Understanding the Study Experiences of Students in Low Agency Profile: Towards a Smart Education Approach

Ville Heilala[1], Päivikki Jääskelä[2], Tommi Kärkkäinen[1], and Mirka Saarela[1]

[1] University of Jyvaskyla, Faculty of Information Technology
P.O. Box 35, FI-40014, University of Jyvaskyla, Finland
ville.s.heilala@jyu.fi,
[2] Finnish Institute for Educational Research, P.O. Box 35, FI-40014, University of Jyvaskyla, Finland

Abstract. In this paper, we use student agency analytics to examine how university students who assessed to have low agency resources describe their study experiences. Students ($n = 292$) completed the Agency of University Students (AUS) questionnaire. Furthermore, they reported what kinds of restrictions they experienced during the university course they attended. Four different agency profiles were identified using robust clustering. We then conducted a thematic analysis of the open-ended answers of students who assessed to have low agency resources. Issues relating to competence beliefs, self-efficacy, student-teacher relations, time as a resource, student well-being, and course contents seemed to be restrictive factors among the students in the low agency profile. The results could provide guidelines for designing systems for smart education.

Keywords: student agency analytics, learning analytics, robust clustering, thematic analysis, knowledge graph

1 Introduction

Digitalization, increased computational power, and advances in data storage have led to vast amounts of data collected from educational domains [5]. It is envisioned that it will soon be possible to store and assess the learning behaviors and outspread the educational history of individual students [4]. Extracting knowledge from these enormous quantities of data and leveraging them to improve education require "smartness" that integrates technology with educational domain knowledge and pedagogical theories.

Learning analytics is a research discipline that emerged with the growing availability of educational data and the demand for understanding these data. It bridges the interface of these large educational datasets and computational visualization and analysis methods for communicating meaningful and actionable patterns that assist individuals in decision making about teaching and learning [17, 19]. Thus, learning analytics provides one viable option to embed smartness into systems of the educational domain.

Smart education — an emergent concept — is currently taking a form under continuous multidisciplinary discussion and there already exists several attempts to define and characterize it [22]. A research framework developed in [22] presents that smart education consists of three elements: smart learners, smart pedagogies, and smart learning environments. Smart learners possess relevant competence: a specific set of skills and knowledge to succeed in modern society. Smart pedagogies take into account the needs of different learners using four different instructional strategies: class-based differentiated instruction, group-based collaborative learning, individual-based personalized learning, and mass-based generative learning. Smart learning environments provide engaging, intelligent, and scalable possibilities for education. In general, the purpose of smart education "is to improve learner's quality of life long learning" [22, p. 15].

Student agency is a multidimensional concept that describes important constituents of intentional and purposeful learning; it emphasizes students' experienced opportunities to influence their learning and their perceptions regarding their capacity to learn in the complex and dynamic learning situations [10, 9]. The data is collected using validated Agency of University Students (AUS) Scale measuring students' experiences of their agency in three resource domains and their respective factors: personal domain (2 factors; Competence beliefs and Self-efficacy); relational domain (3 factors; Equal treatment, Teacher support and Trust); and participatory domain (6 factors; Participation activity, Ease of participation, Opportunities to influence, Opportunities to make choices, Interest and utility value, and Peer support) [10, 11]. The AUS domains and factors assess learners, pedagogical arrangements, and learning environment being, thus, linked to the three constituents of smart education.

In the previous study utilizing learning analytics [9], we applied robust statistics and machine learning to questionnaire data on student agency, with calling this analyzing process as student agency analytics. This article focuses on the experiences of those university students who assessed to have low agency resources. The following research question was set: What kinds of restrictions do the students in the low agency profile experience in the courses they have attended? Besides answering the research question, we also aim to exemplify how student agency analytics relates to smart educational systems in general.

2 Materials and Methods

The research data consist of online questionnaire responses of 292 first and second-year students in three faculties from the University of Jyväskylä, Finland. The data were collected using the AUS Scale [10, 11] consisting of 58 items in a five-point Likert scale (1 = fully disagree, 2 = partly disagree, 3 = neither agree nor disagree, 4 = partly agree and 5 = fully agree). Higher scores on the Likert scale indicated higher levels of agency. Also, the students were given an opportunity to describe their experiences in the course with a few open-ended questions.

We analyzed the data using a mixed-methods approach where we first used robust clustering for deriving student agency profiles and then conducted a qualitative thematic analysis on a selected subset of open-ended question data. All pre-processing, data analysis, and visualization was performed in *Python* 3.7.1 using *Pandas*, *Numpy*, *Matplotlib* and *Seaborn* libraries, except imputation of missing data was done in *R* using *testing* package implementing method described in [12]. Clustering was done using a custom script based on the work done in [8].

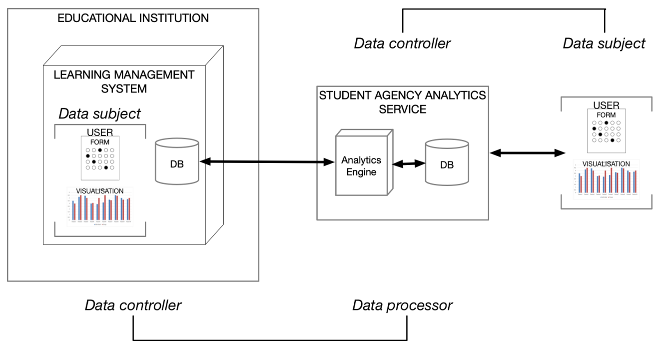

Fig. 1. Agency analytics service is either data controller or data processor depending on the use case.

Part of the data was collected using the Webropol questionnaire tool, and in two courses, we used a questionnaire tool included in our agency analytics service (Fig. 1). The service as a whole is under development, and the main aim of this service is to be able to separate the data controller and data processor [18]. Agency analytics service can be used directly in a web browser, or an educational institution could integrate it into a learning management system. In the latter case, the institution can possess the data in its database and send only the minimum required amount of pseudonymized data to analytics service for processing.

The collected dataset contained missing Likert values (1.43 %). These data were missing at random (MAR) [13] and imputed using the k-nearest neighbor (k-NN) method described in [12]. Inverted questionnaire items were inverted using linear scaling. Factor pattern matrix of the AUS questionnaire factor model was used to calculate the individual student agency factors. The agency factors were scaled to represent the original AUS questionnaire Likert scale from 1 to 5. These factors were then clustered into four student agency profiles (P1-4 in Fig. 2, 4, and 5). The clustering provides the prototype students of each cluster and assigns individual students to these clusters. Clustering was based on a k-means algorithm with the spatial median as a distance measure [2]. A more detailed description of the analysis process is depicted in [9].

In addition to the AUS questionnaire, the students were asked to answer open-ended questions to get more detailed knowledge of their study experiences.

In this paper, we concentrate on analyzing the student-experienced restrictions in their courses about which they wrote in their responses to one particular question, which was: Which factors in this course do you think hindered or limited your learning? The responses were then analyzed using a thematic analysis [3].

Rigorous thematic analysis is a time-consuming research method. Thus, clustering and assigning students' open-ended answers into corresponding profiles helped us to concentrate on an interesting subset (students in the low agency profile) of responses in the thematic analysis. For conducting the thematic analysis for the open-ended answers, we used a procedure by Braun and Clarke [3] consisting of six phases. This approach is argued to be well suited also for educational data [14]. The analysis concentrates on semantic layer [3] of the student answers to find out how they describe their study experiences.

The thematic analysis was performed by the first two authors, both having degrees in the field of education and extensive teaching experience. Intercoder reliability [6] was not formally assessed as the analysis involved the generation of the initial coding. Instead, the analysis was based on the researchers' independent work followed by in-depth discussions and negotiations of the final interpretations several times during the analyzing phases to meet intercoder agreement [7]. By providing the outline of the thematic analysis process, quotations when applicable (quotations have been translated from Finnish by the first author), the explanation of the key codes, and the final thematic map (Fig. 3), we aim to provide evidence for the reader to assess the dependability of our research.

3 Results

Based on our previous research on student agency [9], the individual student agency factors were clustered into four profiles. Fig. 2 presents the general agency profile (GAP) of all students and the deviations of the four individual profiles (P1-P4) from GAP. The different profiles P1-P4 depict the prototype students in each profile. The profile P1 is considered as the low agency profile. As can be seen from the Fig. 2, the students in this profile have lower values in all AUS factors. In particular, they are characterized by weak competence beliefs and self-efficacy as learners. P4, on the other hand, is called the high agency profile. The students in this profile generally have high values in most of the AUS factors. Notably, the students in P4 perceived that they had been treated equally in the study group, and they experienced teacher as more supportive when comparing to students in other profiles.

The low agency profile P1 consisted of 42 students, and 41 of them had answered the open-ended question about their learning restrictions. Fig. 3 presents the thematic map of the themes and respective codings we have derived from the answers of students belonging to the P1 profile. The themes (e.g., competence beliefs, time as a resource) are denoted inside circles surrounded by the codes (e.g., difficult contents, personal obligations) relating to a particular theme. The size of the code represents the number of times the code has occurred in the data. For example, the code *fast instruction pace* occurred more times than the

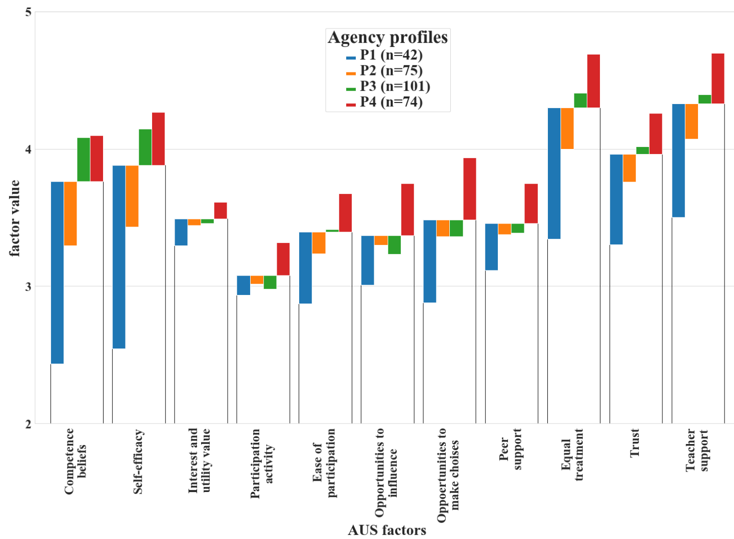

Fig. 2. General agency profile (GAP) and individual profile deviations (P1-P4). Factor values represent the original questionnaire Likert scale.

code *lack of understanding*. The links between codes denote that the students have mentioned them in relation to each other in their answers. Many more links could have been added based on common sense. However, in our analysis, the codes were linked only if the student has explicitly stated them to be interlinked. For example, one student brought out that "overlapping studies ... hard to focus on many things at the same time"; thus, the codes *overlapping studies* and *difficult to concentrate* are linked together.

Next, we describe the results of the thematic analysis and their links to student agency profiles. In the P1 group, the students brought out issues that mostly linked to personal and relational resources of agency. Students in P1 reported having low agency primarily in the factors of competency beliefs, self-efficacy, and in all factors representing the relational resources of agency. These results will be discussed in detail as follows. Furthermore, three other significant themes — time as a resource, student well-being, and course contents — will be elaborated.

Personal resources of agency. As illustrated in Fig. 2, the students in the low agency profile P1 reported lower competence beliefs and self-efficacy when comparing to the students in other profiles. Some students reported even lower values than 2 (partly disagree) (Fig. 4) for both aforementioned dimensions. Low competence beliefs refer, for example, to student-experienced lack of understanding of the course contents, the lack of basic knowledge, and experiences of the course contents as too challenging, while low self-efficacy refers to students' beliefs in not succeeding well in the course and tasks [10].

In their open-ended answers, students in the P1 reported negative past experiences and negative perceptions as a learner. Furthermore, students in the

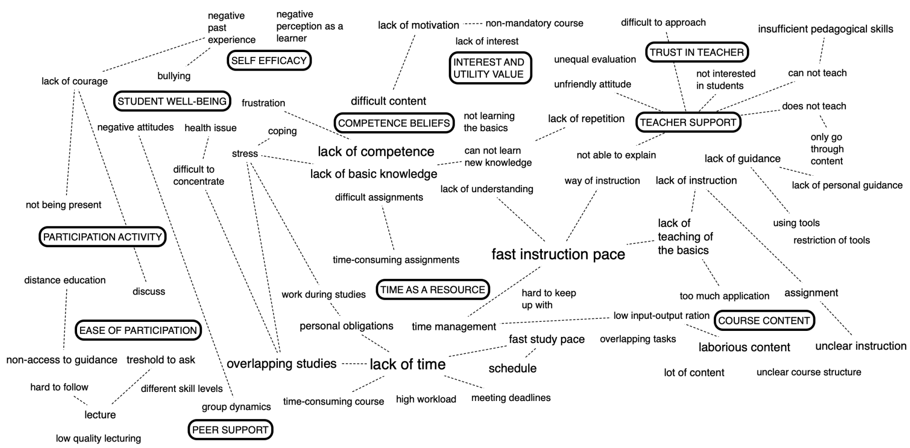

Fig. 3. The results of the thematic analysis: the thematic map and coding of the data in the low student agency profile (P1).

P1 reported lack of competence and lack of basic knowledge and understanding (e.g., "The lack of basic knowledge has been the biggest challenge"). They perceived the course contents as too difficult, and they felt that they could not learn the basics or embrace new knowledge. Some students reported that their experienced lack of competence also led to frustration and stress (e.g., "If I don't know things I get easily frustrated with the course tasks", "If there is no basic knowledge, it is difficult to build new knowledge, and it only creates unnecessary stress").

Relational resources of agency. Even though the GAP (Fig. 2) showed rather high values (near to or over 4) in the factors representing relational resources — equal treatment, teacher support and trust in teacher —, the students in the low agency profile P1 reported somewhat more moderate values. Some students reported values lower than 3 (Fig. 5), which indicates they experience relational resources as less supportive than the students in other profiles. Equal treatment depicts the equality between students and equal treatment of students by the teacher, while trust in teacher and teacher support reflect the attitudes and supportive actions of a teacher [10].

The thematic analysis supported the findings derived from the aforementioned quantitative analysis. The students in P1 reported the teacher being uninterested in students and being difficult to approach (e.g., "A couple of teachers do not seem to be interested in students and they are very difficult to approach with questions"), having an unfriendly attitude, and possessing pedagogical shortcomings (e.g., "I don't see the teacher has the pedagogical skills to teach us who are new to this topic").

Time as a resource. It is argued that student's time is the most precious resource [1]. The argument is supported by our thematic analysis, where the lack

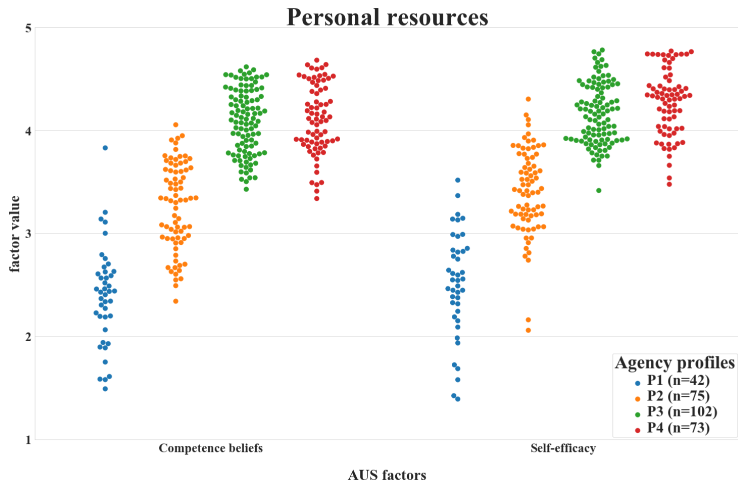

Fig. 4. The personal resources of student agency. Factor values represent the original questionnaire Likert scale.

of time was one of the most cited restrictive aspects of studying. Lack of time was mentioned due to personal obligations (e.g., working during the studies) or issues relating to studying (e.g., high workload). It was also associated with overlapping studies, as some students had many courses going on at the same time. Overlapping studies might be the result of personal choices or curriculum schedule. One major issue in P1 was the experienced fast instruction pace in the course, which was mentioned as, for example, "fast progression" or stating that "new things come at a great pace". To sum up, time seemed to be a complex resource in our material, and its importance depends on the student's situation.

Student well-being. According to a concise definition, student well-being is "a sustainable state of positive mood and attitude, resilience, and satisfaction with self, relationships, and experiences at school" [16, p. 7]. The students in P1 mentioned in their answers several aspects, which we interpreted belonging into a student well-being theme. The students reported, for example, difficulties to concentrate on studying, negative past experiences (e.g., bullying) and stress. Furthermore, the experienced stress was mentioned to be related to overlapping studies and lack of basic knowledge.

Course contents. The students mentioned limitations relating to the contents of the course. A few students complained about the unclear instruction and structure of the course, which was mentioned to be related to the lack of teacher support (e.g., lack of instruction). One interesting point was that some students experienced a low input-output ratio in the course. They felt that even if they work hard, it does not affect the outcome of the course. For example, one student commented that there had been no direct connection between course results and time used.

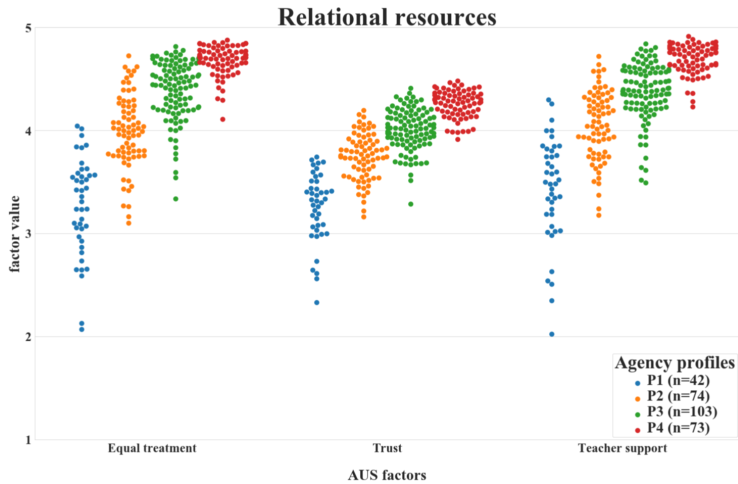

Fig. 5. The relational resources of student agency. Factor values represent the original questionnaire Likert scale.

4 Discussion and Conclusion

Student agency analytics can be considered to support smart education. It utilizes the approach of learning analytics to provide knowledge, which can be used to promote better learning. Moreover, it could be used to help learners to acquire skills they need in a modern and rapidly changing world; help them to become smart learners. By utilizing student agency analytics as a service, it can be embedded into existing learning environments to enhance their smart capabilities.

In terms of quality of education, it is essential to take attention to the students having low agency. They might be unable to benefit from the education in their competence development, or they might otherwise be at risk of "falling behind". Identification of those students is possible by using a validated questionnaire and appropriate learning analytics methods. Also, our qualitative thematic analysis revealed different experiences, which hindered and limited learning among the students belonging to the low agency profile. By using a mixed-methods approach, it is possible to acquire more in-depth insight into students' study experiences.

By identifying the students' different experiences of agency, instructors can provide more personalized support. Especially, meaningful contacts with academic staff are important and recognized in the research literature. For example, students interviewed in [20] found contacts with lecturers problematic because of lecturers being remote, inaccessible, and unable to communicate academic expectations. Same problems were experienced by the low agency students in our analysis. Students reported teachers to be difficult to approach, and they did not get enough instruction and guidance. Also, students in the low agency profile reported lack of competence and lack of basic knowledge. Thus, in the low agency profile, there is an inherent need for support and experienced a lack of support at the same time.

Furthermore, our thematic analysis revealed that the reasons for students mentioning lack of time as a restrictive aspect are manifold. It might not be sufficient to track the time a student has used in a virtual learning environment. Nor it would be "smart" education to send automatic reminders to students, for example, to watch course videos, if they have problems with time management, competence, or well-being. Instead, it would be essential to know, for instance, *why* the student does not have enough time to study or *what* aspect in student's competence is restricting them from learning new.

Providing personalized support for students using smart technologies in education requires that systems must be able to extract and distill the learners' different experiences into useful information. Educators can utilize the information to make pedagogical decisions. The outcome of our thematic analysis is a thematic map (Fig. 3), which starts to resemble and form a knowledge graph. A knowledge graph is a general framework for presenting entities and their relationships [21]. The student-reported restrictions can be seen as nodes and their reported relations as edges in the knowledge graph. From the semantic point of view, many words students use to describe their study experiences are so-called suitcase words [15], which have multiple meanings attached to them. It could be possible to depict these meanings as a knowledge graph. This possibility is the leading idea of our future work as we aim to develop automated handling of open-ended student feedback. Such a system could allow us to process and utilize student feedback at a larger scale. Our thematic analysis contained a limited amount of student answers. Thus, further research is needed to gain more understanding of the learners' experiences in different student agency profiles.

The present study contributes to the discussion of how learning analytics and smart technologies in education can be utilized to benefit the learners as well as educators. We used mixed-methods to analyze university students' agency and study experiences among the students belonging to the low agency profile. In our research data, especially issues related to competence beliefs, self-efficacy, student-teacher relations, time as a resource, student well-being, and course contents were identified as restrictive factors among the students in the low agency profile. To conclude, the "smartness" in education could mean, for example, providing relevant and timely knowledge about the students' individual study experiences for the basis of pedagogical and institutional decision-making.

References

1. Astin, A.W.: Student involvement: A developmental theory for higher education. Journal of College Student Development **40**(5), 518–529 (1999). Original work published 1984.
2. Äyrämö, S.: Knowledge Mining Using Robust Clustering, *Jyväskylä Studies in Computing*, vol. 63. University of Jyväskylä (2006)
3. Braun, V., Clarke, V.: Using thematic analysis in psychology. Qualitative research in psychology **3**(2), 77–101 (2006)

4. Brinkhuis, M., Savi, A., Hofman, A., Coomans, F., van der Maas, H., Maris, G.: Learning as it happens: A decade of analyzing and shaping a Large-Scale online learning system. Journal of Learning Analytics **5**(2), 29–46. (2018)
5. Buder, J., Hesse, F.W.: Informational Environments: Effects of Use, Effective Designs. Springer (2017)
6. Cho, Y.I.: Intercoder reliability. In: P.J. Lavrakas (ed.) Encyclopedia of Survey Research Methods. Sage Publications, Thousand Oaks, CA (2008)
7. Guest, G., MacQueen, K.M., Namey, E.E.: Applied thematic analysis. Sage Publications, Thousand Oaks, CA (2012)
8. Hämäläinen, J., Kärkkäinen, T., Rossi, T.: Scalable robust clustering method for large and sparse data. In: European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. ESANN (2018)
9. Jääskelä, P., Heilala, V., Kärkkäinen, T., Häkkinen, P.: Student agency analytics (2019, in review)
10. Jääskelä, P., Poikkeus, A.M., Vasalampi, K., Valleala, U.M., Rasku-Puttonen, H.: Assessing agency of university students: validation of the aus scale. Studies in Higher Education **42**(11), 2061–2079 (2017)
11. Jääskelä, P., Vasalampi, K., Häkkinen, P., Poikkeus, A.M., Rasku-Puttonen, H.: Students' agency experiences and perceived pedagogical quality in university courses. paper presented in the network of research in higher education (2017). ECER 2017, 22 - 25 August, Copenhagen, Denmark
12. Jönsson, P., Wohlin, C.: Benchmarking k-nearest neighbour imputation with homogeneous likert data. Empirical Software Engineering **11**(3), 463 (2006)
13. Little, R.J., Rubin, D.B.: Statistical analysis with missing data, vol. 793. Wiley (2014)
14. Maguire, M., Delahunt, B.: Doing a thematic analysis: A practical, step-by-step guide for learning and teaching scholars. AISHE-J: The All Ireland Journal of Teaching and Learning in Higher Education **9**(3) (2017)
15. Minsky, M.: The emotion machine: Commonsense thinking, artificial intelligence, and the future of the human mind. Simon and Schuster (2007)
16. Noble, T., Wyatt, T., McGrath, H., Roffey, S., Rowling, L.: Scoping study into approaches to student wellbeing: Final report. Tech. Rep. PRN 18219, Australian Catholic University and Erebus International (2008)
17. Pardo, A., Teasley, S.: Learning analytics research, theory and practice: widening the discipline. Journal of Learning Analytics **1**(3), 4–6 (2014)
18. Regulation [EU] 2016/679: 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 (2016)
19. Saarela, M., Kärkkäinen, T.: Knowledge Discovery from the Programme for International Student Assessment. In: A. Peña-Ayala (ed.) Learning Analytics: Fundaments, Applications, and Trends: A View of the Current State of the Art to Enhance e-Learning, pp. 229–267. Springer International Publishing, Cham (2017)
20. Scanlon, L., Rowling, L., Weber, Z.: 'you don't have like an identity ... you are just lost in a crowd': Forming a student identity in the first-year transition to university. Journal of Youth Studies **10**(2), 223–241 (2007)
21. Yan, J., Wang, C., Cheng, W., Gao, M., Zhou, A.: A retrospective of knowledge graphs. Front. Comput. Sci. **12**(1), 55–74 (2018)
22. Zhu, Z.T., Yu, M.H., Riezebos, P.: A research framework of smart education. Smart learning environments **3**(1), 4 (2016)