementioned variables was also significant, $\chi^2(12, n = 92) = 18.3, p < .001$. The result also indicates a positive link between the level of agency and the performance in the course.

Table 2: The number of lower and higher grades in the four agency profiles (P1–P4) in the course on computer programming, $\chi^2(3, n = 92) = 18.3, p < .001$.

	Agency profiles			
	P1	P2	P3	P4
Lower grade (1-3)	18	10	9	7
Higher grade (4-5)	4	12	27	5

Supervised analysis as depicted in Section 3.4.2 could be used to examine the linkage between student agency and course grades in the basic course on programming. Because of the size of the data, we applied here MLP classifier for the binarized grades in Table 2, and used the mean of the MAS values over the two classes as the sensitivity measure. The four most important agency factors were *i) competence beliefs*, *ii) self-efficacy*, *iii) teacher support*, and *iv) equal treatment*. Classification accuracy over the test folds was 77.2% and the four agency factors explained c. 70% of the total sensitivity of the classifiers.

5.2. Basic course on educational sciences

The agency profiles and their deviations from the GAP are presented in Figure 8. Based on the

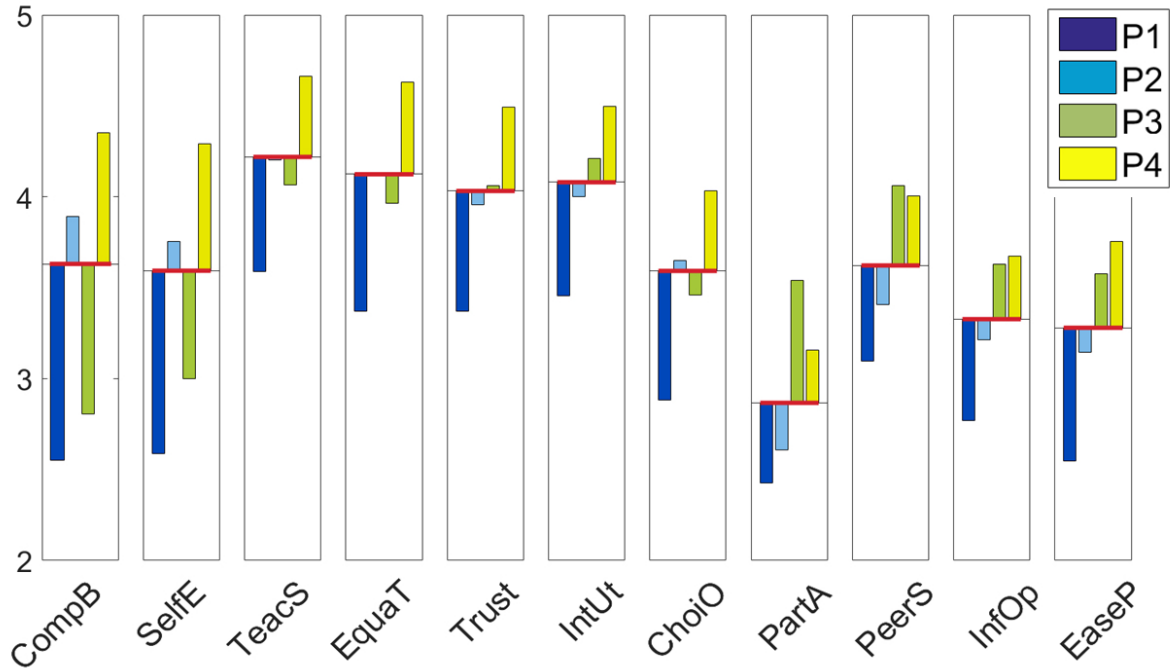


Figure 7: Deviations of the four agency profiles from the GAP in the Faculty of Information Technology.

non-parametric Kruskal–Wallis test, the three most-separating agency factors between the student profiles were opportunities to make choices, participation activity, ease of participation, and peer support. Students’ ratings of their agency resources in the P2 and P3 groups are close to the GAP level. However, the students in the P2 group perceived their agency resources slightly lower than their counterparts in P3. The students in P4 assessed their agency resources close to maximum with respect to all factors. Further, the students in P1 assessed their agency as lower than the GAP, especially in the dimensions measuring the participatory resources of agency (e.g., opportunities to influence). The GAP of the students in the Department of Teacher Education is generally higher compared to the IT students studying programming (see Figure 8 and Figure 7).

Table 3: Thematic study groups (A–F), four agency profiles (P1–P4), and respective student frequency in each profile in the course on educational sciences, $\chi^2(15, n = 64) = 16.3, p = 0.36$.

		Study groups					
		A	B	C	D	E	F
Agency profiles	P1	0	1	1	0	2	2
	P2	2	5	2	4	3	3
	P3	6	6	1	4	4	4
	P4	1	0	1	7	3	2

Students in the course on educational sciences did not receive a numerical grade of their learning. Thus, supervised agency analytics by means of learning results was omitted. However, a chi-square test of independence was performed to examine the relation between student agency profile and study group. The relation between these variables was not statistically significant, $\chi^2(15, n = 64) = 16.3, p = 0.36$. However, Table 3 shows that Group D had more students that represented the profile P4 (high level of agency resources) than other profiles. As mentioned in the context description in Section 3.2, the aforementioned group had agency as their special theme. While the result is not statistically significant, we still consider it as an interesting finding.

6. Discussion

There is a need to support students’ agency construction in higher education to respond to the demands of current working life. However, this presupposes the development of tools for analyzing students’ agency experiences and informing students and teachers about them. We utilized the validated Agency of University Students (AUS) Scale and unsupervised robust clustering methods to analyze student agency. Further, we proposed a service-based system for automating the analysis.

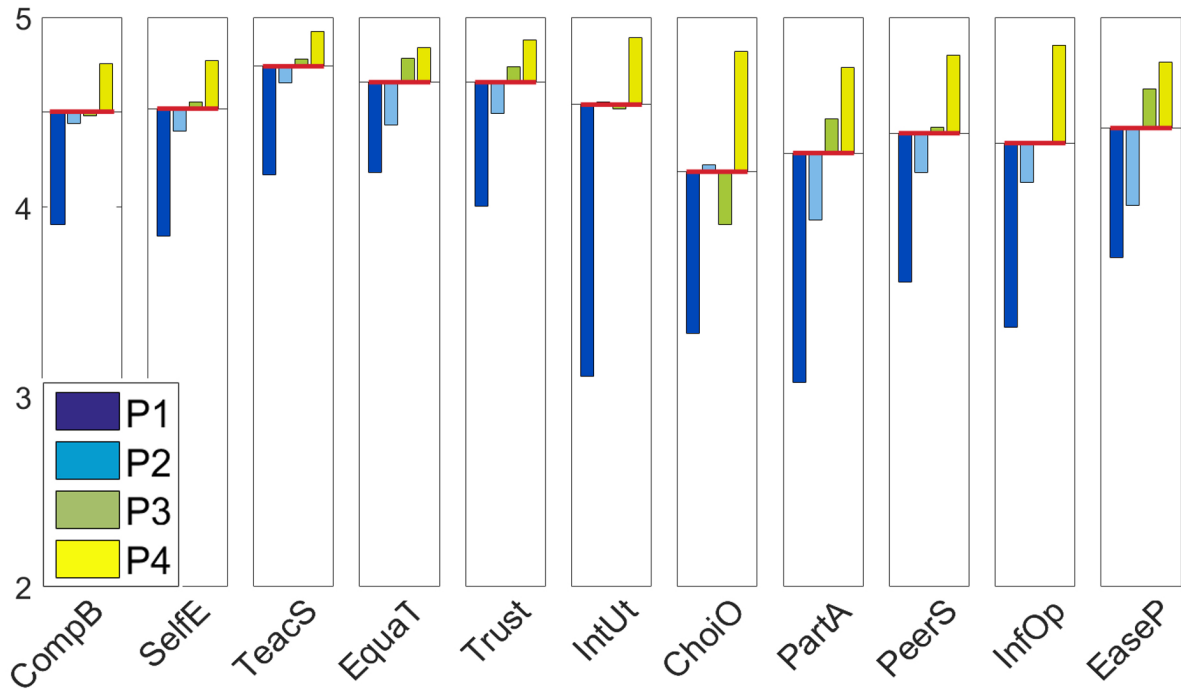


Figure 8: Deviations of the four agency profiles from the GAP in the Department of Teacher Education.

The first aim of this study was to introduce a conceptual and methodological basis for examining student agency. We used a multidimensional conceptualization of student agency, which consists of students' personal resources, participatory resources, and relational resources (Jääskelä et al., 2017b). Data were collected using a validated questionnaire instrument (Jääskelä et al., 2017a). This study adds to previous studies on agency by extending the focus beyond unitary dimensions and/or individual factors (e.g., epistemic agency, competence beliefs) (e.g., Damşa et al., 2010; Schunk and Zimmerman, 2012).

The second aim was to describe statistically robust educational data mining methods for analyzing the data on student agency. As argued in section 3.4, the small amount of data on the Likert scale, with possibly missing values, prevents the theoretically justified (Gaussian assumptions) use of classical second-order statistical methods. Therefore, non-parametric location and clustering methods from previous research in the field were applied in this study.

The third aim of this research was to depict a service-based architecture for supporting the provisioning of student agency analytics in practice. Learning analytics researchers and developers must address issues concerning ethics and privacy. Architectural choices (i.e., microservices) and pseudonymization of learner-

generated data are essential means of ethically processing educational data. Separating the data controller (e.g., the educational institution) and the data processor (the agency analytics service) by architectural means allows the controller to retain full control of personal data while still gaining the benefits of external analytics.

The fourth aim of this study was to examine the applicability of the proposed agency analytics process at the course level. Based on the analyses performed in two different courses, we can conclude that the proposed agency analytics process can be applied at the course level, and different profile groups can be identified. In the present empirical dataset, we found four agency profile groups in both courses.

While considering the profile groups of the two courses in a more detailed way, the following findings stand out: Both courses included a profile group of students who perceived their agency resources as higher than the general agency profile (GAP) in all dimensions of agency. In both courses, there was also a profile group of students who assessed their agency resources as lower than the GAP in all dimensions. These lower profile students might benefit from more tailored support. However, as the information provided to the teacher is supposed to be anonymous from the privacy point of view, the challenge for the teacher is how to recognize these students in the course. One option for

the teacher could be to provide students dialogic spaces (c.f., Lipponen and Kumpulainen, 2011) to reflect on the results.

Furthermore, different agency factors separated the identified profile groups in the two courses: In the computer programming course the factors were trust, self-efficacy, competence beliefs, and ease of participation. Whereas, in the course on educational sciences the factors were participation activity, ease of participation, and peer support. In the computer programming course the students received extensive study support. However, the students' main study method was still doing individual programming tasks, which required sufficient skills and knowledge. In this light, the emphasis on individual performance might explain that student-perceived self-efficacy and competence beliefs appeared to differentiate the profile groups. In the course on educational sciences, the factors related to the participatory resources might be explained by the fact that the students were expected to work in groups and actively participate in the thematic group discussions.

While considering the GAP levels and characteristics of the profile groups in the courses, we observed several features related to both courses in how the students perceived their agency resources. In the computer programming course, especially profile P3 is interesting, because P3 students' competence beliefs and their perceived self-efficacy appear as clearly lower than the GAP level. However, the same group of students assessed their participatory resources (especially opportunities to influence and participate, and getting peer support) near the GAP level or even higher. Furthermore, the P3 students succeeded generally better than other students in the course assessment and more often received grades of 4 or 5. The students in P3 might have benefited from the extensive support offered generally to all students in the course. However, the P3 students would need more individual support for recognizing their own strengths and competences as learners, and acquiring the self-confidence needed in future tasks.

In the course on educational sciences the GAP level was extremely high, indicating that most of the students perceived themselves as well resourced in the course. However, attention is drawn to the P1 students who experienced their participatory resources of agency as clearly lower than other students. The results indicate that the P1 students perceived their opportunities for participation, influencing, and making choices, as well as getting peer support, as lower than the GAP level. Furthermore, these P1 students did not fully find meaningfulness and utility value from the course content.

In the course on educational sciences there was one

interesting study group in which the teacher made an extra effort to implement agency-supportive pedagogy, e.g., by emphasizing the safe atmosphere, encouraging students, giving space for dialogue, maintaining a low threshold for participation, and handling the topic of agency with the students. This group of students belonged more often to the P4 profile with a high perception of their agency resources. This result raises an interest to study further, to what extent it is possible to influence students' experience of agency through pedagogy. In this case, it is not entirely clear to what extent stronger agency experiences resulted from the students' own increasing insight into the role of agency in their education and to what extent stronger agency experiences could be generated by the agency supportive-pedagogy. Our view is that students' cultivation through delivering knowledge of agency and increasing possibilities for their self-assessment of agency, and developing pedagogical practices supportive of agency construction are needed in university education.

6.1. Practical implications

In our analytics process, students receive their own agency profile in comparison to the general agency profile in the course and guidance on how to interpret the results. Teachers receive an analysis containing four different agency profiles in their course. The information about individual agency in comparison to the general agency profile in the course enables students to reflect and critically evaluate their personal learning experiences and their relationships between teachers, fellow students, and the learning environment.

We recommend that student agency analytics provides a tool for students' self-reflection, self-regulation, and academic advising, and for teachers' pedagogical development in higher education. In general, student self-regulation is an essential aim of learning analytics, and institutions should actively enable and encourage students to reflect on their learning and the related data (Greller and Drachsler, 2012). Students and teachers can benefit from learning analytics by self-reflecting on the effectiveness of their learning or teaching practices (Chatti et al., 2012). The visualization of student agency analytics results can be considered, what Baker (2010) calls the distillation of data for human judgment. This kind of analytics is a shift toward a deeper understanding of students' learning experiences in higher education (Viberg et al., 2018).

Another use of student agency analytics is to advance academic advising. The use of technology and data will shape the expectations and delivery of academic advising in higher education (Steele, 2018). Gavriushenko

et al. (2017) discuss the process of academic advising, which is cooperation between the adviser, student, and institution. It involves interactions with a curriculum, a pedagogy, and students' learning outcomes. They conclude that there is a need for personalized and automated academic advising. The AUS Scale concentrates on student-experienced resources of agency (e.g., for ownership and influence; Jääskelä et al. (2017a)), which are also important premises in academic advising. Thus, automated agency analytics could provide a starting point for discussions between the advisee and the advisor, and provide added value to the advising process. In student-centered learning analytics, students are co-interpreters of their own data (Kruse and Pongsa-japan, 2012). In our view, the educational institution enables the use of student agency analytics, and the results could be then interpreted in cooperation between student and advisor.

The last potential benefit we want to note relates to teachers' pedagogical knowledge. Analyzing student agency has the potential to benefit teachers' understanding of their students. Teachers' knowledge base can be divided into multiple categories, including general pedagogical knowledge and the knowledge of learners and their characteristics (Shulman, 1987). General pedagogical knowledge involves "broad principles and strategies of classroom management and organization that appear to transcend subject matter" (Shulman, 1987, p. 8). Further, general pedagogical knowledge can be considered "the knowledge needed to create and optimize teaching-learning situations across subjects," which includes knowledge about student heterogeneity (Voss et al., 2011, p. 953). Considering the definition and the purpose of learning analytics, which is to understand and optimize learning (Conole et al., 2011), it is reasonable to say that pedagogical knowledge and learning analytics have similar objectives. We propose that student agency analytics is one possible option for teachers to acquire information about their students. This information could then be used pedagogically to manage, organize, and optimize learning.

6.2. Limitations

The limitations of the study relate to the lack of previous research on the topic, a small sample size, a long survey instrument, and the selection of the number of profiles. To our knowledge, this is the first study utilizing unsupervised methods in analyzing student agency. Thus, there is very little previous work we can refer to. Furthermore, the present empirical data consisted of only two university courses. The AUS Scale questionnaire is relatively long, and this might have an effect on

the participants' response accuracy in some cases. The number of profiles is based on the CVIs and the knee point (Figure 5). A small number of factors was preferred for the sake of conciseness and easier interpretation from the practitioner point of view. The number of factors could be different in a different dataset. In addition, the results are based on quantitative analysis, and further mixed methods research is needed to validate the students' experiences of the perceived agency resources. Furthermore, in terms of studying the relation between agency experiences and grades, the link between the evaluation framework for grading and learning outcomes should be made explicit.

6.3. Future research

In the discussion, we provided some tentative suggestions for pedagogical use of the analytics process. We see that while students assess their resources of agency, it is primarily a question of student's self-regulation and learning about him-/herself as agentic learner. However, this assessment can be also seen as a reflection on the course implementation and support structures constructed through pedagogy. We intend to utilize the agency analytics process in several courses in the higher education context to obtain more data. One strategy for further research would be then to design an intervention study, which utilizes the individual student agency reports and teacher reports as interventions in a course setting.

7. Conclusion

This study contributes to the research on student agency in the higher education context using learning analytics methods based on unsupervised robust clustering. Furthermore, the study continues the discussion concerning the construct of student agency and offers the person-/subject-oriented approach by emphasizing the multidimensional nature of agency. We proposed and described a process of student agency analytics in a higher education context using a validated instrument, robust statistics, and service-based architecture. The purpose of this approach is to advance learners' commitment to learning by promoting their agentic awareness and informing pedagogical practices. Our demonstration of student agency analytics suggests that it is possible to obtain unique knowledge about the agency of university students using the AUS questionnaire and learning analytics methods described in the research. The findings showed that the proposed method could provide information about student agency at the course level.

Most notably, this is the first study, to our knowledge, to utilize learning analytics methods with a theoretical underpinning in systematically analyzing student agency. The potential of student agency analytics lies, for example, in the areas of students' self-regulation, academic advising, and teachers' pedagogical knowledge. This study was primarily concerned with depicting the overall process of student agency analytics. Although we acknowledge that further research is needed, student agency analytics could provide a bridge between effective learning analytics, students' agentic awareness, and teachers' pedagogical knowledge.

Appendix A. The Agency of the University Students (AUS) Scale

Abbreviated items of the Agency of the University Students (AUS) Scale in the order of dimensions.

- Competence beliefs

- 1 Understanding of the course contents.
- 2 Experiencing course contents as too challenging.^a
- 3 Sufficient basis for participation in discussions in the course.
- 4 Understanding of the constructs presented in the course.
- 5 Course demands have not been excessive.
- 6 Lacking basic knowledge for understanding the course contents.^a
- 7 Experience of a need for revision of basic concepts prior to the course.^a

- Self-efficacy

- 8 Belief in one's ability to succeed in the course.
- 9 Belief in succeeding even in the most challenging tasks.
- 10 Belief in successfully completing the course.
- 11 Confidence in oneself as a learner in spite of challenges.
- 12 Belief in attaining personal goals set for the course.

- Equal treatment

- 13 Equality among students.
- 14 Equal treatment of students by teachers.

- 15 Other students have a stronger influence on the course.^a

- Teacher support

- 16 Teachers' friendly attitude towards students.
- 17 Belittling of students by teachers.^a
- 18 Experience of being oppressed as a student.^a
- 19 Not enough room for discussion given by teachers.^a
- 20 Teachers' contemptuous attitude towards students.^a

- Trust

- 21 Safe course climate.
- 22 Experience of being welcome in the course.
- 23 Experience of being able to trust teachers.
- 24 Approachability of the teachers.
- 25 Possibility to be oneself in the course.
- 26 Experience of teachers' interest in students' viewpoints.
- 27 Encouraging students to participate in discussions.

- Participation activity

- 28 Taking responsibility by being an active participant.
- 29 Asking questions and making comments in the course.
- 30 Expressing opinions in the course.
- 31 Willingness to participate even when having other things to do.
- 32 Enjoyment in taking initiatives and collaborating in the course.

- Ease of participation

- 33 Ease of participation in discussions.
- 34 Difficulties participating in discussions.^a
- 35 Possibility to express thoughts and views without being ridiculed.
- 36 Courage to challenge matters presented in the course.

- Opportunities to influence

- 37 Student viewpoints were listened to.

- 38 Student viewpoints and opinions were taken into account.
 - 39 Experience of having to perform according to external instructions.^a
 - 40 No opportunities to influence the goals set for this course.^a
 - 41 Possibilities to influence the working methods.
 - 42 Opportunity to influence how competence is assessed in the course.
 - 43 No possibilities to influence the course contents.^a
- Opportunities to make choices
 - 44 No possibility to choose contents in line with the learning goals.^a
 - 45 Opportunity to choose course contents based on one's own interest.
 - 46 No possibility to choose between ways of completing the course.^a
 - Interest and utility value
 - 47 The course was not inspiring.^a
 - 48 The course was not inspiring because of unclear utility value.^a
 - 49 High motivation to study in the course.
 - 50 The contents of the course were interesting.
 - 51 Desire to learn in order to understand.
 - 52 Desire to succeed in the course.
 - 53 Maintaining persistence in the face of the high effort demanded.
 - Peer support
 - 54 Experiencing other students as resources for learning.
 - 55 Asking for help from other students when needed.
 - 56 Providing support for other students in challenging study tasks.
 - 57 No possibility to share competence with other group members.^a
 - 58 Opportunities to share competences in the group.

Note: ^a Reversed-coded item. The AUS Scale is copyrighted by the authors, its use requires written permission from the authors; contact information: paivikki.jaaskela@jyu.fi

References

- Archer, M.S., Archer, M.S., 2003. Structure, agency and the internal conversation. Cambridge University Press.
- Ayllón, S., Alsina, Á., Colomer, J., 2019. Teachers' involvement and students' self-efficacy: Keys to achievement in higher education. *PLoS One* 14, e0216865.
- Äyrämö, S., 2006. Knowledge Mining Using Robust Clustering. Ph.D. thesis. Faculty of Information Technology. University of Jyväskylä.
- Baker, R., 2010. Data mining, in: Peterson, P., Baker, E., McGaw, B. (Eds.), *International Encyclopedia of Education*. 3rd ed. ed., Elsevier Science, Amsterdam, pp. 112–118.
- Bandura, A., 1982. Self-efficacy mechanism in human agency. *The American Psychologist* 37, 122–147.
- Bandura, A., 1986. *Social Foundations of Thought and Action: A Social Cognitive Theory*. Prentice-Hall.
- Bandura, A., 2001. Social cognitive theory: An agentic perspective. *Annual review of psychology* 52, 1–26.
- Benson, V., Hand, C., Hartshorne, R., 2018. How compulsive use of social media affects performance: insights from the uk by purpose of use. *Behaviour & Information Technology* 38, 1–15.
- Biesta, G., Tedder, M., 2007. Agency and learning in the lifecourse: Towards an ecological perspective. *Studies in the Education of Adults* 39, 132–149.
- Billett, S., Fenwick, T., Somerville, M., Eteläpelto, A., Saarinen, J. (Eds.), 2006. *Work, subjectivity and learning: understanding learning through working life*. Springer, Dordrecht.
- Bond, M., Zawacki-Richter, O., Nichols, M., 2018. Revisiting five decades of educational technology research: A content and authorship analysis of the british journal of educational technology. *British journal of educational technology* 50, 12–63.
- Bransford, J., Vye, N., Stevens, R., Kuhl, P., Schwartz, D., Bell, P., Meltzoff, A., Barron, B., Pea, R.D., Reeves, B., 2006. Learning theories and education: Toward a decade of synergy, in: Alexander, P.A., Winne, P.H. (Eds.), *Handbook of educational psychology*. second ed., Lawrence Erlbaum Associates, Mahwah, NJ, pp. 209–244.
- Chatti, M.A., Dyckhoff, A.L., Schroeder, U., Thüs, H., 2012. A reference model for learning analytics. *International Journal of Technology Enhanced Learning* 4, 318–331.
- Chen, J., Shao, J., 2000. Nearest neighbor imputation for survey data. *Journal of Official Statistics* 16, 113–131.
- Conole, G., Gašević, D., Long, P., Siemens, G., 2011. Message from the LAK 2011 general & program chairs, in: *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*.
- Cord, A., Ambrose, C., Cocquerez, J.P., 2006. Feature selection in robust clustering based on laplace mixture. *Pattern Recognition Letters* 27, 627–635.
- Damsa, C.I., Kirschner, P.A., Andriessen, J.E.B., Erkens, G., Sins, P.H.M., 2010. Shared epistemic agency: An empirical study of an emergent construct. *Journal of the Learning Sciences* 19, 143–186.
- Dougherty, J., Kohavi, R., Sahami, M., 1995. Supervised and unsupervised discretization of continuous features, in: *Machine Learning Proceedings 1995*. Elsevier, pp. 194–202.
- Eccles, J.S., 2005. Subjective task value and the eccles et al. model of achievement-related choices, in: Elliot, A.J., Dweck, C.S. (Eds.), *Handbook of competence and motivation*. Guilford Press, New York, pp. 105–121.
- Emirbayer, M., Mische, A., 1998. What is agency? *American journal of sociology* 103, 962–1023.
- Eteläpelto, A., Lahti, J., 2008. The resources and obstacles of creative collaboration in a long-term learning community. *Thinking skills and creativity* 3, 226–240.
- Eteläpelto, A., Vähäsantanen, K., Hökkä, P., Paloniemi, S., 2013.

- What is agency? conceptualizing professional agency at work. *Educational Research Review* 10, 45–65.
- Everitt, B.S., 1992. The analysis of contingency tables. Chapman and Hall/CRC.
- Ferguson, R., 2012. Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning* 4, 304–317.
- Ferguson, R., Clow, D., 2017. Where is the evidence? a call to action for learning analytics, in: *LAK '17 Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, ACM, New York, USA. pp. 56–65.
- Ferguson, R., Hoel, T., Scheffel, M., Drachsler, H., 2016. Guest editorial: Ethics and privacy in learning analytics. *Journal of Learning Analytics* 3, 5–15.
- Foucault, M., 1975. *Discipline and Punish: The Birth of the Prison*. Random House.
- Gavriushenko, M., Saarela, M., Kärkkäinen, T., 2017. Towards evidence-based academic advising using learning analytics, in: *International Conference on Computer Supported Education*, Springer. pp. 44–65.
- Giddens, A., 1984. *The Constitution of Society*. University of California Press.
- Goller, M., Paloniemi, S., 2017. Agency at work, learning and professional development: An introduction, in: Goller, M., Paloniemi, S. (Eds.), *Agency at Work: An Agentic Perspective on Professional Learning and Development*. Springer International Publishing, Cham. pp. 1–14.
- Greeno, J.G., 1997. On claims that answer the wrong questions. *Educational Researcher* 26, 5–17.
- Greller, W., Drachsler, H., 2012. Translating learning into numbers: A generic framework for learning analytics. *Journal of Educational Technology & Society* 15, 42–57.
- Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. *Journal of machine learning research* 3, 1157–1182.
- Haggis, T., 2009. What have we been thinking of? a critical overview of 40 years of student learning research in higher education. *Studies in Higher Education* 34, 377–390.
- Hämäläinen, J., Jauhiainen, S., Kärkkäinen, T., 2017. Comparison of internal clustering validation indices for prototype-based clustering. *Algorithms* 10, 1–14.
- Harris, L.R., Brown, G.T., Dargusch, J., 2018. Not playing the game: student assessment resistance as a form of agency. *The Australian Educational Researcher* 45, 125–140.
- Hastie, T., Tibshirani, R., Friedman, J., 2009. *The elements of statistical learning: data mining, inference, and prediction*. Springer Series in Statistics. 2 ed., Springer-Verlag New York.
- Hettmansperger, T.P., McKean, J.W., 1998. *Robust nonparametric statistical methods*. Edward Arnold, London.
- Hökkä, P., Vähäsantanen, K., Mahlakaarto, S., 2017. Teacher educators' collective professional agency and identity – transforming marginality to strength. *Teaching and Teacher Education* 63, 36–46.
- Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. *Neural networks* 2, 359–366.
- Huber, P.J., 1981. *Robust Statistics*. John Wiley & Sons Inc., New York.
- ISO, 2017. *Health informatics — Pseudonymization*. ISO 25237. International Organization for Standardization. Geneva, Switzerland.
- Jääskelä, P., Poikkeus, A.M., Häkkinen, P., Vasalampi, K., Rasku-Puttonen, H., Tolvanen, A., 2019 submitted. Students' agency profiles in relation to perceived teaching practices in university courses.
- Jääskelä, P., Poikkeus, A.M., Vasalampi, K., Valleala, U.M., Rasku-Puttonen, H., 2017a. Assessing agency of university students: validation of the aus scale. *Studies in Higher Education* 42, 2061–2079.
- Jääskelä, P., Vasalampi, K., Häkkinen, P., Poikkeus, A.M., Rasku-Puttonen, H., 2017b. Students' agency experiences and perceived pedagogical quality in university courses. paper presented in the network of research in higher education. ECER 2017, 22 - 25 August, Copenhagen, Denmark.
- Jauhiainen, S., Kärkkäinen, T., 2017. A simple cluster validation index with maximal coverage, in: *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning - ESANN 2017*, ESANN. pp. 293–298.
- Jena, R., 2018. Predicting students' learning style using learning analytics: a case study of business management students from india. *Behaviour & Information Technology*, 1–15.
- John, G.H., Kohavi, R., Pfleger, K., 1994. Irrelevant features and the subset selection problem, in: *Proceedings of the 11th Conference on Machine Learning*, Elsevier. pp. 121–129.
- Juutilainen, M., Metsäpelto, R.L., Poikkeus, A.M., 2018. Becoming agentic teachers: Experiences of the home group approach as a resource for supporting teacher students' agency. *Teaching and Teacher Education* 76, 116–125.
- Kärkkäinen, T., 2002. MLP-network in a layer-wise form with applications to weight decay. *Neural Computation* 14, 1451–1480.
- Kärkkäinen, T., 2014. On cross-validation for MLP model evaluation, in: *Structural, Syntactic, and Statistical Pattern Recognition*. Springer-Verlag. Lecture Notes in Computer Science (8621), pp. 291–300.
- Kärkkäinen, T., 2015. Assessment of feature saliency of MLP using analytic sensitivity, in: *Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning - ESANN 2015*, pp. 273–278.
- Kärkkäinen, T., Äyrämö, S., 2005. On computation of spatial median for robust data mining, in: Schilling, R., Haase, W., Periaux, J., Baier, H. (Eds.), *Proceedings of Sixth Conference on Evolutionary and Deterministic Methods for Design, Optimisation and Control with Applications to Industrial and Societal Problems (EUROGEN 2005)*.
- Kärkkäinen, T., Heikkola, E., 2004. Robust formulations for training multilayer perceptrons. *Neural Computation* 16, 837–862.
- Klemenčič, M., 2017. From student engagement to student agency: Conceptual considerations of european policies on Student-Centered learning in higher education. *Higher Education Policy* 30, 69–85.
- Kruse, A., Pongsajapan, R., 2012. Student-centered learning analytics. *CNDLS Thought Papers*, 1–9.
- Kruskal, W.H., Wallis, W.A., 1952. Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association* 47, 583–621.
- Lave, J., Wenger, E., 1991. *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press, Cambridge.
- Lewis, J., Fowler, M., 2014. Microservices: A definition of this new architectural term. URL: <https://web.archive.org/web/20181202010248/https://martinfowler.com/articles/microservices.html>.
- Lindgren, R., McDaniel, R., 2012. Transforming online learning through narrative and student agency. *Journal of Educational Technology & Society* 15, 344–355.
- Lipponen, L., Kumpulainen, K., 2011. Acting as accountable authors: Creating interactional spaces for agency work in teacher education. *Teaching and Teacher Education* 27, 812–819.
- Liu, H., Motoda, H., 2012. Feature selection for knowledge discovery and data mining. volume 454. Springer Science & Business Media.
- Malmberg, L.E., Hagger, H., 2009. Changes in student teachers' agency beliefs during a teacher education year, and relationships with observed classroom quality, and day-to-day experiences. *The*

- British journal of educational psychology 79, 677–694.
- Martin, J., 2004. Self-regulated learning, social cognitive theory, and agency. *Educational psychologist* 39, 135–145.
- Moreno-Torres, J.G., Sáez, J.A., Herrera, F., 2012. Study on the impact of partition-induced dataset shift on k -fold cross-validation. *IEEE Transactions on Neural Networks and Learning Systems* 23, 1304–1312.
- Namiot, D., Sneps-Sneppé, M., 2014. On micro-services architecture. *International Journal of Open Information Technologies* 2, 24–27.
- Newman, S., 2015. Building microservices: designing fine-grained systems. O'Reilly Media, Inc.
- Niemelä, M., Äyrämö, S., Kärkkäinen, T., 2018. Comparison of cluster validation indices with missing data, in: *Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning - ESAINN 2018*, pp. 461–466.
- O'Meara, K., Eliason, J., Cowdery, K., Jaeger, A., Grantham, A., Mitchall, A., Zhang, K., 2014. By design: How departments influence graduate student agency in career advancement. *International Journal of Doctoral Studies* 9, 155–180.
- Packer, M.J., Goicoechea, J., 2000. Sociocultural and constructivist theories of learning: Ontology, not just epistemology. *Educational psychologist* 35, 227–241.
- Pardo, A., Siemens, G., 2014. Ethical and privacy principles for learning analytics. *British Journal of Educational Technology* 45, 438–450.
- Park, J., Sandberg, I.W., 1991. Universal approximation using radial-basis-function networks. *Neural computation* 3, 246–257.
- Prinsloo, P., Slade, S., 2016. Student vulnerability, agency and learning analytics: An exploration. *Journal of Learning Analytics* 3, 159–182.
- Redecker, C., Johannessen, Ø., 2013. Changing assessment—towards a new assessment paradigm using ICT. *European Journal of Education* 48, 79–96.
- Regulation [EU] 2016/679, 2016. Regulation (EU) 2016/679 of the European parliament and of the council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/EC (general data protection regulation). *Official Journal of the European Union* L119, 1–88.
- Rosé, C.P., 2018. Learning analytics in the learning sciences, in: Fischer, F., Hmelo-Silver, C.E., Goldman, S.R., Reimann, P. (Eds.), *International Handbook of the Learning Sciences*. Routledge, New York, pp. 511–519.
- Ruohotie-Lyhty, M., Moate, J., 2015. Proactive and reactive dimensions of life-course agency: mapping student teachers' language learning experiences. *Language and education* 29, 46–61.
- Ryan, R.M., Deci, E.L., 2000. Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology* 25, 54–67.
- Saarela, M., 2017. Automatic knowledge discovery from sparse and large-scale educational data: case Finland. Ph.D. thesis. Faculty of Information Technology. University of Jyväskylä.
- Saarela, M., Hämäläinen, J., Kärkkäinen, T., 2017. Feature ranking of large, robust, and weighted clustering result, in: *Proceedings of 21st Pacific Asia Conference on Knowledge Discovery and Data Mining - PAKDD 2017*, pp. 96–109.
- Saarela, M., Kärkkäinen, T., 2015. Analyzing student performance using sparse data of core bachelor courses. *Journal of Educational Data Mining* 7, 3–32.
- Saarela, M., Kärkkäinen, T., 2017. Knowledge discovery from the programme for international student assessment, in: *Learning Analytics: Fundamentals, Applications, and Trends*. Springer. Volume 94 of the series *Studies in Systems, Decision and Control*, pp. 229–267.
- Scardamalia, M., 2002. Collective cognitive responsibility for the advancement of knowledge, in: Smith, B. (Ed.), *Liberal Education in a Knowledge Society*. Open Court, Chicago, IL, pp. 67–98.
- Scardamalia, M., Bereiter, C., 1991. Higher levels of agency for children in knowledge building: A challenge for the design of new knowledge media. *Journal of the Learning Sciences* 1, 37–68.
- Schunk, D.H., Zimmerman, B.J., 2012. Competence and control beliefs: Distinguishing the means and ends, in: Alexander, P.A., Winne, P.H. (Eds.), *Handbook of educational psychology*. Routledge, New York, pp. 349–367.
- Seifert, T., 2004. Understanding student motivation. *Educational research* 46, 137–149.
- Shulman, L., 1987. Knowledge and teaching: Foundations of the new reform. *Harvard Education Review* 57, 1–23.
- Siemens, G., 2013. Learning analytics: The emergence of a discipline. *American Behavioral Scientist* 57, 1380–1400.
- Siemens, G., Baker, R.S., 2012. Learning analytics and educational data mining: towards communication and collaboration, in: *Proceedings of the 2nd international conference on learning analytics and knowledge*. ACM. pp. 252–254.
- Starkey, L., 2017. Three dimensions of student-centred education: a framework for policy and practice. *Critical Studies in Education* 60, 1–16.
- Steele, G.E., 2018. Student success: Academic advising, student learning data, and technology. *New Directions for Higher Education* 2018, 59–68.
- Su, Y.H., 2011. The constitution of agency in developing lifelong learning ability: The 'being' mode. *Higher Education* 62, 399–412.
- Thomas, J.W., 1980. Agency and achievement: Self-management and self-regard. *Review of Educational Research* 50, 213–240.
- Thorndike, R.L., 1953. Who belongs in the family? *Psychometrika* 18, 267–276.
- Tuhkala, A., Nieminen, P., Kärkkäinen, T., 2018. Semi-automatic literature mapping of participatory design studies 2006–2016, in: *Proceedings of the 15th Participatory Design Conference (PDC 2018)*. ACM. pp. 1–6.
- Vähäsantanen, K., 2015. Professional agency in the stream of change: Understanding educational change and teachers' professional identities. *Teaching and Teacher Education* 47, 1–12.
- Van Dinther, M., Dochy, F., Segers, M., 2011. Factors affecting students' self-efficacy in higher education. *Educational research review* 6, 95–108.
- Viberg, O., Hatakka, M., Bälter, O., Mavroudi, A., 2018. The current landscape of learning analytics in higher education. *Computers in Human Behavior* 89, 98–110.
- Voss, T., Kunter, M., Baumert, J., 2011. Assessing teacher candidates' general pedagogical/psychological knowledge: Test construction and validation. *Journal of Educational Psychology* 103, 952.
- Waheed, H., Hassan, S.U., Aljohani, N.R., Wasif, M., 2018. A bibliometric perspective of learning analytics research landscape. *Behaviour & Information Technology* 37, 1–17.
- Wise, A.F., 2014. Designing pedagogical interventions to support student use of learning analytics, in: *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge*. ACM, New York, NY, USA. pp. 203–211.
- Wise, A.F., Shaffer, D.W., 2015. Why theory matters more than ever in the age of big data. *Journal of Learning Analytics* 2, 5–13.
- Zhang, J., Zhang, X., Jiang, S., Ordóñez de Pablos, P., Sun, Y., 2018. Mapping the study of learning analytics in higher education. *Behaviour & Information Technology* 37, 1142–1155.
- Zimmerman, B.J., Pons, M.M., 1986. Development of a structured interview for assessing student use of self-regulated learning strategies. *American educational research journal* 23, 614–628.