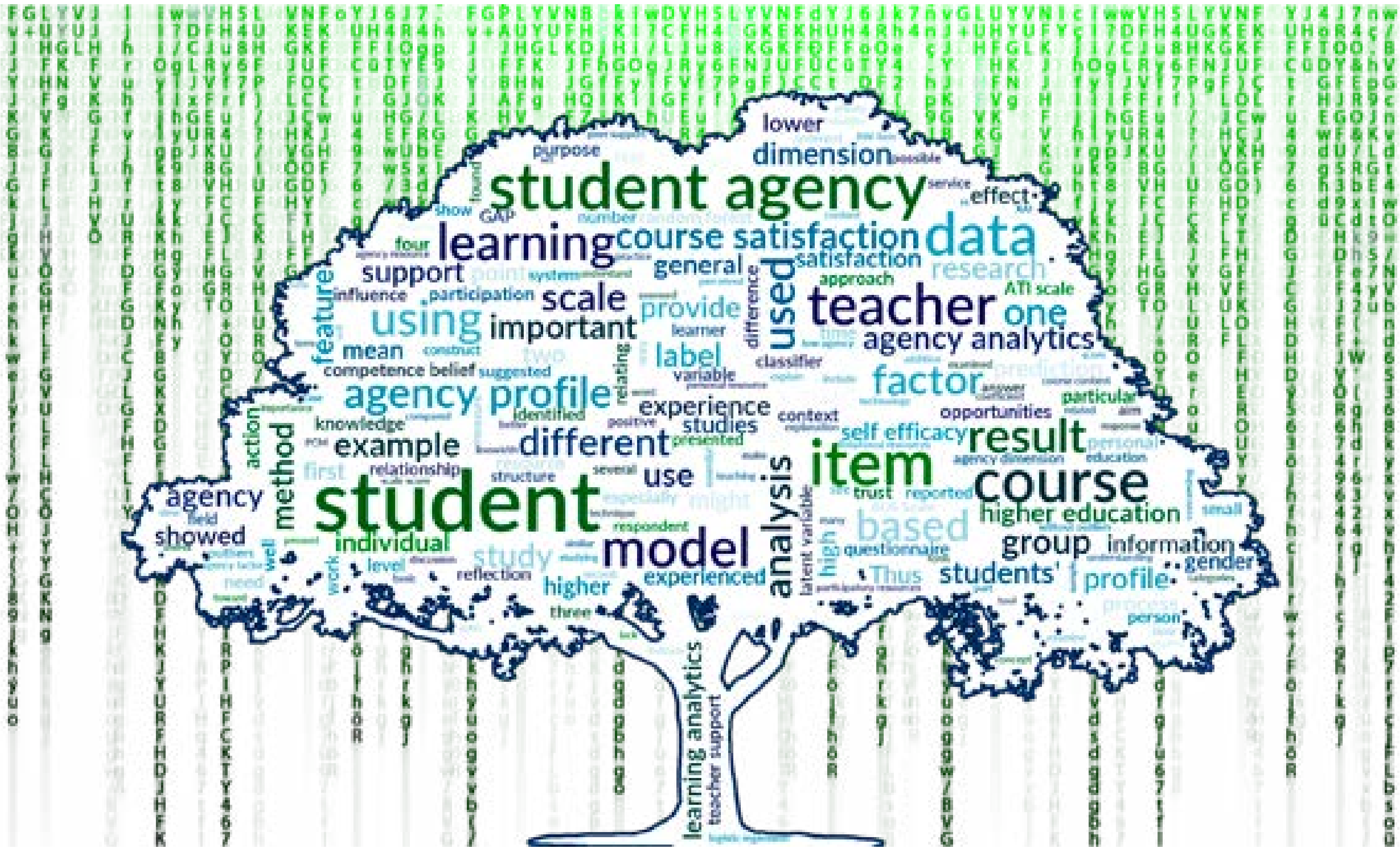
JYU DISSERTATIONS 512

Ville Heilala

Learning Analytics with Learning and Analytics

Advancing Student Agency Analytics



UNIVERSITY OF JYVÄSKYLÄ
FACULTY OF INFORMATION TECHNOLOGY

JYU DISSERTATIONS 512

Ville Heilala

Learning Analytics with Learning and Analytics

Advancing Student Agency Analytics

Esitetään Jyväskylän yliopiston informaatioteknologian tiedekunnan suostumuksella
julkisesti tarkastettavaksi yliopiston vanhassa juhlasalissa S212
toukokuun 23. päivänä 2022 kello 12.

Academic dissertation to be publicly discussed, by permission of
the Faculty of Information Technology of the University of Jyväskylä,
in building Seminarium, Old Festival Hall S212, on May 23, 2022, at 12 o'clock.

JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2022

Editors
Marja-Leena Rantalainen
Faculty of Information Technology, University of Jyväskylä
Päivi Vuorio
Open Science Centre, University of Jyväskylä

ISBN 978-951-39-9121-0 (PDF)
URN:ISBN:978-951-39-9121-0
ISSN 2489-9003

Permanent link to this publication: http://urn.fi/URN:ISBN:978-951-39-9121-0

ABSTRACT

Heilala, Ville
Learning Analytics with Learning and Analytics: Advancing Student Agency Analytics
Jyväskylä: University of Jyväskylä, 2022, 90 p. (+included articles)
(JYU Dissertations
ISSN 2489-9003; 512)
ISBN 978-951-39-9121-0 (PDF)

Pedagogically meaningful, research-based, and ethical learning analytics could foster the values and learning aims we want to advance in our society and educational system. However, it is essential to combine knowledge of the learning sciences and computational sciences when developing and applying learning analytics. This dissertation advances an analytics approach called *student agency analytics* that utilizes learning analytics methods and computational psychometrics. Student agency is a vital characteristic of a learner, especially during times of uncertainty and change. Student agency has been raised to an important position in educational policymaking, and it has been identified as an essential aspect to consider when facilitating lifelong learning.

The research advances the analysis process, examines the results from the student and teacher point of view, and provides novel insights into student agency. Specifically, the research addresses the issue of how to combine theoretical knowledge of learning and analytical methods as a comprehensive process in learning analytics while taking into account teachers' perspectives, methodological issues, and some limitations in learning analytics. The results show that *i*) student agency can be characterized, and different profiles can be generated using robust clustering, *ii*) higher course satisfaction and performance is associated with higher student agency, *iii*) students reporting low agentic resources experience various restrictive aspects in learning, *iv*) explainable artificial intelligence techniques can provide additional insight about the intricacies of student agency, and *v*) teachers can utilize the analytics results in professional reflection and pedagogical decision-making.

Keywords: learning analytics, psychometrics, student agency, higher education

TIIVISTELMÄ (ABSTRACT IN FINNISH)

Heilala, Ville
Oppiminen oppimisanalytiikassa: toimijuusanalytiikkaa edistämässä
Jyväskylä: University of Jyväskylä, 2022, 90 s. (+artikkelit)
(JYU Dissertations
ISSN 2489-9003; 512)
ISBN 978-951-39-9121-0 (PDF)

Pedagogisesti mielekkään, tutkimukseen perustuvan ja eettiset näkökulmat huomioon ottavan oppimisanalytiikan avulla on mahdollista edistää haluamiamme arvoja ja tukea oppimistavoitteita. Oppimisanalytiikan kehittämisessä ja soveltamisessa on kuitenkin tärkeää yhdistää sekä oppimistieteiden että laskennallisten tieteiden tietoa ja osaamista. Tässä väitöskirjassa kehitetään opiskelijan toimijuusanalytiikkaa, joka hyödyntää sekä oppimisanalytiikan menetelmiä että laskennallista psykometriikkaa. Opiskelijan toimijuus on eräs keskeisistä käsitteistä koulutuspoliittisessa päätöksenteossa ja olennainen asia huomioida myös elinikäisessä oppimisessa.

Tässä tutkimuksessa kehitetään opiskelijatoimijuuden analyysiprosessia, tarkastellaan analytiikkaa opiskelijan ja opettajan näkökulmasta sekä selvitetään toimijuuden yhteyksiä eri oppimiskokemuksiin. Tutkimuksessa tarkastellaan erityisesti sitä, miten oppimisteoreettinen tieto ja oppimisanalytiikka voidaan yhdistää ottaen samalla huomioon opettajan näkökulma, menetelmälliset kysymykset sekä oppimisanalytiikkaan liittyvät rajoitteet. Tutkimuksen tulokset osoittavat, että *i*) opiskelijan toimijuutta voidaan analysoida ja profiloida käyttämällä robustia klusterointia, *ii*) kurssityytyväisyys ja akateeminen suoriutuminen ovat yhteydessä opiskelijoiden toimijuuskokemuksiin, *iii*) alhaisimman toimijuusprofiilin opiskelijat kokevat erilaisten tekijöiden rajoittavan oppimistaan, *iv*) selitettävän tekoälyn menetelmät voivat antaa lisätietoa opiskelijoiden toimijuuteen liittyvistä kokemuksista, ja *v*) opettajat voivat hyödyntää analytiikan tuloksia ammatillisessa reflektiossa ja pedagogisessa päätöksenteossa.

Avainsanat: oppimisanalytiikka, psykometriikka, opiskelijan toimijuus, korkeakoulu

Author	Ville Heilala Faculty of Information Technology University of Jyväskylä Finland
Supervisors	Professor Dr. Tommi Kärkkäinen Faculty of Information Technology University of Jyväskylä Finland
	Dr. Mirka Saarela Faculty of Information Technology University of Jyväskylä Finland
	Dr. Päivikki Jääskelä Finnish Institute for Educational Research University of Jyväskylä Finland
Reviewers	Professor Dr. Leonid Chechurin LUT School of Engineering Science LUT University Finland
	Professor Dr. Irene-Angelica Chounta Department of Computer Science and Applied Cognitive Science University of Duisburg-Essen Germany
Opponent	Professor Dr. Dirk Ifenthaler Business School University of Mannheim Germany

ACKNOWLEDGEMENTS

I'm extremely grateful to Professor Tommi Kärkkäinen, Dr. Mirka Saarela, and Dr. Päivikki Jääskelä for their excellent mentoring. Their wide expertice, continuous support, and insightful feedback brought my work to a higher level. I would like to extend my sincere thanks to all the co-authors; it has been a great pleasure to do science with you. I'm very grateful to the Faculty of Information technology and the Doctoral School for supporting my work.

My sincere thanks to the opponent Professor Dirk Ifenthaler and the reviewers of the dissertation Professor Leonid Chechurin and Professor Irene-Angelica Chounta. Thank you anonymous reviewers for reviewing the articles.

I had the great pleasure of working with the brilliant people at "the Rosegarden": many thanks to Professor Raija Hämäläinen, Dr. Ari Tuhkala, and Dr. Joni Lämsä. Special thanks to pilot aces Arto Helovuo and Ilkka Tynkkynen at Finnair Flight Academy. Thanks should also go to Tiina Kullberg, Tanja Välisalo, and Marika Peltonen for exciting projects and collaboration. I also wish to thank colleagues and fellow doctoral researchers in our Human and Machine based Intelligence in Learning research group, especially Dr. Joonas Hämäläinen, Pieta Sikström, Jeri Varjosalo, Joakim Linja, and Marko Niemelä. Thank you, colleagues at the Educational Technology team: Dr. Leena Hiltunen and Dr. Antti Ekonoja. Also, many thanks to all the great minds at the Learning and cognitive science division. Thank you, Marja-Leena Rantalainen, for the editorial process.

I very much appreciate the support and feedback of the organizers and fellow participants of LAK 2020 doctoral school; especially, thank you, Professor Simon Buckingham Shum, Professor Barbara Wasson, Professor Alyssa Wise, and Professor Hendrik Drachsler. Thank you hackers at Mehackit for all the creative technology adventures: Sanna Reponen, Mikko Eloholma, Antti Suomalainen, and Heini Karppinen.

Thank you parents Helena and Sakari. Thank you brother Olli and Emilia. Thank you my dear family: Anni, Vilho, and my wife Jenni; I could have never done this without you. **<3

Jyväskylä 22.02.2022

Ville Heilala

LIST OF FIGURES

LIST OF TABLES

CONTENTS

LIST OF INCLUDED ARTICLES

PI Jääskelä, P., Heilala, V., Häkkinen, P., and Kärkkäinen, T. Student Agency Analytics: Learning Analytics as a Tool for Analyzing Student Agency in Higher Education. *Behavior and Information Technology*, 40(8), 2021.

PII Heilala V., Jääskelä P., Kärkkäinen T., and Saarela M. Understanding the Study Experiences of Students in Low Agency Profile: Towards a Smart Education Approach. *Advances in Smart Technologies Applications and Case Studies. SmartICT 2019. Lecture Notes in Electrical Engineering*, 2020.

PIII Heilala, V., Saarela, M., Jääskelä, P., and Kärkkäinen, T. Course Satisfaction in Engineering Education Through the Lens of Student Agency Analytics. *Proceedings of the 50th IEEE Frontiers in Education Conference (Conference proceedings: Frontiers in Education Conference)*, 2020.

PIV Saarela, M., Heilala, V., Jääskelä, P., Rantakaulio A., and Kärkkäinen, T. Explainable Student Agency Analytics. *IEEE Access*, 2021.

PV Heilala, V., Jääskelä, P., Saarela, M., Kuula., A-S., Eskola, A., and Kärkkäinen, T. ″Sitting at the stern and holding the rudder“: Teachers' reflection on action based on student agency analytics in higher education. *In Leonid Chechurin (Ed.). Digital Teaching and Learning in Higher Education Developing and Disseminating Skills for Blended Learning, London: Palgrave Macmillan*, forthcoming.

PVI Heilala, V., Kelly, R., Saarela, M., Jääskelä, P., and Kärkkäinen, T. The Finnish version of the Affinity for Technology Interaction Scale (ATI): Psychometric properties and an examination of gender differences. *International Journal of Human—Computer Interaction*, 2022.

LIST OF ARTICLES NOT INCLUDED

Heilala, V., Saarela, M., Reponen, S., and Kärkkäinen, T. Let Me Hack It: Teachers' Perceptions About 'Making' in Education. SmartICT 2019: Advances in Smart Technologies Applications and Case Studies, 2020, pp. 509-518

Heilala, V. Learning Analytics and Capability Approach in Education: Analysing Student Agency in Higher Education. Companion Proceedings 10th International Conference on Learning Analytics & Knowledge (LAK20), 2020.

THESIS AT A GLANCE

PII: Examines the experiences of the students exhibiting low student agency resources

Aim: To discern the constraints the students in the low agency profile experience in their studies.

Data: N=292 university students (including N2 and N3 from PI)

Methods: Student agency analytics, thematic analysis

Conclusion: The students in the low agency profile reported specifically low personal and relational resources.The main issues for the students in the low agency profile were competence beliefs, self-efficacy, student-teacher relations, course contents, time as a resource, and student well-being. Clustering can be combined with qualitative analysis.

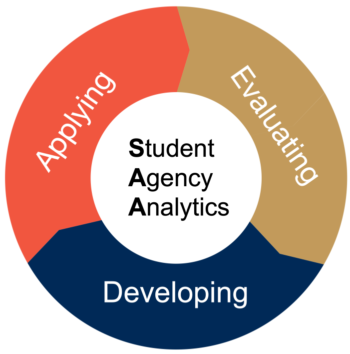
PIII: Examines the link between course satisfaction and the student agency dimensions

Aim: To explore how the course satisfaction is associated with different student agency dimensions

Data: N=293 engineering students

Methods: Statistical and unsupervised analysis

Conclusion: The lower experiences of students agency resources were associated with lower general course satisfaction. Relational resources were critical, indicating that a smaller decrease in those dimensions would more likely affect the experienced general course satisfaction than a decrease in other student agency dimensions. The results highlighted the nuanced and complex relationships between the agency dimensions and general course satisfaction.



PV: Studies how teachers utilize the student agency analytics results in reflecting their professional actions

Aim: To examine how teachers interpret the student agency analytics results and how they reflect on their actions based on the results

Data: four teachers, student data N=141 (same as PIV)

Methods: Semi-structured interviews, pentadic analysis

Conclusion: The results indicated that the teachers were able to reflect on their pedagogical actions by using student agency analytics results as the starting point for their professional thinking.

PVI: Demonstrates the use of psychometric methods in scale development

Aim: Conduct psychometric analysis of the Finnish translation of the ATI Scale. Examine gender differences in ATI.

Data: N=796 university students

Methods: Factor analysis, exploratory graph analysis, Mokken scale analysis, partial credit model, hierarchical regression

Conclusion: The translated version of the scale showed solid psychometric properties, high reliability estimates, and it formed a strong Mokken scale. Men had a slightly higher score on the scale than women when controlling for age and field of study.

PI: Presents the process of student agency analytics

Aim: Apply robust clustering to AUS scale data

Data: N1=270, N2=130 IT students, N3=78 students in teacher education

Methods: Robust clustering, MLP

Conclusion: Robust clustering can be used to create profiles of student agency. Four profiles were suggested based on the cluster validation indices. Differences between disciplines were identified. Competence beliefs, self-efficacy, teacher support, and equal treatment contributed to the course grades among IT students. Clustering and software architectural choices could facilitate security, privacy, and ethical data processing.

PIV: Applies explainable artificial intelligence methods to student agency analytics process

Aim: To clarify the most important characteristics of the different agency profiles (i.e., global explanations) and explain why an individual student was assigned to a particular profile (i.e., local explanations).

Data: N=141 (subset of PIII)

Methods: Linear and nonlinear classifiers, Shapley additive explanations (SHAP), local interpretable model-agnostic explanations (LIME)

Conclusion: The explainable student agency analytics (XSAA) could provide fine-grained information about the agency profiles. The local explanations clarified the issue of how the clustering algorithm assigned individual students to the agency profiles.

1 INTRODUCTION

Learning is the superpower of superpowers,
the one that grows the rest of them.

Rick Hanson

If you could choose one superpower, what would it be? Of course, some of us would like to be able to fly like Superman. Perhaps it would be cool to read thoughts or alter the essence of time like Doctor Strange? On the other hand, according to a psychologist and a New York Times best-selling author Rick Hanson (2018), we already are the potential superheroes of our lives because our ability to learn is the ultimate superpower. In other words, if you are able and willing to learn new things—you possess agency in learning—you could grow your own superpowers. Although it is tempting, I do not discuss supernatural powers further; after all, what you are reading is a mere academic dissertation where we need to retain facts and leave daydreaming to our leisure time.

According to Finnish adults, general education is intrinsically valuable, and learning evokes emotions like curiosity, enthusiasm, and hopefulness (Sitra, 2020). As argued in the OECD Skills Strategy 2019, megatrends like technological change, globalization, and demographic changes emphasize the importance of lifelong learning (OECD, 2019). Continuous learning is needed to develop and renew skills at different stages of life and career; competence is the best security in a changing world (Valtioneuvosto, 2020). To describe learning in terms of financial numbers, it is estimated that the global market for educational services will reach a size of nearly US$2 trillion by 2027 (Global Industry Analysts, 2021). The business potential has also been noted in Finland. For example, Sparkmind.vc[1], the first Nordic venture capital firm focusing on the learning sector, began operations in 2020.

In the past, oil was oil. Next, data were the new oil. Now, it could be justified to say that learning is the newest oil. However, amid the all learning-related

[1] https://www.sparkmind.vc/post/sparkminds-edtech-focused-vc-fund-reaches-eu55-million

hype, it is worth pausing and asking how to refine this "new" essential resource? The dominance of learning raises the fundamental questions of how to measure, analyze, interpret, and understand learning. Considering the megatrend of technological change, we need to investigate how emerging technologies and new applications in educational technology can be harnessed in favor of learning. Specifically, one important topic is how learner-generated data can be utilized effectively and ethically.

Computational data-driven methods have been a common approach to turn data into knowledge (see, Fayyad, Piatetsky-Shapiro, and Smyth, 1996), for example, in business (e.g., Ain et al., 2019) and healthcare (e.g., Y. Wang, Kung, and T. A. Byrd, 2018). Data is seen as a source of wisdom (Ackoff, 1989), efficiency (Siemens and Long, 2011), and competence (Cech, Spaulding, and Cazier, 2018). Therefore, also educational systems at different levels have started to adopt data-driven techniques to improve the learning experience and performance (e.g., Tsai, Rates, et al., 2020). A field called learning analytics combines the developments in the learning sciences and computational sciences to improve, support, and facilitate learning in a wide variety of contexts (Conole et al., 2011; Rosé, 2018).

However, developing and applying learning analytics is not straightforward or unproblematic (e.g., Mathrani et al., 2021). Artificial intelligence is capable of doing incredible things like writing an entire article (GPT-3, 2020), predicting the future (NORAD, 2021), and simulating complete universes (Villaescusa-Navarro et al., 2021), but it is still in its infancy in the field of education (e.g., Gao, 2021; Qin and G. Wang, 2022). Successful development and implementation of learning analytics require research-based knowledge and insights about theories of learning, psychological and educational measurement, analytical and computational methods, technical capabilities, and ethical aspects (e.g., Greller and Drachsler, 2012; Ifenthaler, Gibson, et al., 2021). Furthermore, it is vital to take into account the pragmatic dimension of learning analytics so that the results and outcomes provide actionable information towards the goals of different stakeholders in the educational domain. Specifically, learning analytics research and development should consider addressing students' affective support and "well-being and relatedness rather than focusing on performance measures alone" (Blumenstein, 2020, p. 13). From the practitioner's point of view, analytics could help teachers to analyze, reflect on, and improve their pedagogical practice and increase students' learning experiences (e.g., Greller and Drachsler, 2012; Yau and Ifenthaler, 2021). Zawacki-Richter et al. (2019) pointed out the need of integrating educational perspectives and technological developments and taking into account teacher's perspective in learning analytics and artificial intelligence in education.

In my dissertation, I will scrutinize a learning analytics process called student agency analytics which utilizes a psychometric scale for obtaining student data and robust statistics and machine learning for producing analytical results relating to students' social, affective, and cognitive learning experience. In the context of this dissertation, the learning experience is characterized using a multi-dimensional construct of student agency. The process of student agency analytics involves that *i*) students attend a course and perform their studies, *ii*) they respond to a student agency questionnaire at some point during the course, which is followed by *iii*) an analysis using robust clustering and *iv*) visualization of the results to students and the teacher.

1.1 Research questions

My dissertation deals with the analysis process, examines the results from the student and teacher point of view, and provides novel insight into student agency. Specifically, the research addresses the issue of how to combine theoretical knowledge of learning and analytical methods as a comprehensive process in learning analytics while taking into account teachers' perspective (RQ1), methodological issues (RQ2), and some limitations in learning analytics (RQ3). The research questions of the dissertation are as follows:

RQ1 For what purposes teachers could utilize student agency analytics in higher education?

RQ2 What methodological requirements student agency analytics introduce?

RQ3 How student agency analytics could overcome some of the limitations in learning analytics?

Furthermore, I demonstrate the use of psychometric methods in scale development by conducting a translation process and psychometric analysis for a scale relating to human—technology interaction. Individual differences can affect the use of educational technology (e.g., Sahin and Ifenthaler, 2022) and the translated scale could be useful for assessing individual differences in technological learning environments. Overall, the dissertation examines human learning using machine learning, and it can be placed at the intersection of learning analytics and computational psychometrics.

1.2 Structure of the work

This compilation part of my dissertation reflects on the methods, processes used, and decisions made during the research work. Furthermore, I present the summaries of the articles describing the research aims, data and methods, primary results, research contributions, and my contributions. The Thesis at a Glance section provides a brief overview of the structure of the dissertation. In a broad sense, Articles PI and PIV concentrate on the methodological and analytical approach (i.e., the development of student agency analytics), Articles PII and PIII utilize student agency in a real educational context and provides evidence on how student agency is associated with other learning experiences (i.e., applying), and Article PV examines the analytics results from the teacher point of view (i.e., evaluating).

Finally, Article PVI demonstrates the use of psychometric analysis in scale development. The article relates closely to the technological megatrend and complements the ensemble of articles from a technological and methodological perspective. Section 2 is devoted to dealing with the main research domains covering a brief overview of learning analytics and the theoretical construct of student agency. Section 3 presents the quantitative and qualitative methods used in the research articles. Section 4 summarizes the research aims, data and methods, primary results, research contributions, and my contributions to each accompanying article. Finally, Section 5 concludes with a general discussion, states the limitations, and provides suggestions for future research.

2 MAIN RESEARCH DOMAINS

The idea of using machines and their computational capabilities in advancing human learning emerged already over a century ago. One of the earliest mentions of automatized and personalized teaching was phrased by Edward L. Thorndike, the professor of educational psychology at Columbia University and the father of a learning theory called connectionism (E. L. Thorndike, 1912, p. 165):

> If, by a miracle of mechanical ingenuity, a book could be so arranged that only to him who had done what was directed on page one would page two become visible, and so on, much that now requires personal instruction could be managed by print.

A while later, another professor of educational psychology, Sidney Pressey, devised a teaching machine (see, Benjamin, 1988) for providing automatized drill and practice in the form of multiple-choice questions for his students (Pressey, 1927; Pressey, 1926). However, a certain kind of future of educational technology was visioned by an artist Jean-Marc Côté already in turn of the twentieth century in a series of images titled *En L'An 2000* (*In the Year 2000*) (Figure 1). Perhaps the atmosphere of the image represents the idea of "industrial revolution in education" argued by Pressey (1932), who drew from the achievements of the second industrial revolution. Now, almost a century later, we have been living amid the digital revolution and fourth industrial revolution, and, perhaps, transitioning to imagination age (e.g., Alvarez, 2018). Nevertheless, one of the many manifestations of Côté's futuristic predictions today is learning analytics—the art and science of introducing ubiquitous computation, algorithms, and machine learning for understanding and advancing human learning. An imaginative reader could use the Côté's vision from 1901 for reflecting on the potential advantages and disadvantages of integrating technology and human learning.

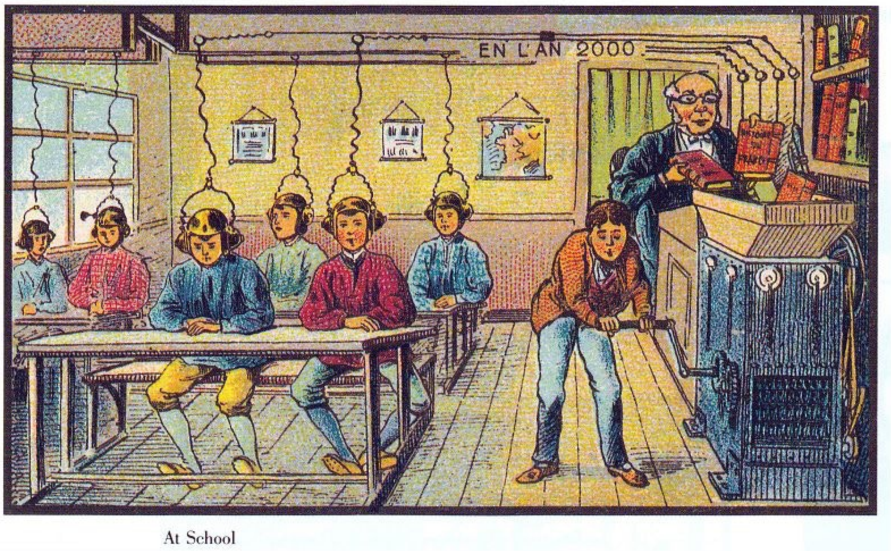


FIGURE 1 Jean-Marc Côté's vision of school in 2000 drawn a hundred years earlier in 1901 (image in the Public Domain).

2.1 Learning analytics

Learning analytics is an interdisciplinary field that aims to advance data-driven approaches in education. As explicated above, the idea of using computation, automatization, and analytics originated among a few educational psychologists and the "teaching machine" movement at the beginning of the twentieth-century (see, Benjamin, 1988). Rosé (2018) traced the scientific roots of learning analytics to machine learning, data mining, applied statistics, intelligent tutoring systems, education, psychology, cognitive science, and computational linguistics. Siemens (2013) mentioned that the fields and activities relating to citation analysis, social network analysis, user modeling, cognitive modeling, intelligent tutors, knowledge discovery in databases, adaptive hypermedia, and E-learning had paved the way for the development of learning analytics. Teasley (2018) and Rosé (2018) considered the synergies between research communities of learning analytics and learning sciences; combining research expertise and crossing disciplinary bounds is vital to advancing how we understand learning.

Perhaps one the first, the most "official" and frequently used definitions of learning analytics comes from the message of chairpersons of the First International Conference on Learning Analytics and Knowledge in 2011, where it was defined as "the measurement, collection, analysis, and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs" (Conole et al., 2011). Guzmán-Valenzuela et al. (2021) point out that the emphasis of learning analytics should be on "using pedagogy-based approaches and educational theories in understanding learning" and the development of learning analytics applications should involve both teachers and students. Also, Rosé (2018, p. 512) highlighted the importance of theoretical frameworks for operationalizing variables and building models instead of "atheoretical empiricism." Buckingham Shum (2018; 2019) positions learning analytics to a so-called Pasteur's quadrant. Pasteur's quadrant refers to research that both increases fundamental understanding and takes into account the potential use of the research results (i.e., use-inspired basic research) (Stokes, 1997, p. 73–74). According to Buckingham Shum (2019), learning analytics aims to "design and deploy analytics that demonstrates how theory can inspire models, algorithms, code, user experiences, teaching practices, and ultimately, learning."

Learning analytics can be depicted as a cyclic process between learners, data, measurement and analytics, and interventions (e.g., Clow, 2012; Di Mitri et al., 2018; Ifenthaler and Greiff, 2021; Saarela and Kärkkäinen, 2017; Siemens, 2013). Also, theories of learning, pedagogical aims, and analytical methods are central aspects to integrate into the process (Heilala, 2018). Using topic modeling, Lemay, Baek, and Doleck (2021) found that during 2015–2019, the most common research topic in learning analytics articles they reviewed was students' engagement and performance in online courses. Other common topics they found related to analytical methods, feedback through assessment and activities, system design, and social network analysis. Also, Blumenstein (2020) found that examinations of cognitive and behavioral engagement formed the largest body of learning analytics research during 2011–2019. The majority of the learning analytics research X. Du et al. (2021) reviewed related to the prediction of learners' performance, decision support for teachers and learners, and detection of behaviors and learner modeling. Furthermore, they found that a greater amount of research targeted higher education than K-12 education and MOOCs. It seems that learning analytics research could benefit from broadening the scope of examination to the affective and emotional domain of learning and beyond the higher education context.

A wide variety of data sources could be used in learning analytics. According to Samuelsen, Chen, and Wasson (2019), the most common sources are learning management systems, questionnaires, and student information systems. They also found that the most common data types were activity logs, student background information, questionnaire data, and performance data. Real-time learning analytics provides possibilites for monitoring and supporting classroom activities (e.g., Chounta and Avouris, 2016). Multimodal data collection (i.e., multimodal learning analytics) could provide additional insight about learning processes (Blikstein, 2013; Scherer, Worsley, and Morency, 2012). Multimodal learning analytics is concerned with the integration of different data sources and modalities, for example, text and speech analysis, analyzing handwriting and sketches, action and gesture analysis, analysis of affective states and neurophysiological signals, eye tracking analytics, and social behavior (Blikstein and Worsley, 2016; Di Mitri et al., 2018; M. X. Li et al., 2021). Di Mitri et al. (2018) pointed out that multimodal data can only capture the observable behavior (i.e., input space), and the challenge is to derive relevant interpretations relating to, for example, emotions, motivation, cognition, and beliefs (i.e., hypothesis space).

Analysis methods in learning analytics cover both descriptive and predictive analytics (X. Du et al., 2021). Statistical modeling (e.g., regression and correlation analysis) is a common approach (Namoun and Alshanqiti, 2020) but also classical machine learning methods and soft computing methods (e.g., Bayesian methods, decision tree, random forest, support vector machines) are widely utilized (Charitopoulos, Rangoussi, and Koulouriotis, 2020). Furthermore, network analytics can be used to provide insight about students' behavior and performance (e.g., Khan, Kaliteevskii, Shnai, and Chechurin, 2020).

A closely related field, educational data mining, is "an emerging discipline, concerned with developing methods for exploring the unique and increasingly large-scale data that come from educational settings and using those methods to understand students better, and the settings which they learn in" (International Educational Data Mining Society, 2022). The focal point of recent educational data mining research seemed to relate more to techniques and methods of data analysis, whereas learning analytics centered more on student engagement, teaching tools, and social network analysis (Lemay, Baek, and Doleck, 2021). In other words, learning analytics deals with sensemaking and action, and educational data mining is more about methodological issues (Siemens, 2013). Saarela (2017) pointed out that the methods in learning analytics and educational data mining are interrelated. Depending on the analytical lens, expected outcomes, and used methods, also other terms can be used to frame the research and practice concerning education and data (Romero and Ventura, 2020): academic analytics and institutional analytics (e.g., Campbell, DeBlois, and Oblinger, 2007) seek to facilitate institutional decision making, teaching analytics examines teacher's actions and the practice of teaching (e.g., Prieto et al., 2016), big data in education deals with data characterized by volume, variety, value, and velocity (e.g., Daniel, 2019), educational data science can be seen as a broad umbrella term (e.g., Romero and Ventura, 2017), data-driven education and data-driven decision-making in education refer to wide and effective use of data in education (e.g., Datnow and Hubbard, 2016).

The main benefits of learning analytics relate, for example, to the promises of providing personalized learning, improving instructor performance, advancing post-educational employment, and, in general, enhancing students' learning outcomes, behavior, and processes (Avella et al., 2016). Ifenthaler and Greiff (2021) pointed out the importance and potential benefits of analytics-enhanced assessment, which include, for example, empovering learners in their learning process, activating peer support, facilitating reflection, and clarifying learning intentions. Chounta, Bardone, et al. (2021) found out that teachers expected artificial intelligence in education to enhance their effectiveness, efficiency, student—teacher relationship, course planning, and personal attributes and skills.

However, several challenges and unresolved obstacles of effectively utilizing learning analytics have been identified in the literature, for example, shortage of valid evidence relating to effectiveness of learning analytics (e.g., R. Ferguson and Clow, 2017; Ifenthaler, 2021; Tsai and Gasevic, 2017; Wilson et al., 2017), in-discriminate use of analytical approaches (e.g., Guzmán-Valenzuela et al., 2021; Wilson et al., 2017), lack of pedagogical implementations (e.g., González-Calatayud, Prendes-Espinosa, and Roig-Vila, 2021; Tsai and Gasevic, 2017), obstacles relating to leadership and institutional adoption (e.g., Tsai and Gasevic, 2017), challenges relating to learning analytics policies and guiding principles (e.g., Ifenthaler, Gibson, et al., 2021; Tsai and Gasevic, 2017), ethical challenges (e.g., R. Ferguson, Hoel, et al., 2016; Ifenthaler, Gibson, et al., 2021; Slade and Prinsloo, 2013), development and implementation processes not user-focused (e.g., Guzmán-Valenzuela et al., 2021; Ifenthaler, Gibson, et al., 2021), and overemphasizing performance evaluation instead of motivation, engagement, and satisfaction (e.g., Guzmán-Valenzuela et al., 2021). Furthermore, Ifenthaler and Yau (2020, p. 1983) highlighted the lack of theoretical clarity and clear definitions of the key constructs used in learning analytics and, thus, "operationalisations of these constructs become blurred and their valid measurement becomes impossible." The identified challenges emphasize the need for an interdisciplinary research approach, close collaboration between different stakeholders, and thoughtful consideration of the purpose and goals of the analytics when researching, developing, and implementing learning analytics applications. As Ifenthaler, Gibson, et al. (2021, p. 2145) summarized, "putting learning back into learning analytics requires a complex set of actions and strategies for policy makers, researchers, and practitioners."

The learning analytics approach presented in this dissertation utilizes a questionnaire to acquire data relating to the subjective learning experience. As stated above, questionnaires were one of the most common sources of data in learning analytics (Samuelsen, Chen, and Wasson, 2019). Questionnaires have been used in learning analytics research as a method to examine learning analytics per se (e.g., Whitelock-Wainwright et al., 2019). Shum and Crick (2012) utilized the Effective Lifelong Learning Inventory questionnaire (Crick, Broadfoot, and Claxton, 2004) and visualization of the results to examine learners' dispositions. Relatedly, dispositional learning analytics combines data from self-report instruments (i.e., learner data) with data recorded from different learning systems (i.e., learning data), which could provide a link between analytics results and actionable interventions (D. Tempelaar, Rienties, and Q. Nguyen, 2020; D. T. Tempelaar, Rienties, and Q. Nguyen, 2017).

2.2 Student agency

Where is agency, there is action and change. As Schlosser (2019) discusses, in a broad sense, it is possible to identify agents and agency in practice everywhere. He mentioned that the intellectual history of agency points back to Hume and Aristotle. However, he states that in contemporary analytic philosophy, agency is often associated with the works of Anscombe (1957) and Davidson (1963) who dealt with agency in terms of the intentionality of intentional action. According to the standard conception of agency, “a being has the capacity to exercise agency just in case it has the capacity to act intentionally, and the exercise of agency consists in the performance of intentional actions and, in many cases, in the performance of unintentional actions” (Schlosser, 2019). The standard theory of agency states that “a being has the capacity to act intentionally just in case it has the right functional organization: just in case the instantiation of certain mental states and events (such as desires, beliefs, and intentions) would cause the right events (such as certain movements) in the right way” (Schlosser, 2019). Emirbayer and Mische (1998, p. 970) emphasized social and relational aspects and defined human agency as “the temporally constructed engagement by actors of different structural environments—the temporal-relational contexts of action—which, through the interplay of habit, imagination, and judgment, both reproduces and transforms those structures in interactive response to the problems posed by changing historical situations.”

Furthermore, different types of agency have been proposed to exist. Mental agency relates to control of mental operations, epistemic agency refers to the control over agent’s beliefs, shared agency emerges in situations where two or more agents collaborate, relational agency takes into consideration interpersonal relationships, and artificial agency raises the question of whether artificial intelligence and robots are capable of agency (Schlosser, 2019). Agency should also be considered in a wider perspective than just from the individual agent’s point of view. For example, according to structuration theory originally proposed by Giddens (1984), people, environment, and social systems bring on the interplay between structure and agency: while people create the social structures, they also conform to the expectations of those structures. Furthermore, it is worth noting that in a post-humanist thinking, agency does not necessarily concern only humans; also, non-humans can be considered as possessing at least some degree of agency. For example, animal agency (e.g., Jamieson, 2018; Steward, 2009) and algorithmic agency (e.g., Peeters, 2020) has been discussed in the literature. Examples of non-human agency in educational technology are intelligent agents (e.g., Baylor, 1999) and pedagogical agents (e.g., Kim and Baylor, 2016).

As briefly depicted above, human agency as a scientific concept has a long and multifaceted history. On the other hand, student agency—human agency explicitly relating to students in various educational settings—seems to be a pretty recent idea in the research literature and in practice. For example, in the Finnish educational context in the nineteenth century, student centeredness and agency was not yet very much considered (e.g., Puranen-Impola, 2022, p. 18). Interestingly, while the idea of student agency reaches back to the time of Enlightenment (Biesta and Tedder, 2007; Jääskelä, Poikkeus, Vasalampi, et al., 2017), the earliest notions explicitly referring to *student agency* in the scientific literature I found dated back to 1980–90s. The following brief treatise on student agency tries to trace its intellectual roots. The presentation does not aim to be a full-coverage systematic literature review; instead, it highlights some of the key findings from the brief explorations of the student agency literature. The search and review strategy was as follows: First, I searched through the literature year by year using

Google Scholar and Scopus starting from 1981, which was chosen as the starting point. Then, if the article discussed student agency more than just mentioning the concept, it was included in the following summary. Unfortunately, in the beginning of the millennium, the overall picture of the whole concept started to get vague (e.g., articles just mentioning the concept without a clear idea or definition). Thus I started to trace the literature year by year from the most recent articles finally reaching the point where I changed the search strategy.

First, student agency was seen as a positive force directing the learning process. King (1983, p. 188) mentioned agency as a motivational force in art learning and proposed that "enhancing student agency through personal choice would affect art learning in a positive manner." Evans and Nation (1989) referred to research in distance education that shows "the power of student agency in the shaping of learning."

The great trailblazers of educational thought—Piaget, Vygotsky, Fraire—were mentioned. Scardamalia and Bereiter (1991, p. 39) dealt with agency in knowledge building in computer supported intentional learning environments. They positioned agency in a learning process to the continuum of constructivism: in behaviorally oriented social psychology in the one extreme agency relates merely to students' responsibility of their learning, in Piagetian view in the other extreme a learner builds knowledge of the world through acting, and between them is the Vygotskian view where an agent functions according to activities, but the knowledge emerges from the social interaction. Briskin and Coulter (1992, p. 259) considered student agency through Freirian perspective and feminist pedagogy, and pointed out the "student responsibility to produce, shape, and interrupt classroom power dynamics."

Different views of education received attention. Also, student agency was seen as a part of curricula, but it also had a reach beyond classroom and schooling. Davies (1994) discussed student agency from transformative education and Giddens' structuration theory point of view. She distinguished two types of student agency: subversive agency relates to the students' ability to challenge the rules in a school, and transformative agency refers to the strive for human rights. Carver (1997) defined agency as one of the three facets forming student experience in experiential education—other facets being belonging and competence. Carver (1996, p. 10) suggested that experiential education can promote student agency, which means that students will become the "change agents in their lives and communities" and they possess the locus of control of their lives which they can use as "a source of power to generate activity". Ottey (1996) treated student agency from the critical pedagogy and curriculum point of view. She suggested that student agency is a vital part of the curriculum and that the school becomes more relevant for the students when they have opportunities to apply their knowledge inside and outside their classroom.

Student agency was described to be contingent on the context, environment, social (or power) relationships, and interpretations of meaning. Ewald and Wallace (1994) discussed student agency in the context of classroom discourse and contrasted the teacher-centered and student-centered approaches. They described agency "as both the ability to interpret events as well as the ability to influence, change, or redirect them within a specific situation," and emphasized that people can have different interpretations of the same issues arguing further that "the evaluation of pedagogical choices *in situ* must consider student agency both at the level of direct impact on a classroom agenda and at the level of interpretive difference" (Ewald and Wallace, 1994, p. 343, p. 349). Fenwick (1996, p. 22) also discussed classroom management and mentioned the tension between teacher control and student agency.

Student agency was also referred to as listening to students' *voices* and considering learning, education, and research from the students' point of view. Lincoln (1995, p. 88) discussed what it means that "student voices reenter educational learning and inquiry", and noted that the assumptions of the society and social context could undermine student agency. Walsh (1994, p. 218) defined teacher and student agency as "their capacity to act in and on their environments." She pointed out that as simple as asking questions but using an impersonal and objective voice could indicate a lack of student agency. From the research point of view, B. J. Smith (2000) criticized the positivist research paradigm in education because it did not take into account student agency.

From the beginning of 2000, the number of articles mentioning student agency started to increase slowly. In some research, student agency was mentioned without a clear conceptual backing (e.g., C. Anderson, Day, and McLaughlin, 2006; Holm, 2010), it was mentioned being somehow "related to" other constructs (e.g., Taub et al., 2020), or it was merely reduced to the level of students' actions, participation, making efforts, or making decisions (e.g., Akos, 2004; Mcintyre, 2006; H. Nguyen et al., 2018). Thus, I changed my search strategy and started tracing the literature from the most recent articles dealing with student agency. I found that still in the 2020s, scholars and educators pondered on the concept.

In an article titled *What is student agency and why is it needed now more than ever?*, Vaughn (2020) built from the theoretical orientations of Dewey, Vygotsky, and Bandura. She formulated a model of student agency, which comprises of three-dimension: dispositional (i.e., agency relates to intentions and purpose), motivational (i.e., agency relates to persisting and choice-making), and positional (i.e., agency relates to interactions, negotiations, and perceptions of self). Vaughn (2020, p. 115) proposed that "agency is structured *with* and *alongside* students and during experiences that allow for students to share their voice, histories, cultural identities, experiences, languages, and interests". She steered that teachers could support student agency by providing students opportunities to "make choices, act on their intentions, and take actions in their efforts to develop their own positions and opinions" (Vaughn, 2021, p. 18). Kangas et al. (2014) considered student agency in elementary education in an out-of-classroom setting. They drew from the previous literature and pointed out that agency "relates to the capacity to initiate purposeful action that implies will, autonomy, freedom, and choice within the affordances of the worlds that they inhabit" and defined agency "as acting authoritatively and accountably" (Kangas et al., 2014, p. 34–35).

In an influential article titled *What is student agency? An ontological explo-*

ration in the context of research on student engagement, Klemenčič (2015) considered student agency by drawing from social psychology and sociology. According to her, agentic orientation—or *will*—covers the temporal aspect of student agency and agentic possibility—or *power*—means the students perceived power to reach the intended goals depending on the context. She proposed that student agency is *i)* developed by interacting within a particular socio-structural and relational context, *ii)* it can be strong or weak depending on the situation, *iii)* it is temporally embedded, *iv)* it is enabled, constrained, and challenged by the interdependent educational, social, cultural, political, and economic conditions, *v)* it takes shape relationally in physical and virtual networks, and *vi)* it is manifested in different modes. Stenalt and Lassesen (2021) reviewed the literature concerning student agency in the higher education context in 1980–2021. All articles they found were published after 2000, and the majority of the 29 articles concerned student-focussed teaching and learning, real-life learning situations, assessment, feedback, globalization, and internationalization. Two of the articles related to learning analytics, of which Article PI was one. The other was a qualitative study by Tsai, Perrotta, and Gašević (2020, p. 555) which critically examined "the extent to which learning analytics can be used to enhance student agency and educational equity." They concluded that learning analytics should be based on the learning sciences, need to leverage instead of replacing human contact, and consider how transparency and visibility of data policies, practices, and algorithms affect agency. In summary, the research topic advanced in this dissertation presents a unique line of research from the student agency and learning analytics point of view.

The brief qualitative conceptual exploration presented in this section provided a perspective to the multifarious topic of student agency. To distill the findings from the diverse literature, I formulate the following summarazing definition: *student agency is an emergent quality of a learner that is formed by the desire, capability, and possibility to drive intentional change for personal development and that reaches to the future by utilizing the resources of the past and the present*. In addition to qualitative definitions, another way to characterize a concept or a construct is to create an instrument for measuring it.

In the context of this dissertation, student agency is approached using a conceptualization behind a psychometric assessment instrument. Jääskelä, Poikkeus, Vasalampi, et al. (2017) have theorized and identified the construct of student agency. They have developed the Agency of university students (AUS) scale, which establishes student agency as a multidimensional construct (Figure 2). Their AUS scale takes into account individual, sociocultural, interactional, and contextual aspects of learning. Jääskelä, Poikkeus, Häkkinen, et al. (2020) synthesize the literature concerning student agency and adopt a view that centers on the "dynamic, contextually situated, and relationally constructed nature of agency while also acknowledging its subjective standpoint and the interplay between resources and a person's capacities." They defined student agency in higher education as "a student's experience of having access to or being empowered to act through personal, relational, and participatory resources, which allow him/her to engage in purposeful, intentional, and meaningful action and learning in study contexts" (Jääskelä, Poikkeus, Häkkinen, et al., 2020). The AUS scale has a link to student-centered education, and it assesses specifically the students' experiences relating to their agentic resources (Jääskelä, Poikkeus, Häkkinen, et al., 2020).

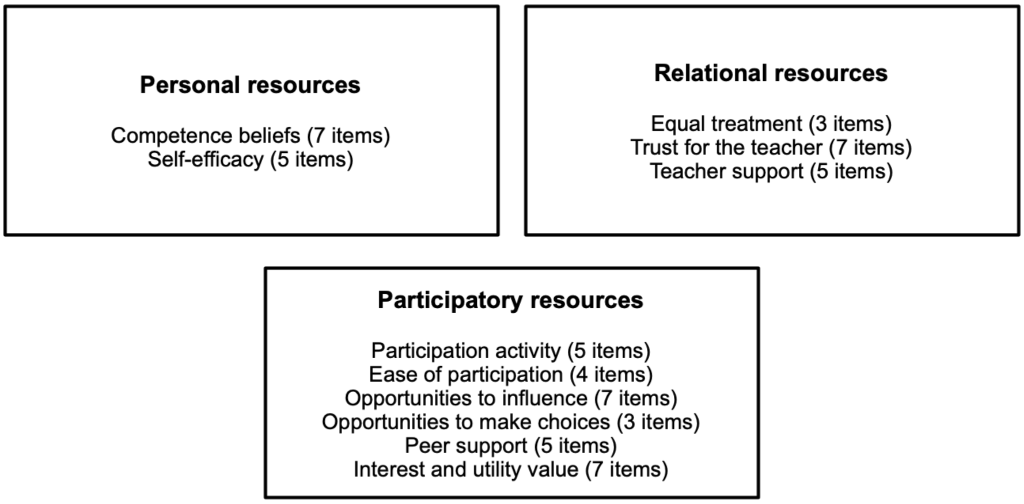


FIGURE 2 Student agency dimensions and resource areas

The AUS scale consists of 58 Likert items assessed in a five-point scale (1 = fully disagree–5 = fully agree). The scale has been considered to have acceptable first-order and second-order structural fit based on confirmatory factor analyses (Jääskelä, Poikkeus, Häkkinen, et al., 2020). As a result, the AUS scale consists of three broad resource areas (i.e., individual/personal, contextual/participatory, relational) comprising of 11 student agency resource dimensions, which are described as follows (Jääskelä, Poikkeus, Häkkinen, et al., 2020; Jääskelä, Poikkeus, Vasalampi, et al., 2017):

- **Personal resources**: relate to student's self-processes and beliefs
  - **competence beliefs**: e.g., sense of understanding and having competence needed for learning the course contents
  - **self-efficacy**: e.g., confidence as a learner, taking up challenges
- **Participatory resources**: relate to interaction, involvement, and engagement in activities
  - **interest and utility value**: e.g., perceiving subject as interesting and useful
  - **participation activity**: e.g., interacting in learning situations, actively engaging in discussions, completing assigned tasks, taking initiatives
  - **ease of participation**: e.g., encouraged to critical thinking, freedom to express own thoughts
  - **opportunities to influence**: e.g., experience of being listened to, can influence on working methods
  - **opportunities to make choices**: e.g., can make choises between ways of completing the course

  - **peer support**: e.g., providing, receiving, and accepting support from peers
- **Relational resources**: relate to power and supportive relations between the teacher and students
  - **teacher support**: e.g., teachers attitude towards students, power relations
  - **trust in teacher**: e.g., psychosocial environment, approachability of the teacher
  - **equal treatment**: e.g., equality among students, experience that the teacher treats students equally

3 METHODOLOGY

This section briefly presents the essential methodological basis of the articles. First, I depict some of the main psychometric methods used in scale construction, namely, latent variable analysis and analysis of reliability. The methods could be used in scale construction and development for learning analytics applications. The Agency of University Students (AUS) scale (Jääskelä, Poikkeus, Häkkinen, et al., 2020; Jääskelä, Poikkeus, Vasalampi, et al., 2017) is a psychometric instrument used to assess students' self-reported experiences relating to agentic resources in learning. The AUS scale was already developed prior to my dissertation. Thus, I wanted to become acquainted with the common psychometric methods used in scale development. Article PIV demonstrates the use of the main psychometric methods presented in this chapter.

Secondly, I will depict the methods used in student agency analytics to analyze and interpret the scale data, namely, robust clustering and selected methods relating to explainable artificial intelligence. Thirdly, I present the methods to analyze qualitative data relating to students' open-ended responses and teachers' interviews. Lastly, I consider the ethical issues relating to the student agency analytics process.

3.1 On scale construction

The renowned William Thomson, also known as Lord Kelvin, declared in a lecture relating to measurement in physical sciences that "when you can measure what you are speaking about, and express it in numbers, you know something about it" (Thomson, 1884, p. 149). The positivist thought had been making headway also in the human sciences; Thomson's contemporary and compatriot, Sir Francis Galton, had just declared a few years before that psychometry "means the art of imposing measurement and number upon operations of the mind (Galton, 1879). Since then, psychometrics has developed into a multifaceted and powerful discipline (Vessonen, 2021).

Psychometrics is a scientific discipline examining measurement, testing, and assessment in education and psychology. Psychometric methods play a key role in educational and psychological measurement and analysis of latent variables, traits, behavior, experiences, and cognitive functions relating to learning. In an interdisciplinary sense, learning analytics could draw from the extensive toolbox and history of research in psychometrics. Recently, computational methods have been gaining popularity also among psychometricians. As a results, a new interdisciplinary field, computational psychometrics (e.g., von Davier, Mislevy, and Hao, 2021b), has emerged that has been also noted in the field of learning analytics (e.g., Drachsler and Goldhammer, 2020; Mislevy, 2019; von Davier, Mislevy, and Hao, 2021a).

Questionnaires and questionnaire data are one of the most common sources of material for learning analytics (Samuelsen, Chen, and Wasson, 2019). Relatedly, psychometric instruments, for example, scales, are common components of different questionnaires and are widely used in survey research. Vessonen (2021, p. 10) notes that the instruments used in questionnaires—"measures of mind", as she figuratively calls them—can seem flimsy outside, but in reality, they are backed with extensive research, analytical knowledge, and methodological know-how. Many instruments are potent and influential in steering policies and affecting people's lives (e.g., measures of happiness, wellbeing, or depression).

Scales are a common way to obtain a numerical value for a mental construct. A measurement procedure consists of three steps: *i*) identifying the object being measured, *ii*) identifying the behavior or property being measured, and *iii*) identifying a numerical rule that assigns a number to the property of the object being measured (Lord and Novick, 1968). In an optimal case, a researcher might find an existing and well-constructed scale suitable for his or her research purpose. However, finding an existing scale might not be possible, especially in a specific and small language. Another option would then be to conduct a translation and psychometric analysis process for the scale existing in a different language (e.g., Brislin, Lonner, and R. M. Thorndike, 1973). If no suitable scale exists, the last option is to develop a new scale from scratch. However, developing psychometric instruments poses a chicken or the egg problem: one needs a theory to support the development of the measure, and, on the other hand, one needs a measure to develop and support the theory (Vessonen, 2021). DeVellis (2017) outlines the steps involved in scale development: *i*) determine the aims of the measurement, *ii*) generate an item pool, *iii*) select the format for measurement, *iv*) ask experts to review the initial item pool, *v*) add possible validation items, *vi*) collect initial data, *vii*) analyze the data, and *viii*) optimize the scale length. The scale items are constructed so that they reflect the theoretical construct. The scale developer should take into account, for example, redundancy of the items, number of items, reading difficulty and comprehensibility, and item wording (DeVellis, 2017). After the initial data collection, the data should be exposed to a rigorous psychometric analysis in case of a new scale. The analysis provides insight into the properties of the measurement instrument.

The internal structure is of vital interest when constructing new measurement instruments. One would naturally want the structure of the empirical data collected using the measurement instrument to comply with the theoretical construct being measured. Unidimensional concepts and instruments are convenient from the practical point of view as they are more straightforward to deal with (Hattie, 1985; McDonald, 1981). Often psychometric analysis concerns the analysis of unidimensionality. However, unidimensional concepts can be used to form multidimensional constructs and measurement instruments. For example, the scale analyzed in Article PVI represents a unidimensional concept, whereas the AUS scale can be considered to consist of 11 unidimensional scales representing the 11 dimensions of student agency.

Latent variables can be thought of "as the unobserved determinants of a set of observer scores," and the latent variables are "considered to be the common cause of the observed variables" (Borsboom, 2005, p. 4). Factor analysis is a common approach in latent variable analysis. Another approach to latent variable analysis is the item response theory (Borsboom, 2005, p. 50) which includes, for example, Mokken scale analysis (Mokken, 1971; Molenaar, 1991) and partial credit model (Masters, 2016; Masters, 1982). Classical test theory (a.k.a., true score theory) deals with the reliability and the true score, which is the expected value of the observed variable, and it is equated with the measured construct (Borsboom, 2005). In the following treatment, the reliability of a scale is considered from the CTT and factor analysis point of view.

Latent variable analysis

Parallel analysis The first step when analyzing the structure and the dimensionality of data is determining the plausible number of dimensions. Parallel analysis is a common method for assessing the initial dimensionality and choosing the number of factors to retain. In general, the parallel analysis compares randomly sampled data with the original data. Parallel analysis with PCA extraction (PA-PCA) also called as the Horn's PA (Horn, 1965) using polychoric correlation has been suggested for different types of data (Garrido, Abad, and Ponsoda, 2013). Also, another variant of PA, parallel analysis using minimum rank factor analysis as an extraction method (PA-MRFA), has been proposed to be used when assessing the number of common factors underlying ordered polytomous scored variables (Timmerman and Lorenzo-Seva, 2011). Furthermore, a complementary method to parallel analysis, MAP, is based on the matrix of partial correlations, and it is also suggested to be an accurate method for assessing the number of dimensions (Zygmont and M. R. Smith, 2014), especially in a unidimensional case (Golino et al., 2020). The smallest MAP value designates the number of dimensions in the data (Zwick and Velicer, 1986).

Figure 3 demonstrates the use of PA-MRFA for data used in Article PVI. The first factor component explains a greater amount of the variance in the original sample than in the permuted random samples. On the other hand, the second component explains less amount of variance in the original sample than in the

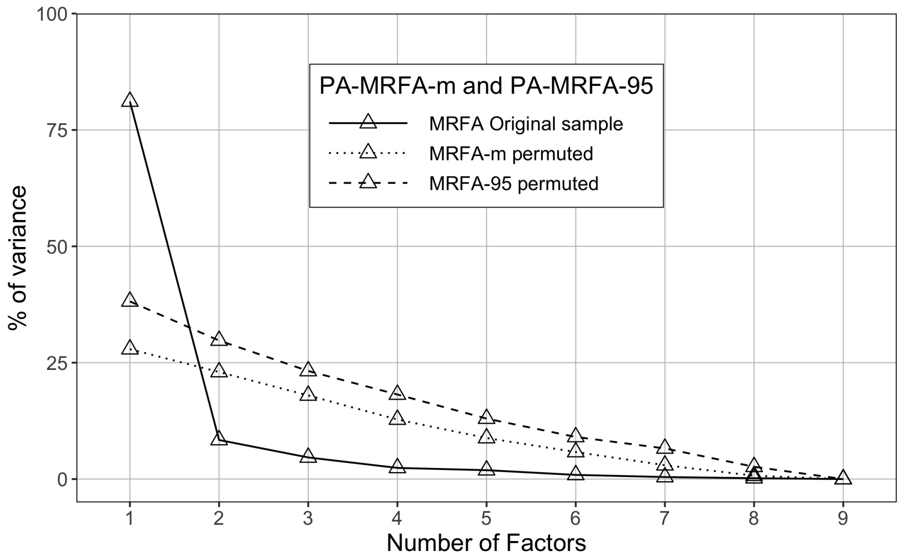


FIGURE 3 Parallel analysis for data in Article PVI using PA-MRFA suggested unidimensional structure. PA-PCA was used in the original Article PVI.

permuted samples. As a result, the parallel analysis suggested a unidimensional structure. However, $k \pm 1$ dimensional structure is suggested to be examined, k being the number of dimensions suggested by the parallel analysis (Lim and Jahng, 2019). Thus, one-dimensional and two-dimensional structures should be selected for further analysis in this case. Permuted samples were obtained using column permutation (500 random data sets), polychoric correlation, and quantile thresholds 50% (median, PA-MRFA-m) and 95% (PA-MRFA-95). PA-PCA provided similar results as described in PVI (Auerswald and Moshagen, 2019; Buja and Eyuboglu, 1992).

Exploratory factor analysis Exploratory factor analysis (EFA) is traditionally the next step after the initial dimensionality analysis. Previous literature suggests examining ($k \pm 1$)-factor structures suggested by the dimensionality assessment methods (e.g., PA) (Lim and Jahng, 2019). Thus, for example, a two-dimensional structure should also be examined if the scale is suggested to be unidimensional. For discrete Likert-type data polychoric correlation (Holgado–Tello et al., 2010), weighted least squares (WLS) estimation (Goretzko, Pham, and Bühner, 2019), and direct oblimin rotation to allow the factors to be correlated (Brown, 2015) could be used. Different EFA models can be compared with each other using model fit indices (e.g., RMSEA) (Fabrigar and Wegener, 2012, p. 120). Item quality can be assessed using the values of communality and complexity. Communality is the variance of the item that is explained by the factors in the model. A low communality value might indicate that the item exhibits high levels of random error or that the item does not belong to the same domain as the other items in the model (Fabrigar and Wegener, 2012, p. 134). The index of variable complexity in a multidimensional model indicates the number of common factors involved with a particular variable (Hofmann, 1978). The index ranges from unity to the number of factors, and values close to unity indicate that, in general, the item is related to a single factor. Article PVI examined an existing a priori scale and, thus, EFA was not applied in that research.

Exploratory graph analysis An approach in network psychometrics, exploratory graph analysis (EGA), combines the methods behind latent variable and network models (Epskamp, Rhemtulla, and Borsboom, 2017). As proposed by Golino et al. (2020), EGA creates a weighted network by applying the Gaussian graphical model (Lauritzen, 1996) with graphical least absolute shrinkage and selection operator (GLASSO) estimation (Friedman, Hastie, and Tibshirani, 2008) or by applying the triangulated maximally filtered graph approach (TMFG) (Massara, Di Matteo, and Aste, 2016). Nodes represent psychometric objects (e.g., scale items) and edges in the network represent the associations (e.g., partial correlations) between nodes. When the network is identified, a community detection algorithm is applied. Communities of nodes are detected using a Walktrap algorithm (Pons and Latapy, 2006). Golino et al. (2020) proposed an approach for mitigating the Walktrap algorithm's tendency to penalize unidimensional structures by adding simulated unidimensional data to the original data when analyzing unidimensional structures. The communities of nodes are suggested to depict the dimensions in the data (Golino et al., 2020) and, as a result, EGA provides a graphical representation of the structure of the network.

Confirmatory factor analysis Confirmatory factor analysis (CFA) is a structural equation modeling technique that deals with measurement models (e.g., construct validation, psychometric evaluation of measurement instruments) (Brown, 2015, p. 25). For example, CFA can be used—as the name suggests—to confirm the structure defined in an exploratory analysis. Thus, optimally and according to good practice, CFA should be conducted for an independent new data set not used in EFA.

The common factor model is based on the linear combinations of the factor loadings λ and factors η for the indicator y_j and m factors (Brown, 2015, p. 17):

$$y_j = \lambda_{j1}\eta_1 + \lambda_{j2}\eta_2 + ... + \lambda_{jm}\eta_m + \epsilon_j,$$

which can be expressed in a matrix form as

$$\Sigma = \Lambda_y \Psi \Lambda_y^{'} + \Theta\epsilon,$$

where Σ is the correlation matrix of the indicators, Λ is the matrix of factor loadings, Ψ is the correlation matrix of the factor correlations in case of non-orthogonal rotation, and $\Theta\epsilon$ is the matrix of unique variances. Different estimators are used to estimate the CFA model. For ordinal data (e.g., Liker-type data) a suggested approach for estimation is to use polychoric correlation (Holgado–Tello et al., 2010) and robust diagonally weighted least squares (DWLS) estimation with test statistics adjusted in terms of mean and variance (i.e., scale-shifted approach, a.k.a., WLSMV) (Beauducel and Herzberg, 2006; DiStefano and Morgan, 2014; Foldnes and Grønneberg, 2021; Forero, Maydeu-Olivares, and

Gallardo-Pujol, 2009; C.-H. Li, 2016a,b, 2021; El-Sheikh, Abonazel, and Gamil, 2017). After estimation, the factor score estimates for each observation can be calculated using non-refined (e.g., sum scores by factor) or refined methods (e.g., model-based approach) (e.g., DiStefano, Zhu, and Mîndrilã, 2009; Grice, 2001, p. 305–306; Bollen, 1989; Hershberger, 2014).

Different goodness-of-fit indices are used to evaluate the model fit in factor analysis. The goodness-of-fit indices provide "a global, descriptive indication of the ability of the model to reproduce the observed relationships among the indicators in the input matrix" (Brown, 2015, p. 97). The common absolute goodness-of-fit measure χ^2 is stringent and sensitive, for example, to sample size (Brown, 2015). χ^2 test is based on the normal distribution, and it renders a significant result when the model does not fit the data. It commonly rejects the null hypothesis (i.e., H_0 implies the model fits perfectly), meaning the model does not fit perfectly and, thus, a set of other kinds of indices are recommended in addition (Brown, 2015, p.67–73). There exist other commonly used fit indices: standardized root mean squared residuals (SRMR) is an indicator of the absolute fit, the root mean square error of approximation (RMSEA) is an indicator of the parsimony corrected fit, and the comparative fit index (CFI) and Tucker—Lewis index (TLI) is indicators of the comparative fit.

SRMR was initially selected as the basis for the combinational rule-based "2-index presentation strategy" in the highly influential paper by Hu and Bentler, where it was supplemented with another fit index to evaluate the model fit (Hu and Bentler, 1999). SRMR is calculated based on the residual correlation matrix, and it is the square root of the average of the squared standardized variances and covariances of the residuals (Brown, 2015, p. 70). In other words, it is based on the difference between the model-implied matrix and the covariance matrix, which makes it relatively insensitive to different estimators and appropriate to use in the case of ordinal models (Shi and Maydeu-Olivares, 2020). In the case of five or six categories (as in ATI), the SRMR based on WLSMV estimation was not found to be very different from SRMR based on ML (Beauducel and Herzberg, 2006).

RMSEA is an χ^2-based parsimony corrected index, which is affected by the number of the model parameters. More precisely, RMSEA compensates for model complexity by favoring models with fewer freely estimated parameters (i.e., mode degrees of freedom), is relatively insensitive to sample size, and "assesses the extent to which a model fits reasonably well in the population" (Brown, 2015, p. 71–72). As the names suggest, comparative fit indices CFI and TLI compare the actual model to a baseline model, which assumes no relationships among the variables. The indices are also χ^2-based, and TLI compensates for model complexity similar to RMSEA (Brown, 2015, p. 71–72).

However, various aspects relating to, for example, estimation method, sample size, and model complexity, affect the choice of what cutoff values could be used to confirm a model is having an acceptable fit (Brown, 2015, p. 74; Shi and Maydeu-Olivares, 2020). To make the analytical decision making based on the cutoff values of the fit indices more intricate from the ordinal CFA estimation point of view, for example, it is worth noting that the highly influential cutoff values suggested by Hu and Bentler (Hu and Bentler, 1999) were developed for maximum likelihood (ML) -based estimation. Cutoff values designed to work for ML-based, continuous, and normal data might not unequivocally generalize to other estimators (e.g., DWLS) suggested for ordinal and nonnormal data (Beauducel and Herzberg, 2006; Savalei, 2018; Shi and Maydeu-Olivares, 2020; Xia and Yang, 2019). Despite the extensive use of the fit indices in the literature, there exists no "golden rules" or cut off values that determine the fit or misfit of a model (e.g., Greiff and Heene, 2017; Shi, T. Lee, and Maydeu-Olivares, 2019) and various values and rules have been proposed (e.g., TLI or CFI > 0.95 and SRMR < 0.09 (Hu and Bentler, 1999), dynamic fit index (McNeish and Wolf, 2021)).

Instead of relying purely on the fit indices, it is a good practice to examine the local model fit using the standardized residual covariance matrix. The residual covariance matrix is the difference between the sample and model-implied matrices. Absolute values of standardized residuals over 1.96 are usually considered statistically significant at the $p < 0.05$ level (Brown, 2015). However, a large sample size might also be associated with larger standardized residuals (Brown, 2015). Therefore, it is recommended to report the absolute values of the six largest standardized residual covariances (Maydeu-Olivares, 2017). A confirmatory model can also be scrutinized by exploratory means using model modifications. Modification means applying, for example, correlated errors to the a priori model and estimating the model again. Modification indices (i.e., Lagrange multipliers) and the expected parameter changes can be used to examine the local misspecifications (e.g., Greiff and Heene, 2017). However, the applied modifications should be justified based on logical reason or theoretical assumptions (Brown, 2015).

Mokken scale analysis The aim of Mokken scale analysis is to determine how the analyzed scale compares with an ideal scale given a specific definition of error. Mokken (1971, p. 42–43) defined scalability of a set of scale items as "the degrees to which a set of items may be said to fit the model of a perfect scale." The perfect scale refers to the Guttman model (Guttman, 1950), which underlies a deterministic unidimensional scale where a person is expected to agree or respond correctly to a less difficult item if the person has agreed or responded correctly to a more difficult item. The error in the Guttman model context refers to deviations from the response mentioned above pattern (e.g., Goodenough, 1955; Suchman, 1950).

The scalability (i.e., goodness-of-fit) can be assessed using different coefficients of scalability and for dichotomous data Mokken (1971, p. 59) chose to use the coefficient of homogeneity H (Loevinger, 1947, 1948; Mokken, 1971, p. 148–153). Molenaar (1991, 1997) extended the method to cover ordered polytomous categorical variables. For polytomous item pairs, the extended method is based on the weighted count of Guttman errors given marginal distribution. In other words, the version of the coefficient H by Molenaar (1991, p. 97) "is always equal to the ratio of the correlation between the two item scores and the maximum possible correlation given the marginal distributions per item." Consequently, for item pairs j and k, the coefficient of homogeneity H_{jk} is defined as the ratio between the covariance of the item pair and the maximum possible covariance (Sijtsma and Ark, 2020, p. 126; Ark L, 2007):

$$H_{jk} = \frac{\sigma_{jk}}{\sigma_{jk}^{\max}},$$

where the maximum of the covariance is obtained by finding a joint distribution of the item pair with a given marginal distribution which corresponds the perfect Guttman pattern (Molenaar, 1991).

For individual item j, the coefficient H_j is the sum of the covariances of item j with the rest of the items divided by the corresponding maximum value or, stated differently, the covariance between the item score with the rest score, $R_{(j)} = \sum_{k:k\neq j} X_k$, divided by the corresponding maximum value (Sijtsma and Ark, 2020, p. 126):

$$H_j = \frac{\sum_{k:k\neq j} \sigma_{jk}}{\sum_{k:k\neq j} \sigma_{jk}^{\max}} = \frac{\sigma(X_j, \sum_{k:k\neq j} X_k)}{\sigma^{\max}(X_j, \sum_{k:k\neq j} X_k)} = \frac{\sigma(X_j, R_{(j)})}{\sigma^{\max}(X_j, R_{(j)})} = \frac{\rho(X_j, R_{(j)})}{\rho^{\max}(X_j, R_{(j)})}.$$

As the last part of the equation above denotes, H_j can be also stated using the corrected item-total correlations, $\rho(X_j, R_{(j)})$. For the complete scale, the coefficient H is the sum of the item-rest covariances divided by the sums of the maximum covariances of the item pairs:

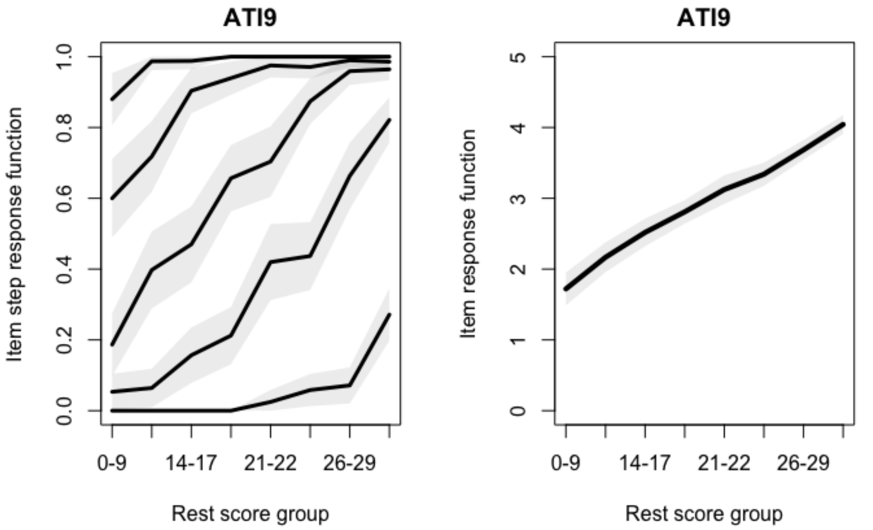
$$H = \frac{\sum_{j=i}^{J} \sigma(X_j, R_{(j)})}{\sum\sum_{j\neq k} \sigma_{jk}^{\max}}.$$

In essence, H refers to the degree the scale resembles a perfect Guttman scale. According to Mokken (1971, p, 185), the coefficient of homogeneity H is also a criterion of scalability for the complete scale in terms of MHM, and $H \geq 0.50$ refers to a strong scale, $0.40 \leq H < 0.50$ referst to a medium scale, and $0.30 \leq H < 0.40$ refers to a weak scale. Coefficient H is also used to define a scale: "a scale is a set of items that are all positively correlated and with the property that every item coefficient of scalability (H_i) is larger than or equal to a given positive constant (c)" (Mokken, 1971, p. 184). A constant $c = 0.30$ is commonly used for the lower bound (Mokken, 1971, p. 184; Sijtsma and Ark, 2020, p. 129).

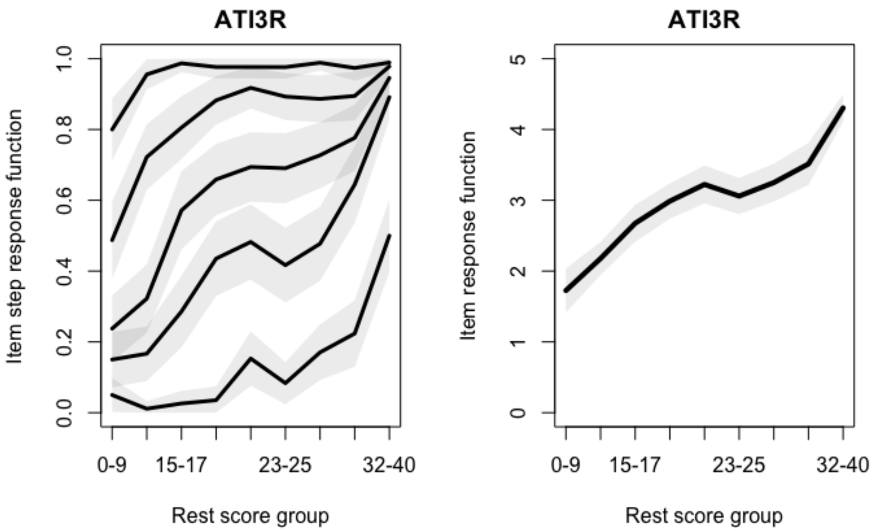
Mokken scale analysis is based on the monotone homogeneity model (MHM), which has three underlying assumptions: unidimensionality, local independence, and monotonicity (Sijtsma and Ark, 2020). In the Mokken scale context, unidimensionality is closely related to the idea of scalability. If the scale items form a Mokken scale, they are thought to measure the same latent variable (Molenaar, 1991). Local independence means that the variance of the scale items is caused only by the variance in the latent variable. Local independence can be assessed using the conditional association (CA) procedure which is based on indices of conditional covariance (Straat, Ark L, and Sijtsma, 2016). Indices defined as outliers indicate the possible existence of local dependency.

The monotonicity assumption can be assessed visually by plotting (Figure 4) the individual item step response functions (ISRFs) and item response functions

(IRFs) including their binomial proportion confidence intervals (a.k.a., Wald confidence interval) (Ark L, 2013). The response functions should be monotonically nondecreasing (Figure 4a) and there should not be many significant violations of monotonicity among the scale items (Figure 4b). Furthermore, by definition, the ISRFs within the same item should not intersect (Sijtsma and Molenaar, 2002). A special case of MHM, and a more strict model, a double monotonicity model (DMM), also assumes that the ISRFs of different items do not intersect (Sijtsma and Molenaar, 2002).



(a) The response functions of ATI9 showed a clear monotonic pattern as assumed for MHM.



(b) The response functions of a reverse-worded item ATI3R showed deviations from monotonicity.

FIGURE 4 The ISRFs and the IRFs of items ATI9 and ATI3R with Wald 95% confidence intervals analyzed in Article PVI.

Partial credit model In latent variable analysis, the partial credit model (PCM) is a parametric item response theory model for polytomous scored data belonging to the class of adjacent category models (Masters, 2016; Masters, 1982; Sijtsma and Pfadt, 2021). PCM was chosen to be used in Article PVI as it is the least restrictive of the parametric IRT models still having the property of stochastic ordering of latent variable by the total scale score (i.e., SOL by X_+) (Hemker et al., 1997; Sijtsma and Hemker, 2000). The model estimates the probability that a respondent has the item score x instead of the previous score $x-1$, and the probability should increase monotonically with respect to latent variable (Masters, 2016, p. 109–112; Sijtsma and Pfadt, 2021, p. 216). PCM is constructed using Rasch's dichotomies of adjacent categories (Masters, 2016, p. 111; Sijtsma and Pfadt, 2021, p. 216):

$$\frac{P(X_j = x|\theta)}{P(X_j = x-1|\theta) + P(X_j = x|\theta)} = \frac{e^{(\theta-\delta_{jx})}}{1+e^{(\theta-\delta_{jx})}},$$

where X_j is the person's response category for item j, θ is the person's value of a latent variable (i.e., person's ability), and δ_{jx} is the location parameter of the item step response function (ISRF) of the item j for the category x. PCM can be visualized using the ISRFs, which depict the most probable response categories (Masters, 2016). The ISRFs should show an ordered structure as described in Article PVI. The goodness of fit of the PCM can be assessed using unweighted mean-square (a.k.a., outfit) and weighted mean-square (a.k.a., infit), which are based on the standardized differences between observed and expected responses of a person (Masters, 2016, p. 118–119).

Differential item functioning Study subjects exhibiting the same latent variable level are expected to respond similarly to an item in an instrument. Differential item functioning (DIF) occurs "when people in different groups perform differently on an item, even though the people have been matched on some relevant criterion" (Dorans and Cook, 2016, p. 23). For example, an item is is easier or more difficult for some respondents than others. In other words, an item shows different statistical properties with respect to different groups (Angoff, 1993). In practice, DIF means that an item is more difficult (e.g., to answer correctly, to agree with) to a certain group of subjects. DIF needs to be carefully examined and assessed as it could affect to fairness and impartiality of the instrument; one would not want to have a measurement instrument that is biased and falsely discriminating when conducting scientific research, not to mention other high-stakes situations.

There exist two broad categories of DIF: uniform differential item functioning and nonuniform differential item functioning (e.g., Walker, 2011). Uniform DIF refers to a situation where the difference between groups remains the same for all levels of a latent variable. On the contrary, nonuniform DIF arises when the difference is not constant across all levels of a latent variable.

DIF is relatively difficult to detect, and several methods have been proposed for analyzing dichotomous and polytomous items using parametric and nonparametric methods (e.g., Walker, 2011). Article PVI demonstrated analyzing uniform

DIF concerning different age groups and gender. The applied method utilized a regularization approach based on the lasso principle and partial credit model (Schauberger and Mair, 2020).

Reliability

Reliability can be interpreted in two ways (AERA, APA, and NCME, 2014); First, it can be used to denote the reliability coefficient in terms of classical test theory (CTT). Second, it can be used more broadly to describe the consistency of scores when the measurement process is replicated. Also, Sijtsma and Pfadt (2021) reminded us that the definition of reliability is based on the idea of replicability, meaning a situation where the measurement procedure is repeated under the same circumstances.

The underlying assumption of the CTT mode is that a person's observable measurement score X is the sum of the expected true score τ and some error random variable E (Lord and Novick, 1968, p. 55):

$$X = \tau + E.$$

Furthermore, the CTT model assumes that the expected value of E is zero and E is uncorrelated with the true score τ and other error random variables. For parallel measurements X and X' when replicating the measurement procedure, the observed scores are assumed to have identical true scores and linearly experimentally independent errors meaning the two measurements are uncorrelated (Lord and Novick, 1968, p. 45, 47). Deriving from the assumptions, Lord and Novick (1968) provide a detailed account on different quantities of how reliability can be defined. They emphasize that the reliability can be defined as the coefficient of determination $\rho^2_{X\tau}$ which is the squared correlation between the observed score and true score. Furthermore, they derive other definitions: "the reliability of a test is a measure of the degree of true-score variation relative to observed-score variation", "reliability is a number in the interval from 0 to 1", "reliability is an inverse measure of error variance relative to observed variance", and reliability is "equal to the correlation between parallel measures" (Lord and Novick, 1968, p. 55–61):

$$\rho^2_{X\tau} = \frac{\sigma^2_\tau}{\sigma^2_X} = 1 - \frac{\sigma^2_E}{\sigma^2_X} = \rho_{XX'}.$$

In practice, the true score can not be directly observed, and replicating the measurement procedure under the same conditions can be challenging. However, psychometric instruments are often composed of several items, which provides a way to approach reliability. A composite measurement instrument is composed of several parts called component measurements (Lord and Novick, 1968). When considering the reliability of a composite measurement, coefficient α provides an estimate of a lower bound to the reliability of a composite measurement

(Lord and Novick, 1968; Sijtsma and Pfadt, 2021):

$$\alpha = \frac{n}{n-1}\left(1 - \frac{\sum_{i=1}^{n}\sigma^2_{Y_i}}{\sigma^2_X}\right) \leq \rho_{XX'},$$

where σ^2_X is the variance of the composite observed score and $\sigma^2_{Y_i}$ is the variance of the i:th component. Sijtsma and Pfadt (2021) summarized the usefulness of coefficient α by stating that it is a mathematical lower bound to the reliability of a composite score and, in case of approximate unidimensional model, it is close to reliability, $\rho_{XX'}$. Coefficient α is widely used but often misunderstood (Sijtsma and Pfadt, 2021).

In addition to the CTT approach above, reliability can also be analyzed in the context of structural equation modeling (SEM). Green and Yang (2009) proposed a nonlinear SEM approach for estimating reliability for ordered categorical items which Kelley and Pornprasertmanit (2016) termed as categorical omega, ω_c. Omega for ordered categorical items is calculated using the sample estimates obtained from a SEM model estimated using polychoric correlations and robust weighted least squares estimation (Green and Yang, 2009). Kelley and Pornprasertmanit (2016) recommend using categorical omega with a bias-corrected and accelerated bootstrap confidence interval for estimating reliability in case data is ordered categorical. They emphasized that a congeneric unidimensional model is assumed when estimating reliability in the context of SEM using categorical omega. In SEM, congeneric models and more restrictive tau-equivalent models are characterized by equality constraints, and the congeneric model is more common in applied research (Brown, 2015, p. 207–208). Factor loadings and error variances are allowed to vary in the congeneric model. The tau-equivalent model is more restrictive than the congeneric model, and it constrains the factor loadings to be equal while allowing for varying error variances.

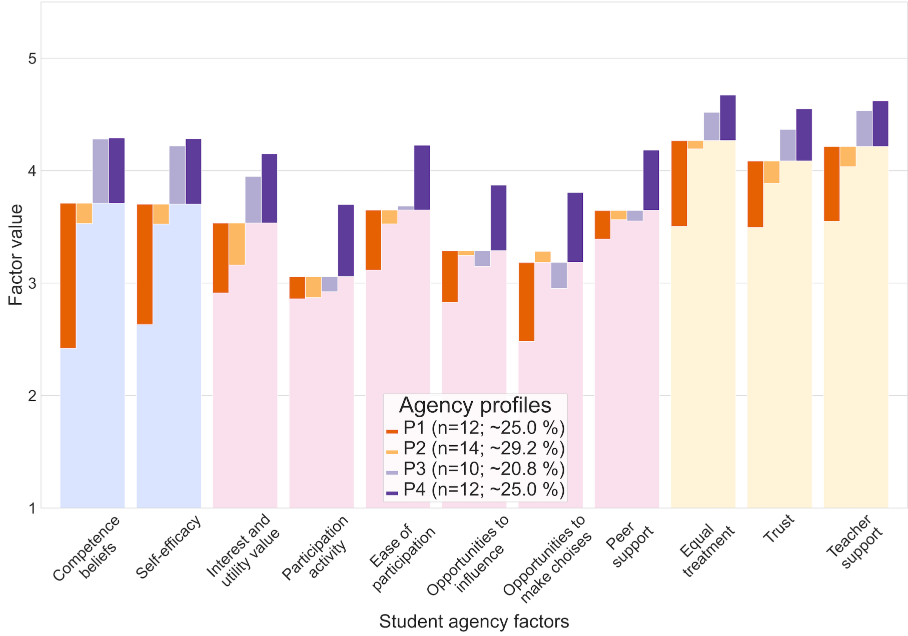
3.2 Robust clustering

When conducting analyses of statistical inference, one has to consider the validity of different underlying assumptions relating to, for example, data-generating processes and analysis procedures. If there is a possibility that there exist deviations from the assumptions, one has the option to utilize robust, nonparametric, and distribution-free procedures. Huber (1981) makes distinctions between the procedures and defines that "robustness signifies insensitivity to small deviations from the assumptions." A nonparametric procedure refers to situations where the procedure can be applied for a range of distributions having no specific parameters, and a distribution-free means that a statistical test falsely rejects the null hypothesis for all underlying continuous distributions with the same probability (Huber, 1981, p. 6). For example, the sample median is a robust and nonparametric estimate of the population median that is insensitive (i.e., robust, resistant) to changes in the observed values in a sample (e.g., Huber, 1981; Kärkkäinen and Heikkola, 2004; Rousseeuw and Hubert, 2018).

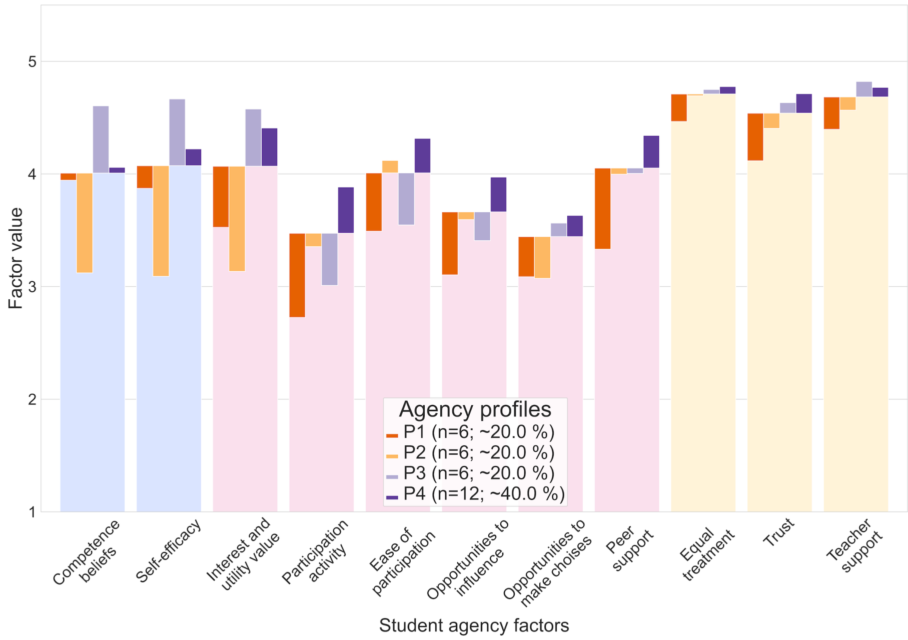
Clustering is an unsupervised machine learning method that aims to form groups (i.e., clusters) of similar observations based on some distance measure (Jain, Murty, and Flynn, 1999). The observations can then be labeled and described according to the characteristics of the group they were assigned to. A procedure called k-means is the simplest and one of the most popular partitional clustering method aiming to minimize the squared error criterion (Jain, Murty, and Flynn, 1999; MacQueen, 1967). The basic k-means algorithm *i*) selects k centers for clusters in an N-dimensional space based on the possible number of clusters in N-dimensional data, ii) assigns each observation to the closest cluster center based on a distance measure, *iii*) calculates the new cluster centers based on the assigned observations, and *iv*) continues to update the cluster centers and assigning the observations until a convergence criterion is attained.

The student agency analytics process utilizes a variant of the k-means procedure based on the spatial median. The spatial median (a.k.a., geometric median) generalizes the median to a multidimensional space and minimizes the sum of euclidian distances between the observations and their location estimate. The k-spatialmedians algorithm calculates the cluster centers exploiting the spatial median instead of the mean, making it robust against outliers and missing data (Hämäläinen, Kärkkäinen, and Rossi, 2018; Kärkkäinen and Äyrämö, 2004). The initialization method (i.e., selecting the initial cluster centers) is an important step in the algorithm affecting the efficiency and accuracy. A popular approach for initialization is to select the initial clusters based on the probability proportional to the squared distances to the centers (i.e., k-means++) (Arthur and Vassilvitskii, 2007; Ostrovsky et al., 2006). Specifically, the student agency analytics process utilizes the k-spatialmendians++ algorithm by applying clustering based on the spatial median and probability-based initialization described in Article PI.

A cluster validation index (CVI) is a measure that quantifies the quality of the clustering result. The internal examination of validity refers to methods that examine how well the clustering structure fits the data intrinsically (Jain, Murty, and Flynn, 1999). Hämäläinen, Jauhiainen, and Kärkkäinen (2017) analyzed different internal clustering validation indices for different variants of the k-means algorithm. One of the suitable internal CVIs for k-spatialmedian they proposed is the Pakhira-Bandyopadhyay-Maulik (PBM-index) cluster validation index which was used in Article PI. PBM-index takes into account the number and compactness of the clusters, the maximum separation between a pair of clusters, and "ensures the formation of a small number of compact clusters with a large separation between at least two clusters" (Pakhira, Bandyopadhyay, and Maulik, 2004, p. 487). Figure 5 shows en example of clustering results provided for the teachers in Article PV. These figures were omitted from the original article because of space limitations.



(a) Student agency profiles in Helen's course in PV



(b) Student agency profiles in Katie's course in PV

FIGURE 5 Student agency analytics results presented as a bars-in-bar graph show the general average profile and the deviations of the four profiles from the average profile.

3.3 Explainable artificial intelligence

Algorithms of artificial intelligence are capable of handling loads of high-dimensional data and conducting complex analyses. Unfortunately, the downside is that we humans seldom fully comprehend the intricate rules behind the often very complex and esoteric algorithms. A branch of research in AI, explainable artificial intelligence (XAI), aims to cast light inside the black boxes of the algorithms.

Explainable artificial intelligence has gained traction quite recently (e.g., Adadi and Berrada, 2018). However, the idea behind explainable AI is not new. Already Shortliffe et al. (1975) introduced a medical recommendation system having an interactive explanation capability. The system aimed to give physicians advice on suitable antimicrobial medication for patients with bacterial infections. They made several essential observations relating to the purpose and use of such systems (Shortliffe et al., 1975, p. 309–310):

> Physicians often voice pessimism about the potential usefulness of a computer-based diagnostic or consultation system, asserting that few clinicians will ever be willing to place life-and-death decisions in the hands of a computer. Many clinicians feel that if errors are going to be made, they would prefer to have made the mistakes themselves rather than to have put misplaced confidence in a machine. It is our belief, therefore, that a consultation program will gain acceptance only if it serves to augment rather than replace the physician's own decision making processes. ... An important way to emphasize a program's role as a helpful tool, and to establish its credibility, is to permit the clinician to evaluate the program's advice before he acts upon it. Such a capability permits the physician to reject advice which he feels is based upon incomplete or incorrect decision criteria. In addition, the capability can serve an educational role by pointing out decision rules that the physician may wish to incorporate into his own knowledge of clinical medicine.

The quote above highlights that the purpose of the AI system was not supposed to replace human decision-making. Instead, the system was thought to provide interpretable results and a chance for professional development. When considering from teachers' perspectives, explainable student agency analytics (XSAA) shares the same aim. Article PIV applies XAI methods to the process of student agency analytics.

Understandability in the context of AI is a "characteristic of a model to make a human understand its function – how the model works – without any need for explaining its internal structure or the algorithmic means by which the model processes data internally" and comprehensibility "refers to the ability of a learning algorithm to represent its learned knowledge in a human-understandable fashion" (Barredo Arrieta et al., 2020, p. 84–85). Transparency, interpretability, and explainability are distinct features that oppose the opaque—"black box"—models; Transparent models are potentially understandable, interpretable models can provide explanations understandable to humans, and explainable models relate to "explanation as an interface between humans" and AI algorithms (Angelov et al., 2021, p. 13). Generally, applicable XAI methods that can be used in various situations are called model agnostic as opposed to model-specific methods that are suitable for a limited number of situations (Angelov et al., 2021). XAI methods can utilize several approaches for disclosing the inner workings of an algorithm. For example, explanation by feature relevance means that the XAI method points out the important features relevant for the AI algorithm's internal decision-making and local explanations provide insight of the algorithm's operational principles around a specific observation (Angelov et al., 2021). Articles PI, PII, and PIV utilized explanation by feature relevance, and Article PIV also applies a method relating to local explanations.

SHapley Additive exPlanation (SHAP) (Lundberg and S.-I. Lee, 2017) is a XAI method utilizing Shapley values (Shapley, 1953) originating from cooperative game theory. It is a feature-based approach where SHAP values "attribute to each feature the change in the expected model prediction when conditioning on that feature" (Lundberg and S.-I. Lee, 2017, p. 5). Another XAI method, local interpretable model-agnostic explanations (LIME) (Ribeiro, Singh, and Guestrin, 2016), examines the features important for predicting a specific observation (i.e., local explanation). The LIME approach creates a surrogate model to analyze the local behavior of the original model around the specific prediction (Angelov et al., 2021). LIME and SHAP methods were used in Article PIV.

Some machine learning algorithms provide at least a certain degree of transparency natively. For example, the relevance of the features in a random forest (Breiman, 2001) classification model can be quantified using the impurity importance (a.k.a, the mean decrease in impurity (MDI), Gini importance) (Nembrini, König, and Wright, 2018). Furthermore, a traditional method like Kruskal-Wallis H (Kruskal and Wallis, 1952) can be used to quantify the feature separation in a clustering approach (Cord, Ambroise, and Cocquerez, 2006; Saarela, Hämäläinen, and Kärkkäinen, 2017). Impurity importance relating to a random forest model was used in Article PIII, and Kruskal-Wallis H was used in Articles PI and PIII.

3.4 Interpreting qualitative data

One of the best ways to examine persons' individual experiences is to ask them. The approach above was utilized in three articles: Articles PII and PIII analyzed students' open-ended questionnaire responses, and Article PV dealt with transcribed semi-structured interviews of the teachers interpreting their student agency analytics results. Braun and Clarke (2006) claim that there exists no perfect theoretical framework for conducting qualitative research, and the central aspect to consider is choosing methods that align with the research questions. Thematic analysis was chosen for the analysis method for the open-ended responses as the main aim was to identify topics and themes from the data. Pentadic analysis—a form of narrative analysis—was used to analyze teachers' interviews as the main aim was to find out how teachers reflected on their pedagogical actions.

My stance is that it is inadequate to examine the complex psychosocial phenomena merely in terms of quantity. Assuming a pragmatist approach (Onwuegbuzie and Leech, 2005), I appreciate the interpretivist view and methodological pluralism. Quality is an essential aspect when examining the reality which is accessed "through social constructivism such as language, consciousness, shared meanings, and instruments" (Myers, 2020, p. 45). However, it is important to consider that the researcher's interpretations are affected by subjective bias. Thus, qualitative analyses were processed by two researchers to reduce subjective bias and enhance intercoder reliability.

Thematic analysis Thematic analysis is a qualitative method that aims to identify, analyze, and report patterns (i.e., themes) in the data (Braun and Clarke, 2006). Thematic analysis can be approached inductively (a.k.a., bottom-up) or using a theory as a basis (a.k.a., top-down). Braun and Clarke (2006, p. 84) point out that while "data are not coded in an epistemological vacuum", the inductive approach aims to link themes accurately to the data without specific theoretical presuppositions, which is a somewhat similar approach as in grounded theory. On the other hand, in a theory-driven approach, the analysis process is guided by the specific research questions and underlying theoretical assumptions (Braun and Clarke, 2006). Semantic level of the thematic analysis concerns with the surface meaning of the data and latent level of the analysis bores beyond the surface meaning by examining "the underlying ideas, assumptions, and conceptualizations—and ideologies—that are theorized as shaping or informing the semantic content of the data" (Braun and Clarke, 2006, p. 84). The thematic analysis process covers broadly the following phases: *i*) familiarize yourself with the data, *ii*) construct initial codes, *iii*) search for themes, *iv*) review themes, *v*) define themes, and *vi*) write the results (Braun and Clarke, 2006; Maguire and Delahunt, 2017).

I used thematic analysis in Article PII to analyze the open-ended written responses of the students assigned to the low agency profile. The central assumption was that there exist aspects that restrict students' agency from actualizing and aspects that enable or empower students' agentic experiences. The particular interest was the experiences of students in the low agency profile. In other words, student agency analytics was used as an analytical data reduction tool to focus on an interesting subset of respondents. Theory-driven approach with respect to the research questions and utilizing student agency as the underlying construct guided the formation of the themes. However, other ad-hoc themes were also allowed to emerge. The analysis concentrated on what the respondent explicitly wrote in the open-ended response (i.e., semantic level).

The data concerning restricting and empowering aspects of learning in Article PII consisted of short phrases organized as an Excel file. In the first analysis step, each of the phrases was assigned codes best describing the content of the phrase. To strengthen the validity of the results, the initial coding was conducted independently by me and the second author. After deriving the initial codes, we discussed the findings together, refined the codes as second-order codes, and constructed the themes emerging from the coding process. The results were visualized as a map where the second-order codes were grouped as themes. The themes were given a name that encompassed all the related codes. The links drawn between the codes of the same respondent depicted the qualitative associations between the codes and themes. The associations could provide interesting topics for future research. For example, a respondent stated that combining work and studying was stressful. The statement could provide research ideas relating to student well-being. Methodologically Article PII provides an example of how quantitative and qualitative analysis can be combined advantageously.

Content analysis Krippendorff (2013, p. 24) defines content analysis as ´´a research technique for making replicable and valid inferences from texts (or other meaningful matter) to the context of their use." The technique bears similarities with thematic analysis as Myers (2020, p. 317) defines content analysis as ´´a systematic approach to qualitative data analysis that looks for structures and patterned regularities in the text." The content analysis follows the top-down approach where the researcher constructs a set of codes that are then applied to the data (Myers, 2020). A common way to distill and represent the results is to count the number of occurrences of each code (Krippendorff, 2013).

Content analysis was used in Article PIII to classify the open-ended responses, and the quantified measures were used in the subsequent statistical analysis. In Article PV, content analysis was used to discover that the teachers pointed out more negative aspects than positive aspects when reflecting on their student agency analytics results from their course (Figure 6). First, the phrases containing reflections were identified from the transcribed interviews. Then the individual phrases were classified as containing negative content (e.g., self-criticizing, pointing out contradictions) and positive content (e.g., pointing out success). Similar approach was used to quantify the content relating to the student agency dimensions and profile groups. The results pointed out that two interviewees, John and Helen paid less attention to the profile groups and participatory resources than the other interviewees (Figure 6 and Figure 7). The figures mentioned above were omitted from the original article due to space limitations and, thus, presented here.

Pentadic analysis The narrative is "talk organized around consequential events" that the teller uses to explain to the listener what has happened in the past (C. Riessman, 2002, p. 219). Research interviews can be considered as one form of narrative. In research interviews, respondents "narrativize particular experiences in their lives, often where there has been a breach between ideal and real, self and society" (C. Riessman, 2002, p. 219). One analytical approach to study research interviews is to use Burke's dramatistic pentad (Burke, 1945), which is also called a pentadic analysis (Allen, 2017). Burke (1945, p. xv) states that human actions and motives can be described using the five core terms (i.e., the pentad): "what was done (act), when or where it was done (scene), who did it (agent), how he did it (agency), and why (purpose)." The pentadic analysis identified the key pentadic terms from the transcribed interviews and examined the meaningful links between the terms (i.e., ratios). The ratios are ordered pairs of the pentadic terms, and they can point out interesting structural points in a narrative (Allen, 2017; F. D. Anderson and Prelli, 2001; Burke, 1945; C. K. Riessman, 1993). The ratios can also bring out imbalance and tension relating to the narrativized experiences.

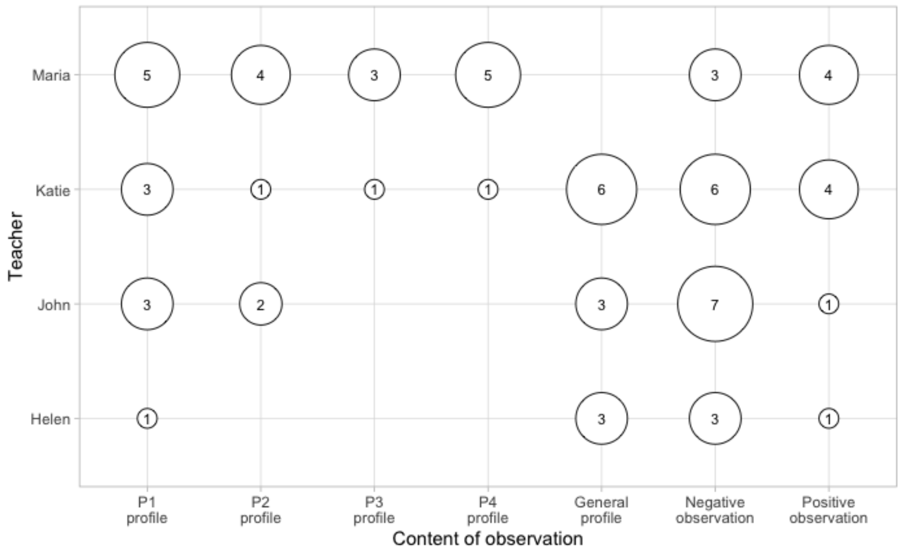


FIGURE 6 The number of reflections relating to the profile groups and valence of the reflection

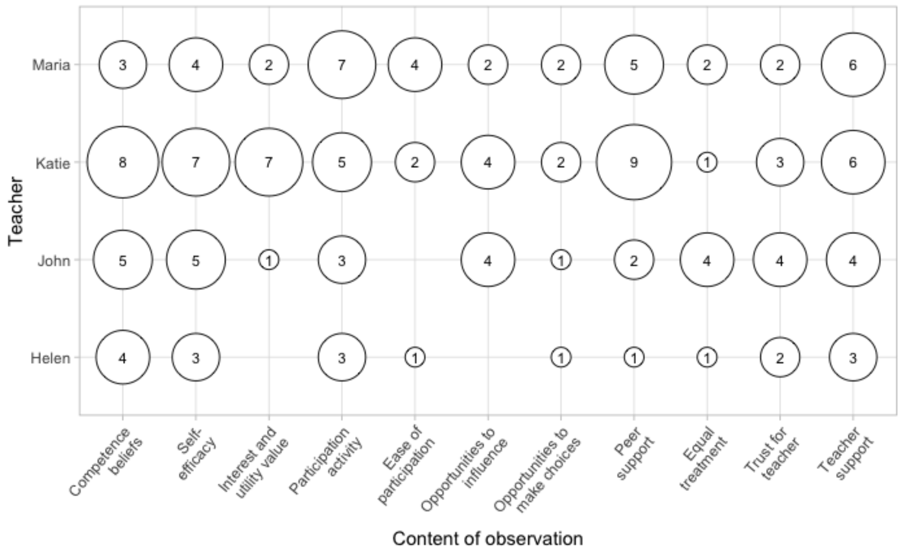


FIGURE 7 The number of reflections relating to the student agency dimensions

Bruner (1991, p. 16) outlines that the balance between the terms is determined by cultural conventions and "when conventional expectation is breached, Trouble ensues." For example, student-centered educational conventions dictate certain kinds of expectations (e.g., freedom to choose from different options, support for individual goals and difficulties), and the interesting observations could be identified by finding imbalanced pentadic ratios.

Pentadic analysis was used in Article PV to analyze teachers' reflection-on-action (see, D. A. Schön, 1983) based on the student agency analytics results they received from their courses. The analytics results functioned as a form of stimulated recall where the teachers discussed and reflected on their own actions and students' actions. The aim was to determine what student agency dimensions the teachers paid attention to and what kinds of reflections they highlighted.

3.5 Ethical considerations

Ethical issues are important to consider in all research having direct or indirect effects on humans. Tzimas and Demetriadis (2021) identified technological, pedagogical, and institutional dimensions consisting of six key ethical issues in learning analytics: privacy, transparency, labeling, data ownership, algorithmic fairness, and the obligation to act. Pargman and McGrath (2021) found out that transparency, privacy, and informed consent are the top three topics most often addressed in learning analytics literature. Kitto and Knight (2019) highlight the questions of how to balance consequential risks of harm and benefits of learning analytics, how to deal with conflicting regulatory principles and frameworks, and how to relate to learning analytics as research and as an institutional intervention.

The key issue is that learning analytics deals with personal and potentially sensitive data. Personal data refers to any data that can be directly or indirectly attributed to an identifiable person. Special categories of personal data (i.e., biometric data, health data) should be handled extra carefully. Data is the source of innovations and a highly appreciated resource in many fields of science and business. Personal data can "tell some of the most personal stories" of our lives (Mikk, Sleeper, and Topol, 2017). In learning analytics, the data is mainly generated by the learners and—like in the case of health data (e.g., Mikk, Sleeper, and Topol, 2018)—one might argue that the individuals themselves should control the data.

Furthermore, data processing should be transparent, unbiased, and based on valid methods. As per the General Data Protection Regulation (GDPR), people have the right to obtain an explanation when they are exposed to automated analysis and decision making (European Union, 2016). A. Nguyen et al. (2021) found out that majority of the students using a learning analytics information system expressed concerns relating to ethical issues like the transparency of the process, data security, and data storage. In general, data protection, privacy, and other ethical aspects should be taken into account by design and by default when developing learning analytics processes and applications (Heilala, 2018). Therefore, it is essential to consider privacy and data ownership.

Privacy The right to privacy is a fundamental human right that is secured, for example, by The Constitution of Finland (Finlex, 1999) and The Universal Declaration of Human Rights (Nations, 1948). A notable legal framework concerning privacy and personal data is GDPR (European Union), which sets the stage for data controlling and processing, and that has several implications on learning analytics (Heilala, 2018). However, the complex nature of learning analytics could pose challenges for applying regulative frameworks (Kitto and Knight, 2019).

Ifenthaler and Schumacher (2016) propose that students evaluate the potential risks and benefits of learning analytics, and the decision to disclose information is affected by risk-minimizing (e.g., control over data) and risk-maximizing factors (e.g., non-transparency). As learning-related data reflects personal behavior attitudes and possibly contain information about cognitive strengths and weaknesses, it is essential to address privacy by design and by default. Student agency analytics process aims to protect students' privacy by utilizing data aggregation: individual analytics results are provided only for the students, and the teachers receive the profiled results from which the individual students can not be identified. In other words, the teacher's visualized results are a compromise between privacy and data granularity.

Data ownership There exist discussions about whether the data and information can be a subject of property in the first place and should the issue be addressed more from the security and legitimate use perspective (Liddell, Simon, and Lucassen, 2021). Be as it may, as Asswad and Marx Gómez (2021) conclude, "data ownership is always dependent on the case and its requirements, including legal, ethical, organizational, and technical requirements."

Software architectural design decision (i.e., microservices) was proposed to be used as a technical solution for separating the data controller and data processor (e.g., Bolognini and Bistolfi, 2017; Ianculescu and Alexandru, 2020) as depicted in Articles PI, PII, and already in Heilala (2018). The data processor possessing the knowledge and algorithms for refining the data receives pseudonymized data, which provides a layer of privacy protection compared to using data without pseudonymization (e.g., Bolognini and Bistolfi, 2017; D. Schön and Ifenthaler, 2018). However, pseudonymized data is still personal data if there is any chance for the reidentification of a natural person. Furthermore, software architectural design can help protect the analysis algorithms and related intellectual property rights; providing learning analytics-as-a-service permits the use of analytics applications without being forced to release the algorithms and models (e.g., analytics engine (Ifenthaler and Greiff, 2021)) that have required a significant amount of research and development.

4 SUMMARY OF ARTICLES

In this chapter, I summarize the research aims, data, methods, and the main results of the articles. The details of each research can be found in the original articles, and my emphasis here is on the research contributions. There are several strategies that a researcher can use to make a theoretical contribution. The following list presents some strategies outlined by Jaccard and Jacoby (2020, p. 37–45) and how I connect the strategies with the contributions of my dissertation:

– develop typologies/taxonomies (Article PI),
– extend an existing theory or idea to a new context (Articles PI and PIII),
– clarify, refine, or challenge the conceptualization of a variable/concept (Articles PII, PIV, and PV),
– identify nuanced functional forms of relationships (Articles PII, PIII, and PIV), and
– identify the boundary conditions of an effect of one variable on another (Article PIII).

4.1 Article PI: Presents the student agency analytics process

Jääskelä, P., Heilala, V., Häkkinen, P., and Kärkkäinen, T. Student Agency Analytics: Learning Analytics as a Tool for Analyzing Student Agency in Higher Education. *Behavior and Information Technology*, 40(8), 790-808, 2020, doi:10.1080/0144929X.2020.1725130.

Research aims This article was foundational to my dissertation as it presented the overall and novel process of student agency analytics which involves *i*) acquiring the student agency data using a psychometric scale, *ii*) deriving the agency dimension scores of individual students and analyzing the group data using a robust educational data mining method, *iii*) visualizing the results, and *iv*) depicting a service-based architecture for automating the process. The research aimed to demonstrate what kinds of characterizations of student agency could be found utilizing the process and how the results could inform pedagogical practices at the course level.

Data and methods Two different data sets were utilized in this research. The first data set containing the AUS scale responses (N = 270) of Finnish university students (used in the AUS scale validation study by Jääskelä, Poikkeus, Häkkinen, et al. (2020)) was used to develop the student agency analytics workflow. A second data set containing AUS scale responses (N = 208) of students from the faculties of information technology (n = 130) and teacher education (n = 78) was used to examine the applicability of the process.

The data were analyzed using nonparametric methods relying on robust statistics (i.e., robust clustering). Cluster validation indices were used to choose the optimal number of clusters (i.e., student agency profiles). Furthermore, supervised machine learning, in this case, a multilayer perceptron (MLP), was used to evaluate the contributions of student agency dimensions on course grades.

Main results The results based on the robust clustering technique and the cluster validation index (PBM) proposed that an optimal number of student agency profiles would be four. From the practical point of view, the model of four profiles could also provide enough information without being too complex. The profiles of the IT students and the students in teacher education showed different characteristics and separating student agency factors. The characteristics of the profiles could reflect the students' individual differences and the differences in educational and pedagogical practices. The supervised analysis suggested that the four most crucial agency factors contributing to course grades among the IT students were competence beliefs, self-efficacy, teacher support, and equal treatment. Clustering and software architectural choices could facilitate students' privacy and ethical data processing.

Research contributions This research was novel in the sense that it interdisciplinarily combined some distinct developments in the fields of education, psychometrics, software technology, and data science as a coherent proposition for an approach in learning analytics that was not done before. As a result, a typology of student agency profiles was derived using robust clustering: the four student agency profiles classify the students in terms of differences in agentic experiences.

My contributions I was involved in developing the analysis process, writing the initial draft and the final article. My main contributions were the development of the microservices-based architectural design and privacy-related issues. Furthermore, I collected the second data set and conducted the statistical analyses relating to associations between the student agency dimensions, the course grades, and the study groups. I invented the visualization method for presenting the student agency profile results.

4.2 Article PII: Examines aspects restricting student agency

Heilala V., Jääskelä P., Kärkkäinen T., and Saarela M. Understanding the Study Experiences of Students in Low Agency Profile: Towards a Smart Education Approach. In: El Moussati A., Kpalma K., Ghaouth Belkasmi M., Saber M., Guégan S. (eds) *Advances in Smart Technologies Applications and Case Studies. SmartICT 2019.* Lecture Notes in Electrical Engineering, vol 684, 2020, doi:10.1007/978-3-030-53187-4_54.

Research aims From the teacher's point of view, the students needing support and the students being successful in their studies are interesting groups to consider. The information relating to the students' specific support and the conductive aspects of learning provide insights a teacher could utilize in pedagogical decision-making. This research examined the experiences of students categorized to the low student agency profile. Specifically, the research aimed to find out what kinds of constraints the students in the low agency profile experience in their studies. Furthermore, the research exemplified how student agency analytics relates to smart educational systems in general.

Data and methods The data consisted of AUS scale responses of students (N = 292) studying in three departments in a Finnish public research university (Faculty of information technology, Centre for multilingual academic communication, Department of teacher education). In addition to AUS scale responses, the students were asked to list openly any aspects that were improving or restricting their learning in the course. The research approach used mixed methods. Student agency analytics as a quantitative method was used to derive the student agency profiles. The students' open-ended responses in the low agency profile were analyzed qualitatively using thematic analysis. In other words, clustering was used to focus on an interesting subset of respondents.

Main results The analysis of the profiles revealed a group of students that showed lower student agency scores in general in all dimensions—especially in personal and relational resources. Based on the visualization of the individual AUS scores, the students in the low agency profile seemed to be "falling behind" compared to other profiles. Based on the thematic analysis, the main issues for the students in the low agency profile were competence beliefs, self-efficacy, student-teacher relations, course contents, time as a resource, and student well-being. However, the results also pointed out the complex and interrelated graph of topics depicting the restricting aspects. From the learning analytics point of view, the ability of a teacher to acquire and utilize analytical information could provide a basis for "smart education", which considers the needs of different learners (i.e., idiographic approach).

Research contributions This research elucidated some of the nuanced, complex, and nonlinear relationships of the experiences among the students in the low agency profile. Also, it clarified the typology of a low agency profile by providing an overview of the possible restrictions the students encounter.

My contributions I invented the research idea, defined the research aims, collected the data, conducted the analyses, synthesized the thematic analysis results in collaboration with Dr. Jääskelä, created the visualizations, and wrote the initial draft of the article. I was the corresponding author and responsible for the submission and review process. I presented the article at an international conference.

4.3 Article PIII: Links student agency with course satisfaction

Heilala, V., Saarela, M., Jääskelä, P., and Kärkkäinen, T. Course satisfaction in engineering education through the lens of student agency analytics. *Proceedings of the 50th IEEE Frontiers in Education Conference (Conference proceedings: Frontiers in Education Conference)*, 2020, doi:10.1109/FIE44824.2020.9274141.

Research aims This article examined the links between the general course satisfaction and the student agency dimensions. For measuring course satisfaction, I decided to utilize the idea behind the net promoter score (see, Reichheld, 2003). The first thing was to determine if course satisfaction could be categorized similarly to the willingness to promote a product or a service. Next, we investigated the differences between the dimensions of student agency and the course satisfaction categories. Lastly, we identified the most important and critical dimension contributing to the general course satisfaction.

Data and methods Data consisted of AUS scale responses of $N = 293$ engineering students from a Finnish university of applied sciences. We measured net promoter score and general course satisfaction using 10-point single Likert items. Respondents were divided into three groups based on their general course satisfaction. We examined the differences in student agency dimensions between the three satisfaction groups using nonparametric statistics and four different classification methods.

Main results The results showed that the measure of general course satisfaction followed the idea of net promoter score because the measure was associated with the willingness to recommend the course to a fellow student. The main finding was that the lower experiences of students' agency resources were associated with lower general course satisfaction. The finding was not surprising; however, our research clarified the intricate relationships of the complex underlying phenomena. The three general course satisfaction categories (i.e., dissatisfied, neutral, and satisfied) were associated with the level of experienced resources of student agency. Lower levels of student agency indicated lower general course satisfaction. Kruskal-Wallis test statistics and the feature importances of the best classifier—random forest in this case—indicated that teacher support, trust for teacher, interest, and utility value, self-efficacy, and competence beliefs were the most important dimensions predicting general course satisfaction.

Not all agency dimensions were equally important. The results showed that relational resources (i.e., trust for the teacher, teacher support, and equal treatment) were critical, indicating that a smaller decrease in those dimensions would more likely affect the experienced general course satisfaction than a decrease in other student agency dimensions. Also, personal resources (i.e., self-efficacy and competence beliefs) and the dimension of interest and utility value proved to be essential features when predicting course satisfaction. Lastly, the results highlighted the nuanced and complex relationships between the agency dimensions and general course satisfaction, prompting further research.

Research contributions First, the research contributed by extending an existing idea to a new context. Net promoter score was developed for business and marketing (see, Reichheld, 2003). In this research, we applied a similar idea for measuring general course satisfaction. The research contributed to understanding the boundary conditions of how student agency affects general course satisfaction.

The findings in this article complement the qualitative findings in the article PII: as lower student agency was associated with lower course satisfaction, the students in the lower agency profile also reported a wide variety of difficulties in their learning, especially concerning personal and relational resources. From the practical point of view, the results shed light on how experiences of course satisfaction could be supported by enhancing student agency, especially in terms of personal and relational resources.

My contributions I invented the research idea, defined the research aims, devised the measurement and analysis strategy, collected the data, conducted all the statistical analyses excluding the classification analysis, created most of the visualizations, and wrote the initial draft of the article. I was the corresponding author and responsible for the submission and review process. I presented the article at an international conference.

4.4 Article PIV: Applies explainable artificial intelligence methods

Saarela, M., Heilala, V., Jääskelä, P., Rantakaulio A., and Kärkkäinen, T. Explainable Student Agency Analytics for Higher Education. *IEEE Access*, 2021, pp. 137444–137459, doi:10.1109/ACCESS.2021.3116664.

Research aims The relationships in educational data are diverse and multi-faceted. Thus, we need the understanding of the complex nature of educational processes at the levels of individuals and groups (e.g., Dawson et al., 2019; Jovanović et al., 2021). The complexity is often addressed by using complex analytical methods, and explainable artificial intelligence (XAI) aims to provide more understandable, comprehensible, and ethical models (e.g., Adadi and Berrada, 2018). This research aimed to integrate XAI techniques into the student agency analytics process. Specifically, the research aimed to clarify the most important characteristics of the different agency profiles (i.e., global explanations) and explain why an individual student was assigned to a particular profile (i.e., local explanations).

Data and methods The data consisted of AUS scale responses of engineering students (N = 141) from a Finnish university of applied sciences (a subset of the data used in PIII). The students in the subset were studying the same subject under four different teachers. Thus, we could consider the studied subject (mathematics) as "constant" for all students. The underlying assumption then is that the student agency emerged as a result of students' individual differences and the pedagogical practices in the four courses.

Firstly, the student agency analytics process was conducted to acquire the values of the student agency dimensions and the profile categories for each student. Secondly, three classifiers were trained to predict the student agency profiles based on the agency dimensions and course instances. Multinomial logistic regression was the linear classification method and the nonlinear classifiers were the multilayer perceptron (MLP) and the random forest. In other words, clustering as the unsupervised method first created a representation of the students' agentic experiences based on the AUS model. Subsequently, the supervised classifiers created a representation—or an explanation— of how the clustering functioned. The best-performing predictive models based on the independent test set (the random forest as a nonlinear classifier and the multinomial logistic regression with l_1 penalization as a linear classifier) were chosen to explain the clustering results. Thirdly, XAI methods Shapley additive explanations (SHAP) (Lundberg and S.-I. Lee, 2017; Shapley, 1953) and local interpretable model-agnostic explanations (LIME) (Ribeiro, Singh, and Guestrin, 2016) were used to interpret the best predictive model. Global explanations refer to the features explaining the complete model, and local explanations refer to the features explaining individual students' classification results. In general, the results can be considered as glimpses to the inner workings of the clustering algorithm.

Main results The XAI approach in this research introduced additional layers of modeling and, thus, complexity to the process. However, it would be very difficult or impossible for a human to comprehend the clustering process in a multidimensional (in this case, 11 dimensional) space. Thus, in the case of this research, XAI is a trade-off between a "black box" model and "comprehensibility" is attained through additional layers of modeling. In general, the results indicated that the explainable student agency analytics (XSAA) could provide fine-grained information about the agency profiles.

SHAP and LIME provided estimates for the feature importances. In other words, they indicated how strongly the agency dimensions and course instances contributed to the clustering results. Especially, the local explanations clarified the issue of how the clustering algorithm assigned individual students to the agency profiles. Also, the local explanations do not necessarily always resemble the global explanation. From the teacher's perspective, the explainable student agency analytics could provide insight for pedagogical planning. Furthermore, the ethical use of machine learning and legal regulation (e.g., GDPR) incorporates the right to receive an explanation. Thus, XAI methods in LA could facilitate transparent and ethical analytics. However, further research should be conducted to examine the XAI methods in LA.

Research contributions Based on the brief literature review presented in the article, the existing research combining XAI and learning analytics in higher education was found to be scarce. Thus, the main contribution of this research was to extend the idea and methods of XAI to the field of learning analytics in higher education. Also, the research clarified the typology of the student agency profiles and provided means to examine and identify the differences between profiles.

My contributions I collected the data set and developed the analytics scheme employing the subset concerning the same course context. I conducted student agency analytics for the data and the statistical analyses relating to the differences between courses. I synthesized and visualized the XSAA process and contributed to the pedagogical implications. I was involved in writing the initial draft and the final article.

4.5 Article PV: Analyzes teachers' perspective

Heilala, V., Jääskelä, P., Saarela, M., Kuula., A-S., Eskola, A., and Kärkkäinen, T. "Sitting at the stern and holding the rudder": Teachers' reflections on action based on student agency analytics in higher education. In Leonid Chechurin (Ed.). Digital Teaching and Learning in Higher Education: Developing and Disseminating Skills for Blended Learning, *Palgrave Macmillan*, forthcoming.

Research aims The purpose of student agency analytics is to provide information relating to agentic resources in learning to students and teachers. The information could facilitate, for example, reflection in teachers' professional practice. This research considered the student agency analytics from the teacher's point of view. The main aim was to examine how four teachers in higher education reflected on their pedagogical actions based on the analytics results.

Data and methods Four teachers in a Finnish university of applied sciences each taught a course about mathematics for engineers. The courses were the same courses that were analyzed in Article PIV. The teachers received student agency analytics results after their courses. The teachers participated individually in a semi-structured interview in which they interpreted the results and tried to relate them to their pedagogical actions and observations during the course. The data consisted of transcribed interviews of the four teachers. The interview data were analyzed using pentadic analysis.

Main results The results indicated that the teachers were able to reflect on their pedagogical actions by using student agency analytics results as the starting point for their professional thinking. The qualitative findings pointed toward the complex nature of student agency. In their interviews, the teachers pointed out that the student agency dimension could be contingent on each other, teacher's actions could both promote and prevent student agency, temporality plays a role in student agency, and a student exhibiting high agentic resources might not eventually need pedagogical guidance. Also, they proposed actions they could have done differently, which indicates reflective thinking. Lastly, one teacher used a boat metaphor to describe her purpose in the classroom: in her view, a teacher is "sitting at the stern and holding the rudder" and the students' places in a boat describe what kinds of agents they are in a course.

Research contributions This research is the first examination and an example of how teachers utilized student agency analytics as a tool for professional reflection. The research clarified how the teachers interpret student agency analytics results and utilize them in their reflective thinking.

My contributions I invented the research idea, defined the research aims, constructed the outline of the interview in collaboration with the second author, devised the analysis strategy, transcribed the interviews, conducted the initial analyses, and wrote the initial draft of the article. I was the corresponding author and responsible for the submission and review process.

4.6 Article PVI: Demonstrates psychometric methods

Heilala, V., Kelly, R., Saarela, M., Jääskelä, P., and Kärkkäinen, T. The Finnish version of the Affinity for Technology Interaction Scale (ATI): Psychometric properties and an examination of gender differences. *International Journal of Human—Computer Interaction*, 2022, doi:10.1080/10447318.2022.2049142.

Research aims Measurement is fundamental to the sciences. In education and psychology, scales are essential measurement tools. Unfortunately, translated and comprehensively analyzed scales can be scarce in small language regions such as Finnish. Thus, translating or adapting an existing measure is often the first step when starting research work in psychological and educational assessment (e.g., Ziegler, 2020).

The main aim of this research was to create a Finnish translation and present a comprehensive analysis of psychometric properties of the existing Affinity for technology interaction (ATI) scale initially developed in German and English (see, Franke, Attig, and Wessel, 2019). The scale could be used, for example, in research relating to educational technology. A key aim was also professional development: to learn the psychometric methods used in the research.

Data and methods The scale was first translated using forward-backward translation and a committee approach. Students ($N = 796$) studying in a Finnish public multidisciplinary research university responded to an online questionnaire containing the translated ATI scale. The scale data was analyzed using multiple psychometric methods: factor analysis, exploratory graph analysis, non-parametric item response theory, and parametric item response theory. Furthermore, the differences in ATI between genders were examined using hierarchical linear regression.

Main results The Finnish version of the ATI scale showed essential unidimensionality high reliability estimates and formed a strong Mokken scale. Men showed slightly higher ATI scores than women when controlling for age and field of study. However, the effect size was small and comparable to the findings identified in a meta-study (see, Cai, Fan, and J. Du, 2017). As a result, the scale could be used to find differences between individuals and groups in terms of ATI when applied in a population similar to the sample used in the research.

Research contributions A significant research contribution was to extend the use of the ATI scale to a new language region by providing a translated version of the scale for the research community. In addition, the research advanced the theory of measurement by synthesizing multiple psychometric methods as a comprehensive analytics scheme.

My contributions I invented the research idea, defined the research aims, acquired the funding, organized the translation process and created the initial translations with the second author, devised the analysis strategy, collected the data, conducted all the statistical analyses, created the visualizations, and wrote the initial draft of the article. I was the corresponding author and responsible for the submission and review process.

4.7 Summary of the research contributions

The previous section provided summaries of each included article. Table 1 summarizes the research contributions. The contributions are linked with the respective research questions.

TABLE 1 Summary of the research contributions

P	RQ	Research contributions
PI	RQ1, RQ2, RQ3	*i*) described the technique for constructing profiles using robust clustering, *ii*) demonstrated a novel visualization method for presenting the clustering results, *iii*) outlined a technical solution for analytics implementation taking into account security, privacy, and legal regulation, *iv*) showed associations between student agency and academic outcomes, *v*) showed different profiles between two fields of study
PII	RQ1, RQ2	*i*) provided insight about the characteristics of the low agency profile, *ii*) demonstrated how clustering-assisted qualitative analysis can provide detailed information about a specific subgroup of interest
PIII	RQ1, RQ2	*i*) showed a method for measuring general course satisfaction, *ii*) showed associations between general course satisfaction and student agency dimensions, *iii*) demonstrated how learning analytics can be linked with affective experience in learning, *iv*) provided insight about how course satisfaction could be supported through utilizing student agency analytics
PIV	RQ1, RQ2, RQ3	*i*) showed that explainable artificial intelligence methods can deepen the insight about the characteristics of the profiles and individual students, *ii*) considered the results from the teachers' perspective, *iii*) showed that explainable artificial intelligence methods can advance ethical learning analytics
PV	RQ1, RQ3	*i*) showed that teachers can reflect on their pedagogical actions based on student agency analytics results, *ii*) provided insight about how teachers interpret the different profiles, *iii*) indicated that teachers can use the student agency analytics results in pedagogical decision making
PVI	RQ3	*i*) demonstrated a comprehensive psychometric analysis scheme that could be also used for scales in learning analytics, *ii*) provided a Finnish translation of a human—technology interaction scale for use, for example, in educational technology research

5 GENERAL DISCUSSION

Pedagogically meaningful, research-based, and ethical learning analytics could foster the values and learning aims we want to advance in our society and educational system. Jane Tompkins (1990, p. 656) stated in her article *Pedagogy of the Distressed* that "classroom is a microcosm of the world." In some sense, a teacher's purpose is to mold and facilitate the formations of structure and agency inside these "small worlds" through professional knowledge, practice, experience, and perception. However, Tompkins (1990, p. 653) also argues that as teachers, "our practice in the classroom doesn't often come very close to instantiating the values we preach." Perhaps learning analytics could help us, educators, to enhance our *pedagogical awareness*, by which I mean the capability of a teacher to be cognizant—having awareness and knowledge—about learners' experiences and suitable pedagogical means to support learning in a given situation. As Chounta, Bardone, et al. (2021) metaphorically envisioned, maybe innovative learning analytics and artificial intelligence in education could equip educators with "teacher superpowers."

However, interdisciplinary collaboration is vital for innovations in learning analytics. Theoretical knowledge without practical applicability and analytics without theoretical grounding are both detached from the reality of learning and teaching. When using machine learning to understand human learning, it is imperative to integrate both educational and computational knowledge; in learning analytics, one cannot succeed without another.

5.1 Learning analytics with learning and analytics

The aim of student agency analytics advanced in this dissertation was to provide insight into the agentic learning experiences among a group of students, in other words, inside a "microcosm". The aim was achieved by utilizing machine learning and a psychometric scale. According to Buckingham Shum (2019), learning analytics aims to "design and deploy analytics that demonstrates how theory can inspire models, algorithms, code, user experiences, teaching practices, and ultimately, learning." Relatedly, the main benefits of the student agency analytics are that:

1) it is based on a theorized construct relating to student-centered learning,
2) it can be automatized and applied for a reasonable group size,
3) it utilizes a novel visualization for providing results for the teacher,
4) it respects students' privacy,
5) it is shown to be associated with positive affect (i.e., course satisfaction) and performance,
6) teachers could use the results in pedagogical reflection and decision-making.

Greller and Drachsler (2012) identified six critical dimensions of learning analytics: objectives, stakeholders, data, instruments, internal limitations, and external limitations. I will review the results of the individual articles in terms of the six critical dimensions (Figure 8). The treatment provides answers to the overall research questions of this dissertation as stated in Section 1 and repeated below:

RQ1 For what purposes teachers could utilize student agency analytics in higher education?

RQ2 What methodological requirements student agency analytics introduce?

RQ3 How student agency analytics could overcome some of the limitations in learning analytics?

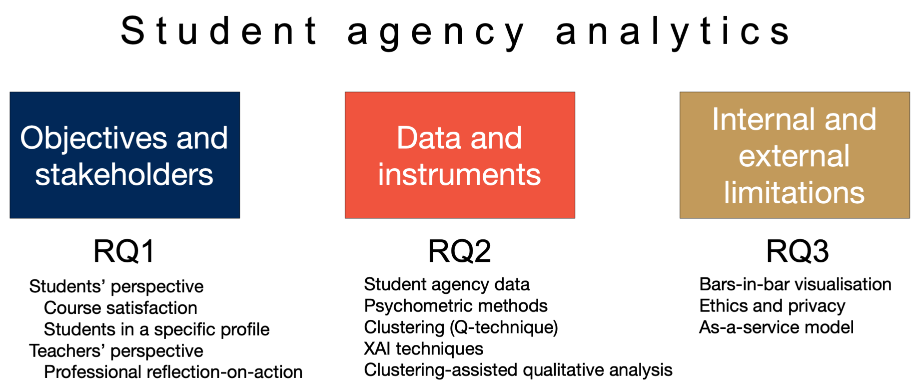


FIGURE 8 Student agency analytics reflected through the critical dimensions of learning analytics depicted by Greller and Drachsler (2012).

RQ1 Objectives and stakeholders The emphases in learning analytics should be on theoretical knowledge of education, pedagogy, and the inclusion of students and teachers in the process of developing and applying analytics (Guzmán-Valenzuela et al., 2021). When starting a learning analytics project, one should choose a theoretical basis for the analytics and an objective to advance. Blumenstein (2020) found out that learning personalization did not have a large impact on academic performance or course completion. Thus, she suggested that perhaps the emphasis should be shifted to student well-being and emotional support. The theoretical basis for the learning analytics process advanced in this dissertation is student agency conceptualized as a multidimensional construct of available resources in learning (Jääskelä, Poikkeus, Häkkinen, et al., 2020; Jääskelä, Poikkeus, Vasalampi, et al., 2017) which relates closely to students' well-being and emotional aspects of learning. A major benefit of learning analytics is that the results can provide possibilities for self-reflection among stakeholders (i.e., students, teachers, institutions) (e.g., Greller and Drachsler, 2012; Ifenthaler and Greiff, 2021). Yau and Ifenthaler (2021) pointed out that teaching analytics—the analytics methods and tools for teaching staff—could help teachers to analyze, reflect on, and improve their pedagogical practice and increase students' learning experiences. My dissertation specifically addressed the teachers' point of view in utilizing the analytics results (Article PV). The results indicated that the teachers were able to reflect on (see, D. A. Schön, 1983) their pedagogical actions based on the presented visualization of the student agency analytics results of their courses: they pointed out the contextual, situational, and temporal aspects of agency. Also, the teachers suggested tentative actions that might have improved learning experiences in their courses.

Knight, Shum, and Littleton (2014) outlined that subjectivist or affect-based approach in learning analytics emphasizes personal affect instead of learning in a traditional academic sense and, for example, tries to understand why a learner is or is not engaged in different contexts. Thus, from the learning analytics point of view in general, student agency analytics can be considered as an example of subjectivist and student-centered learning analytics. In other words, the stakeholders of student agency analytics are the students themselves and educators wishing to promote student-centered learning. The analysis process was also examined from the students' point of view. The results showed that students in the low agency profile experienced several restrictions in their studies (Article PII), higher scores on student agency were associated with higher course satisfaction (Article PIII), and course satisfaction and performance in terms of higher course grades were associated with higher student agency (Article PI). From the pedagogical and learning analytics point of view, the question would be how teachers' actions and learning analytics could support student agency. The results indicated that pedagogical actions and learning analytics for helping students in time management, providing them personalized, just-in-time suggestions for learning materials, and optimizing teachers' support could facilitate students' agentic experiences and increase general course satisfaction.

From the research point of view, student agency analytics could be considered as a research tool for examining learning experiences in a real-world educational setting (e.g., practitioner research). The developed approach provides a convenient way to explore the different aspects of student agency "in the wild." By utilizing student agency analytics in a wide range of educational settings, one could obtain a more comprehensive view of the multifaceted phenomenon. From the research methods point of view, student agency analytics can be used in *clustering-assisted qualitative analysis* (Article PII) to select a particular subset of respondents for a detailed analysis.

RQ2 Data and instruments Instruments in learning analytics include the actual tools and processes for collecting data but also conceptual instruments such as theoretical constructs and algorithms (Greller and Drachsler, 2012). The theoretical construct in student agency analytics was the AUS model (Jääskelä, Poikkeus, Häkkinen, et al., 2020; Jääskelä, Poikkeus, Vasalampi, et al., 2017). However, the theoretical construct and measurement instrument can be replaced and the analytics process could follow a broadly similar approach in a generalized case. Data in student agency analytics was collected using an online questionnaire and a psychometric scale (Articles PI–PV). When utilizing scales, the quality of the data depends on the properties of the scale. Thus, the scales used in learning analytics should be exposed to a comprehensive psychometric analysis. The AUS model was analyzed in previous studies (Jääskelä, Poikkeus, Häkkinen, et al., 2020; Jääskelä, Poikkeus, Vasalampi, et al., 2017). The scale translation process and psychometric analysis in Article PVI demonstrated a comprehensive analysis scheme utilizing factor analysis and nonparametric and parametric item response theory.

The purpose of robust clustering in student agency analytics (Articles PI, PII, PIV, and PV) was to group students in terms of the second-order factorial structure of the AUS scale (i.e., the dimensions of student agency) (Figure 5). The approach relates to a so-called Q-technique (Cattell, 1952; Nunnally and Bernstein, 1994, p. 526–530). Q-technique is "a type of factor analysis used to understand the major dimensions or 'types' of people by identifying how they perceive different variables" (APA, n.d.). In other words, factors or clusters are created by grouping related objects of interest. Clustering can be thought of as a form of Q-technique (e.g., Miller, 1978; Vogler et al., 1989). The other types of techniques are the R-technique which aims to understand the associations between variables and how they group together, and the P-technique, which aims to understand the associations between variables over time (Cattell, 1952). R-technique refers to traditional factor analysis, which was used to construct the AUS scale (i.e., Jääskelä, Poikkeus, Häkkinen, et al., 2020; Jääskelä, Poikkeus, Vasalampi, et al., 2017). I have conducted an initial analysis utilizing the idea of P-technique by collecting and analyzing repeated intraindividual student agency analytics measurements during the same course. Initial results indicated that the agentic experiences could exhibit temporal variation. The initial results align with the conception of agentic orientation (Klemenčič, 2015) and that agency is a temporally embedded process (Emirbayer and Mische, 1998). Thus, an idiographic approach concerning student agency could provide interesting insights into personal learning experiences. Relating to individual differences, the clustering-assisted qualitative analysis provided a convenient way to select a subset of respondents for a detailed analysis (Article PII).

From the teacher's perspective, the population of interest is the students in the course, or at least the students who decide to answer the questionnaire and disclose their learning experience for analysis. Thus, the profiles formed using clustering provide valuable information relating to those *particular* students. In other words, the teacher might be specifically interested in the experiences of her or his students and not necessarily about statistical generalizations to all "general students". In that sense, for student agency analytics, it is not feasible to use methods that require large amounts of observations (e.g., latent profile analysis (S. L. Ferguson, G. Moore, and Hull, 2020)) because then the process would be applicable only in very large courses. The results based on the student agency analytics process relying on robust clustering showed that it is possible to distill relevant information from groups of around 30 students (Articles PIV and PV). Therefore, the reduced requirements concerning the necessary amount of data of robust analytics (Kärkkäinen and Heikkola, 2004) conveniently support the possibilities to produce course-level and temporally varying agency information for the stakeholders.

Methods of explainable artificial intelligence (XAI) were applied and examined in the context of student agency analytics (Article PIV). XAI methods can provide more detailed information about a single student (i.e., local explanation) and profile characteristics (i.e., global explanation). From the teacher's perspective, the information can clarify the inter-individual differences and intra-profile similarities relating to students' learning experiences. Therefore, the teacher and other relevant stakeholders could make more personalized pedagogical decisions based on the XAI results. From the methodological point of view, the complex nonlinear methods performed better in terms of accuracy, whereas the traditional linear techniques performed worse but provided more informative explanations (Articles PIII and PIV). While XAI methods can provide some transparency to the complex models, they can be considered models on top of models. Thus, the model selection is a tradeoff between performance and explainability as XAI methods can introduce an additional layer of complexity.

RQ3 Internal and external limitations Learning analytics is affected by enabling or restricting human factors like competence and acceptance of the users (i.e., internal restrictions) and environmental factors like conventions and norms (i.e., external restrictions) (Greller and Drachsler, 2012). Visualizing the learning analytics results is a powerful way to convey relevant information and reduce internal limitations by advancing comprehension and interpretation. Ifenthaler and Yau (2020, p. 1984) highlighted the need for teachers' educational data literacy which means "ethically responsible collection, management, analysis, comprehension, interpretation, and application of data from educational contexts." Vieira, Parsons, and V. Byrd (2018) pointed out that more emphasis should be put on connecting educational theory and sophisticated visualizations. In general, the visualization of the learning analytics results should align with the data literacy ability of the intended stakeholders (Sarikaya et al., 2018). Thus, a novel visualization method (Articles PI, PII, PIV, and PV), *bars-in-bar*, was used to depict the general average level of each student agency dimension and the respective deviations. A clear benefit of learning analytics is that it makes an abstract concept—like student agency—visible so that it can be brought under mutual discussion (Articles PI, PII, PIV, and PV). Furthermore, the clustering method and the visualization used to inform the teacher aggregates the information and provides a certain degree of privacy for the students.

Students' privacy in learning analytics is a central concern (e.g., Ifenthaler and Schumacher, 2016). The proposed microservices-based model and pseudonymization of the data (Articles PI, PII, and PIV) aimed to address some of the data ownership and security issues: data controller and data processor could be separated if needed, and pseudonymization could provide a layer of security by complicating re-identification in case of a data breach.

Psychometric properties of a scale affect the reliability and validity of a learning analytics process. The scale must provide consistent and unbiased data from the ethical learning analytics perspective. Scales utilized in learning analytics should be analyzed rigorously to prevent biased data from affecting and degrading the results. For example, differential item functioning is an important property to evaluate (Article PVI). There can exist gender differences in the usage of educational technology (Sahin and Ifenthaler, 2022) and generally in technology interaction (Article PVI). Thus, it is essential to have rigorously examined instruments for assessing different experiences towards technology. The translated version (Article PVI) of the Affinity for technology interaction (ATI) scale (Franke, Attig, and Wessel, 2019) showed promising properties for measuring technology interaction in Finnish.

5.2 Reliability and validity

Triangulation based on data, analysis, and methods (Denzin, 1978) can promote reliability and validity of a research (e.g., Hussein, 2009; Risjord, Moloney, and Dunbar, 2001; Wimsatt, 1981). In this research, data were obtained from different sources and contexts (i.e., fields of study, institutions) and in different formats (i.e., quantitative and qualitative). From the analysis point of view, cluster validation (Article PI), cross-validation (Articles PI, PIII, and PIV), and different classifiers (Articles PIII and PIV) were used or compared in the analyses. Also, XAI methods provided transparency and interpretability relating to the student agency profiles obtained using robust clustering (Article PIV). The robust clustering (Article PI) produced a similar typology as the latent profile analysis in Jääskelä, Poikkeus, Häkkinen, et al. (2020). Methodologically both quantitative and qualitative approaches were used to examine the agencic experiences of the students in the low agency profile (Article PII). In the case of the translated ATI scale, different psychometric methods (Article PVI) demonstrated how multiple analysis approaches could provide complementary information about the functioning of a scale. From the scale development point of view, the AUS scale (Jääskelä, Poikkeus, Häkkinen, et al., 2020; Jääskelä, Poikkeus, Vasalampi, et al., 2017) and translated ATI scale (Article PVI) showed high reliability estimates and solid psychometric properties in terms of factorial structure.

The results relating to course satisfaction (Article PIII) and performance (Article PI) can be considered indicators of predictive validity. The results relating to the student's experiences in the low agency profile (Article PII) can be considered indicators of construct validity. Furthermore, the student agency analytics process, in general, can be assessed through the lens of pragmatic validity. Worren, Moore, and Elliott (2002) concluded that "pragmatic validity is fundamentally about whether the use of certain tools helps guide action to attain goals." In essence, the teachers' interviews (Article PV) showed that they could utilize student agency analytics results in their professional reflection which can be counted towards pragmatic validity of the analytics process. Lastly, the data in all articles were collected from a natural educational setting in higher education which can be considered to enhance the ecological validity of the research as defined by Orne (1962) (see also, Kihlstrom, 2021).

5.3 Limitations and recommendations for future research

The research in my dissertation consisted of one "cycle" of learning analytics research and development as depicted in the Thesis at a Glance section. More research is needed, for example, to examine how students perceive their personal analytics results and how teachers utilize the information. Student agency analytics could be utilized in a large-scale study and longitudinal research in several courses spanning several semesters. Also, student agency analytics could be utilized at the institutional level: time series data relating to student agency could be collected from different courses, or profiles could be created at the course level or even at the institutional level. Furthermore, student agency analytics could be applied as a research tool in quasi-experimental research where pedagogical interventions are developed based on the analytics results. The presented analytics approach and the results could pave the way for theorizing and exploring student agency in other educational levels (e.g., K12 and vocational) and subject areas.

The presented research has shed light on the complex nature of student agency. However, associations of student agency with other constructs and their relationships need further investigation. Future research could examine also the temporal intraindividual variation in the student agency dimensions in different learning contexts. I have conducted the research "inside the system" as I have been a student, a teacher, and currently a researcher working in the Scandinavian student-centered educational system. Thus, while aiming at objectivity, my scientific thinking can be biased and distorted by contextual reasons, motives, and incitements. The results are applicable in a Finnish educational context, and in another cultural context, the process of student agency analytics can be conceived differently. Lastly, the analytics approach and methods presented in this dissertation for localising and explainably analyzing psychometric scale data could be applied for any relevant learning-related scale in future research.

YHTEENVETO (SUMMARY IN FINNISH)

Väitöskirjani käsittelee toimijuusanalytiikaksi kutsuttua oppimisananalytiikan menetelmää, joka hyödyntää psykometrista mittaria oppimistiedon keräämiseksi ja robustia koneoppimista tiedon analysoimiseksi. Toimijuusanalytiikan tarkoituksena on tarjota tietoa oppimiskokemuksesta ja sen sosiaalisista, affektiivisista ja kognitiivisista ulottuvuuksista. Oppimiskokemusta käsitellään opiskelijan toimijuuden (Jääskelä, Poikkeus, Häkkinen et al., 2020; Jääskelä, Poikkeus, Vasalampi et al., 2017) käsitteen kautta. Väitöskirjassa tarkastellaan toimijuusanalytiikalla saatua tietoa erityisesti opettajan näkökulmasta. Väitöskirja kokonaisuutena vastaa seuraaviin tutkimuskysymyksiin:

RQ1 Mihin tarkoitukseen korkeakoulujen opettajat voivat käyttää toimijuusanalytiikkaa?

RQ2 Millaisia menetelmälisiä vaatimuksia toimijuusanalytiikkaa liittyy?

RQ3 Miten toimijuusanalytiikassa voidaan ratkaista joitakin oppimisanalytiikan rajoitteita?

Väitöskirja koostuu kuudesta osatutkimuksesta. Artikkeli PI esitteli toimijuusanalytiikan prosessin ja analyysimenetelmän, jossa toimijuustieto hankitaan psykometrisen mittarin avulla, analysoidaan robustin koneoppimisen avulla ja esitetään käyttämällä uutta visualisointimenetelmää. Toimijuusanalytiikka tuottaa neljä toimijuusprofiilia sekä yleisen toimijuusprofiilin opiskelijaryhmästä. Profiilit näyttävät liittyvän opintomenestykseen: vahvempi toimijuus liittyi parempaan opintomenestykseen. Artikkeli PII eritteli heikomman toimijuusprofiilin opiskelijoiden oppimiskokemuksia. Tämän toimijuusprofiilin opiskelijat raportoivat erilaisia opiskeluun liittyviä rajoittavia tekijöitä, kuten aiemman osaamisen puute sekä opettajan tuen ja opintoihin käytettävissä olevan ajan puute. Artikkeli PIII selvitti toimijuuden ja kurssityytyväisyyden välistä yhteyttä. Tulosten perusteella voidaan sanoa, että koetut vahvemmat toimijuuden resurssit—erityisesti opettajan tuki, luottamus opettajaan ja tasapuolinen kohtelu—ovat yhteydessä korkeampaan koettuun tyytyväisyyteen opinnoissa. Tulosten perusteella toimijuutta tukevia pedagogisia toimenpiteitä voidaan kohdistaa erityisesti näihin kriittisiin resurssialueisiin. Artikkeli IV sovelsi selittävän tekoälyn menetelmiä toimijuusanalytiikan prosessiin. Selittävän tekoälyn menetelmät nostivat tarkemmin esiin toimijuusprofiilien ja niille tyypillisten opiskelijoiden toimijuuden piirteitä. Artikkeli V eritteli neljän opettajan tulkintoja oman kurssinsa toimijuusanalyysistä. Tulokset osoittivat, että opettajat pystyivät reflektoimaan omaa ammatillista toimintaansa toimijuusprofiilien perusteella. Artikkeli PVI esitteli erään ihmisen ja teknologian vuorovaikutukseen liittyvän psykometrisen mittarin suomenkielisen käännöksen ja sen psykometrisen analyysin. Mittari osoittautui päteväksi instrumentiksi kohderyhmässään ja sen avulla voitiin tunnistaa pieni ero miesten ja naisten välillä mitattuun käsitteeseen liittyen.

Yleisesti ottaen toimijuusanalytiikka perustuu teoreettiseen ja oppijakeskeiseen käsitteeseen. Analytiikkaprosessi voidaan automatisoida ja sitä voidaan soveltaa myös pienelle opiskelijamäärälle. Tulosten esittämisessä käytetään uutta visualisointitapaa. Toimijuusanalytiikassa on otettu huomioon yksityisyyden vaatimukset. Analytiikan tulokset näyttävät liittyvän sekä opintomenestykseen että kurssityytyväisyyteen. Opettajat voivat käyttää toimijuusanalytiikan tuloksia pedagogisessa ja ammatillisessa reflektiossa ja päätöksenteossa.

BIBLIOGRAPHY

Ackoff, R. L. 1989. From data to wisdom. Journal of Applied Systems Analysis 16 (1), 3–9.

Adadi, A. and Berrada, M. 2018. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). IEEE Access 6, 52138–52160. DOI: 10.1109/ACCESS.2018.2870052.

AERA, APA, and NCME 2014. Standards for educational and psychological testing. Washington, DC: American Educational Research Association.

Ain, N., Vaia, G., DeLone, W. H., and Waheed, M. 2019. Two decades of research on business intelligence system adoption, utilization and success – A systematic literature review. Decision support systems 125, 113113. DOI: 10.1016/j.dss.2019.113113.

Akos, P. 2004. Advice and Student Agency in the Transition to Middle School. RMLE Online 27 (2), 1–11. DOI: 10.1080/19404476.2004.11658169.

Allen, M. 2017. Pentadic Analysis. In The SAGE Encyclopedia of Communication Research Methods. Ed. by Allen, M. Thousand Oaks, CA: SAGE Publishing. DOI: 10.4135/9781483381411.n420.

Alvarez, M. G. 2018. Can character solve our problems? Character qualities and the imagination age. Creative Education 9 (2), 152–164.

Anderson, C., Day, K., and McLaughlin, P. 2006. Mastering the dissertation: lecturers' representations of the purposes and processes of Master's level dissertation supervision. Stud. High. Educ. 31 (2), 149–168. DOI: 10.1080/03075070600572017.

Anderson, F. D. and Prelli, L. J. 2001. Pentadic cartography: Mapping the universe of discourse. The Quarterly Journal of Speech 87 (1), 73–95. DOI: 10.1080/00335630109384319.

Angelov, P. P., Soares, E. A., Jiang, R., Arnold, N. I., and Atkinson, P. M. 2021. Explainable artificial intelligence: an analytical review. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 11 (5). DOI: 10.1002/widm.1424.

Angoff, W. H. 1993. Perspectives on differential item functioning methodology. In Differential item functioning. Ed. by Holland, P. W. Vol. 453. Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc, xv, 3–23. DOI: 10.1075/z.62.13kok.

Anscombe, G. E. M. 1957. Intention. Oxford: Basil Blackwell.

APA n.d. Q-technique factor analysis. https://dictionary.apa.org/q-technique-factor-analysis. Accessed: 2022-2-3. Washington, DC.

Ark L, A. van der 2007. Mokken scale analysis in R. J. Stat. Softw. 20 (11), 1–19. DOI: 10.18637/jss.v020.i11.

Ark L, A. van der 2013. Visualizing Uncertainty of Estimated Response Functions in Nonparametric Item Response Theory. In New Developments in Quantitative Psychology. Springer New York, 59–68. DOI: 10.1007/978-1-4614-9348-8_5.

Arthur, D. and Vassilvitskii, S. 2007. k-means++: The Advantages of Careful Seeding. In SODA '07: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Asswad, J. and Marx Gómez, J. 2021. Data Ownership: A Survey. Information 12 (11), 465. DOI: 10.3390/info12110465.

Auerswald, M. and Moshagen, M. 2019. How to determine the number of factors to retain in exploratory factor analysis: A comparison of extraction methods under realistic conditions. Psychol. Methods 24 (4), 468–491. DOI: 10.1037/met0000200.

Avella, J. T., Kebritchi, M., Nunn, S. G., and Kanai, T. 2016. Learning Analytics Methods, Benefits, and Challenges in Higher Education: A Systematic Literature Review. Online Learning 20 (2), 13–29.

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., and Herrera, F. 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inf. Fusion 58, 82–115. DOI: 10.1016/j.inffus.2019.12.012.

Baylor, A. 1999. Intelligent Agents as Cognitive Tools for Education. Educational technology research and development: ETR & D 39 (2), 36–40.

Beauducel, A. and Herzberg, P. Y. 2006. On the Performance of Maximum Likelihood Versus Means and Variance Adjusted Weighted Least Squares Estimation in CFA. Structural Equation Modeling 13 (2), 186–203. DOI: 10.1207/s15328007sem1302_2.

Benjamin, L. T. 1988. A history of teaching machines. Am. Psychol. 43 (9), 703–712. DOI: 10.1037/0003-066X.43.9.703.

Biesta, G. and Tedder, M. 2007. Agency and learning in the lifecourse: Towards an ecological perspective. Studies in the Education of Adults 39 (2), 132–149. DOI: 10.1080/02660830.2007.11661545.

Blikstein, P. 2013. Multimodal learning analytics. In Proceedings of the Third International Conference on Learning Analytics and Knowledge. LAK '13. Leuven, Belgium: Association for Computing Machinery, 102–106. DOI: 10.1145/2460296.2460316.

Blikstein, P. and Worsley, M. 2016. Multimodal Learning Analytics and Education Data Mining: using computational technologies to measure complex learning tasks. Journal of Learning Analytics 3 (2), 220–238. DOI: 10.18608/jla.2016.32.11.

Blumenstein, M. 2020. Synergies of learning analytics and learning design: A systematic review of student outcomes. J. learn. anal. 7 (3), 13–32. DOI: 10.18608/jla.2020.73.3.

Bollen, K. A. 1989. Structural Equations with Latent Variables. Hoboken, NJ: John Wiley & Sons.

Bolognini, L. and Bistolfi, C. 2017. Pseudonymization and impacts of Big (personal/anonymous) Data processing in the transition from the Directive 95/46/EC to the new EU General Data Protection Regulation. Computer Law & Security Review 33 (2), 171–181. DOI: 10.1016/j.clsr.2016.11.002.

Borsboom, D. 2005. Measuring the Mind: Conceptual Issues in Contemporary Psychometrics. Cambridge, NY: Cambridge University Press.

Braun, V. and Clarke, V. 2006. Using thematic analysis in psychology. Qual. Res. Psychol. 3 (2), 77–101. DOI: 10.1191/1478088706qp063oa.

Breiman, L. 2001. Random Forests. Mach. Learn. 45 (1), 5–32. DOI: 10.1023/A:1010933404324.

Briskin, L. and Coulter, R. P. 1992. Introduction: Feminist Pedagogy: Challenging the Normative. Canadian Journal of Education / Revue canadienne de l'éducation 17 (3), 247–263. DOI: 10.2307/1495295.

Brislin, R. W., Lonner, W. J., and Thorndike, R. M. 1973. Cross-cultural research methods. Comparative studies in behavioral science. New York, NY: John Wiley & Sons.

Brown, T. A. 2015. Confirmatory Factor Analysis for Applied Research. second. Methodology in the Social Sciences. New York, NY: Guilford Publications.

Bruner, J. 1991. The Narrative Construction of Reality. Critical inquiry 18 (1), 1–21. DOI: 10.1086/448619.

Buckingham Shum, S. 2018. Transitioning education's knowledge infrastructure: Shaping design or shouting from the touchline? In Proceedings of International Conference of the Learning Sciences, ICLS.

Buckingham Shum, S. 2019. Transitioning Education's Knowledge Infrastructure. https://www.youtube.com/watch?v=89569EPNFtQ. Accessed: 2022-2-6.

Buja, A. and Eyuboglu, N. 1992. Remarks on Parallel Analysis. Multivariate Behav. Res. 27 (4), 509–540. DOI: 10.1207/s15327906mbr2704_2.

Burke, K. 1945. A Grammar of Motives. first. New York, NY: Prentice-Hall.

Cai, Z., Fan, X., and Du, J. 2017. Gender and attitudes toward technology use: A meta-analysis. Comput. Educ. 105, 1–13. DOI: 10.1016/j.compedu.2016.11.003.

Campbell, J. P., DeBlois, P. B., and Oblinger, D. G. 2007. Academic analytics: A new tool for a new era. EDUCAUSE review 42 (4), 40.

Carver, R. L. 1996. Theory for Practice: A Framework for Thinking about Experiential Education. Journal of Experiential Education 19 (1), 8–13. DOI: 10.1177/105382599601900102.

Carver, R. L. 1997. Theoretical underpinnings of service learning. Theory into practice 36 (3), 143–149. DOI: 10.1080/00405849709543760.

Cattell, R. B. 1952. The three basic factor-analytic research designs—their interrelations and derivatives. Psychol. Bull. 49 (5), 499–520. DOI: 10.1037/h0054245.

Cech, T. G., Spaulding, T. J., and Cazier, J. A. 2018. Data competence maturity: developing data-driven decision making. Journal of Research in Innovative Teaching & Learning 11 (2), 139–158. DOI: 10.1108/JRIT-03-2018-0007.

Charitopoulos, A., Rangoussi, M., and Koulouriotis, D. 2020. On the Use of Soft Computing Methods in Educational Data Mining and Learning Analytics Research: a Review of Years 2010–2018. International Journal of Artificial Intelligence in Education 30 (3), 371–430. DOI: 10.1007/s40593-020-00200-8.

Chounta, I.-A. and Avouris, N. 2016. Towards the real-time evaluation of collaborative activities: Integration of an automatic rater of collaboration quality in the classroom from the teacher's perspective. Education and Information Technologies 21 (4), 815–835. DOI: 10.1007/s10639-014-9355-3.

Chounta, I.-A., Bardone, E., Raudsep, A., and Pedaste, M. 2021. Exploring teachers' perceptions of artificial intelligence as a tool to support their practice in Estonian K-12 education. International journal of artificial intelligence in education. DOI: 10.1007/s40593-021-00243-5.

Clow, D. 2012. The Learning Analytics Cycle: Closing the Loop Effectively. In Proceedings of the 2Nd International Conference on Learning Analytics and Knowledge - LAK '12. New York, NY: ACM, 134–138. DOI: 10.1145/2330601.2330636.

Conole, G., Gašević, D., Long, P., and Siemens, G. 2011. Message from the LAK 2011 General & Program Chairs. In Proceedings of the 1st International Conference on Learning Analytics and Knowledge. New York, NY: ACM.

Cord, A., Ambroise, C., and Cocquerez, J.-P. 2006. Feature selection in robust clustering based on Laplace mixture. Pattern Recognit. Lett. 27 (6), 627–635. DOI: 10.1016/j.patrec.2005.09.028.

Crick, R. D., Broadfoot, P., and Claxton, G. 2004. Developing an Effective Lifelong Learning Inventory: the ELLI Project. Assessment in Education: Principles, Policy & Practice 11 (3), 247–272. DOI: 10.1080/0969594042000304582.

Daniel, B. K. 2019. Big Data and data science: A critical review of issues for educational research. Br. J. Educ. Technol. 50 (1), 101–113. DOI: 10.1111/bjet.12595.

Datnow, A. and Hubbard, L. 2016. Teacher capacity for and beliefs about data-driven decision making: A literature review of international research. Journal of Educational Change 17 (1), 7–28. DOI: 10.1007/s10833-015-9264-2.

Davidson, D. 1963. Actions, Reasons, and Causes. The Journal of Philosophy 60 (23), 685–700. DOI: 10.2307/2023177.

Davies, L. 1994. Can Students Make a Difference? International Perspectives on Transformative Education. International Studies in Sociology of Education 4 (1), 43–56. DOI: 10.1080/0962021940040103.

Dawson, S., Joksimovic, S., Poquet, O., and Siemens, G. 2019. Increasing the Impact of Learning Analytics. In Proceedings of the 9th International Conference on Learning Analytics & Knowledge. LAK19. Tempe, AZ, USA: Association for Computing Machinery, 446–455. DOI: 10.1145/3303772.3303784.

Denzin, N. K. 1978. The Research Act: A Theoretical Introduction to Sociological Methods. second. New York: McGraw-Hill.

DeVellis, R. F. 2017. Scale Development: Theory and Applications. fourth. Los Angeles, CA: SAGE Publications.

Di Mitri, D., Schneider, J., Specht, M., and Drachsler, H. 2018. From signals to knowledge: A conceptual model for multimodal learning analytics. J. Comput. Assist. Learn. 34 (4), 338–349. DOI: 10.1111/jcal.12288.

DiStefano, C. and Morgan, G. B. 2014. A Comparison of Diagonal Weighted Least Squares Robust Estimation Techniques for Ordinal Data. Struct. Equ. Modeling 21 (3), 425–438. DOI: 10.1080/10705511.2014.915373.

DiStefano, C., Zhu, M., and Mîndrilã, D. 2009. Understanding and Using Factor Scores: Considerations for the Applied Researcher. Practical Assessment, Research, and Evaluation 14 (1), 20. DOI: 10.7275/da8t-4g52.

Dorans, N. J. and Cook, L. L. 2016. Fairness in Educational Assessment and Measurement. Milton Park, Abingdon-on-Thames, Oxfordshire, England: Routledge.

Drachsler, H. and Goldhammer, F. 2020. Learning Analytics and eAssessment—Towards Computational Psychometrics by Combining Psychometrics with Learning Analytics. In Radical Solutions and Learning Analytics: Personalised Learning and Teaching Through Big Data. Ed. by Burgos, D. Singapore: Springer Singapore, 67–80. DOI: 10.1007/978-981-15-4526-9_5.

Du, X., Yang, J., Shelton, B. E., Hung, J.-L., and Zhang, M. 2021. A systematic meta-Review and analysis of learning analytics research. Behaviour & information technology 40 (1), 1–14. DOI: 10.1080/0144929X.2019.1669712.

Emirbayer, M. and Mische, A. 1998. What is agency? American journal of sociology 103 (4), 962–1023. DOI: 10.1086/231294.

Epskamp, S., Rhemtulla, M., and Borsboom, D. 2017. Generalized Network Psychometrics: Combining Network and Latent Variable Models. Psychometrika 82 (4), 904–927. DOI: 10.1007/s11336-017-9557-x.

European Union 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 (OJ) 59 (L119), 1–88.

Evans, T. and Nation, D. 1989. Dialogue in practice, research and theory in distance education. Open Learning: The Journal of Open, Distance and e-Learning 4 (2), 37–42. DOI: 10.1080/0268051890040207.

Ewald, H. R. and Wallace, D. L. 1994. Exploring Agency in Classroom Discourse or, Should David Have Told His Story? College Composition and Communication 45 (3), 342–368. DOI: 10.2307/358815.

Fabrigar, L. R. and Wegener, D. T. 2012. Exploratory Factor Analysis. Series in Understanding Statistics. Oxford: Oxford University Press.

Fayyad, U. M., Piatetsky-Shapiro, G., and Smyth, P. 1996. Data mining and knowledge discovery in databases. AI Magazine 17 (3), 37–54. DOI: 10.1145/240455.240463.

Fenwick, T. 1996. Balancing on the Edge of Surprise: Managing Junior High Classrooms. In 24th Annual Conference of the Canadian Society for the Study of Education. Ontario, Canada.

Ferguson, R. and 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 International Conference Proceeding Series. Vancouver, BC, Canada: ACM, 56–65. DOI: 10.1145/3027385.3027396.

Ferguson, R., Hoel, T., Scheffel, M., and Drachsler, H. 2016. Guest Editorial: Ethics and Privacy in Learning Analytics. Journal of Learning Analytics 3 (1), 5–15. DOI: 10.18608/jla.2016.31.2.

Ferguson, S. L., G. Moore, E. W., and Hull, D. M. 2020. Finding latent groups in observed data: A primer on latent profile analysis in Mplus for applied researchers. Int. J. Behav. Dev. 44 (5), 458–468. DOI: 10.1177/0165025419881721.

Finlex 1999. The Constitution of Finland.

Foldnes, N. and Grønneberg, S. 2021. The sensitivity of structural equation modeling with ordinal data to underlying non-normality and observed distributional forms. Psychol. Methods. DOI: 10.1037/met0000385.

Forero, C. G., Maydeu-Olivares, A., and Gallardo-Pujol, D. 2009. Factor Analysis with Ordinal Indicators: A Monte Carlo Study Comparing DWLS and ULS Estimation. Struct. Equ. Modeling 16 (4), 625–641. DOI: 10.1080/10705510903203573.

Franke, T., Attig, C., and Wessel, D. 2019. A Personal Resource for Technology Interaction: Development and Validation of the Affinity for Technology Interaction (ATI) Scale. International Journal of Human–Computer Interaction 35 (6), 456–467. DOI: 10.1080/10447318.2018.1456150.

Friedman, J., Hastie, T., and Tibshirani, R. 2008. Sparse inverse covariance estimation with the graphical lasso. Biostatistics 9 (3), 432–441. DOI: 10.1093/biostatistics/kxm045.

Galton, F. 1879. Psychometric Experiments. Brain 2 (2), 149–162. DOI: 10.1093/brain/2.2.149.

Gao, H. 2021. Reform of College English Teaching Model under the Background of Artificial Intelligence. J. Phys. Conf. Ser. 1744 (4), 042161. DOI: 10.1088/1742-6596/1744/4/042161.

Garrido, L. E., Abad, F. J., and Ponsoda, V. 2013. A new look at Horn's parallel analysis with ordinal variables. Psychological Methods 18 (4), 454–474. DOI: 10.1037/a0030005.

Giddens, A. 1984. The Constitution Of Society: outline of the theory of structuration. Cambridge: Polity Press.

Global Industry Analysts 2021. Educational Services - Global Market Trajectory & Analytics. Tech. rep. San Jose, CA.

Golino, H., Shi, D., Christensen, A. P., Garrido, L. E., Nieto, M. D., Sadana, R., Thiyagarajan, J. A., and Martinez-Molina, A. 2020. Investigating the performance of exploratory graph analysis and traditional techniques to identify the number of latent factors: A simulation and tutorial. Psychological Methods 25 (3), 292–320. DOI: 10.1037/met0000255.

González-Calatayud, V., Prendes-Espinosa, P., and Roig-Vila, R. 2021. Artificial Intelligence for student assessment: A systematic review. Appl. Sci. 11 (12), 5467. DOI: 10.3390/app11125467.

Goodenough, W. H. 1955. A Technique for Scale Analysis. Educ. Psychol. Meas. 4 (3), 179–190. DOI: 10.1177/001316445500400301.

Goretzko, D., Pham, T. T. H., and Bühner, M. 2019. Exploratory factor analysis: Current use, methodological developments and recommendations for good practice. Current Psychology. DOI: 10.1007/s12144-019-00300-2.

GPT-3 2020. A robot wrote this entire article. Are you scared yet, human? https://web.archive.org/web/20200908090812/https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3.

Green, S. B. and Yang, Y. 2009. Reliability of summed item scores using structural equation modeling: An alternative to coefficient alpha. Psychometrika 74 (1), 155–167. DOI: 10.1007/s11336-008-9099-3.

Greiff, S. and Heene, M. 2017. Why Psychological Assessment Needs to Start Worrying About Model Fit. European Journal of Psychological Assessment 33 (5), 313–317. DOI: 10.1027/1015-5759/a000450.

Greller, W. and Drachsler, H. 2012. Translating Learning into Numbers: A Generic Framework for Learning Analytics. Journal of Educational Technology & Society 15 (3), 42–57.

Grice, J. W. 2001. Computing and evaluating factor scores. Psychol. Methods 6 (4), 430–450.

Guttman, L. L. 1950. The basis for scalogram analysis. In Measurement and prediction. Ed. by Stouffer, S. A. Studies in social psychology in World War II. Princeton, NJ: Princeton University Press.

Guzmán-Valenzuela, C., Gómez-González, C., Rojas-Murphy Tagle, A., and Lorca-Vyhmeister, A. 2021. Learning analytics in higher education: a preponderance of analytics but very little learning? Int J Educ Technol High Educ 18 (1), 23. DOI: 10.1186/s41239-021-00258-x.

Hämäläinen, J., Jauhiainen, S., and Kärkkäinen, T. 2017. Comparison of Internal Clustering Validation Indices for Prototype-Based Clustering. Algorithms 10 (3), 105. DOI: 10.3390/a10030105.

Hämäläinen, J., Kärkkäinen, T., and Rossi, T. 2018. Scalable robust clustering method for large and sparse data. In European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning.

Hanson, R. 2018. The Importance of Learning. https://web.archive.org/web/20180327113559/https://www.rickhanson.net/the-importance-of-learning/. Accessed: 2022-2-1.

Hattie, J. 1985. Methodology Review: Assessing Unidimensionality of Tests and ltems. Appl. Psychol. Meas. 9 (2), 139–164. DOI: 10.1177/014662168500900204.

Heilala, V. 2018. Framework for pedagogical learning analytics. PhD thesis. University of Jyväskylä.

Hemker, B. T., Sijtsma, K., Molenaar, I. W., and Junker, B. W. 1997. Stochastic ordering using the latent trait and the sum score in polytomous IRT models. Psychometrika 62 (3), 331–347. DOI: 10.1007/BF02294555.

Hershberger, S. L. 2014. Factor Score Estimation. In Wiley StatsRef: Statistics Reference Online. Chichester, UK: John Wiley & Sons, Ltd. DOI: 10.1002/9781118445112.stat06532.

Hofmann, R. J. 1978. Complexity And Simplicity As Objective Indices Descriptive Of Factor Solutions. Multivariate Behavioral Research 13 (2), 247–250. DOI: 10.1207/s15327906mbr1302_9.

Holgado–Tello, F. P., Chacón–Moscoso, S., Barbero–García, I., and Vila–Abad, E. 2010. Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. Quality & Quantity 44 (1), 153–166. DOI: 10.1007/s11135-008-9190-y.

Holm, A.-S. 2010. Gender Patterns and Student Agency: Secondary School Students' Perceptions over Time. European Educational Research Journal 9 (2), 257–268. DOI: 10.2304/eerj.2010.9.2.257.

Horn, J. L. 1965. A Rationale and Test for the Number of Factors in Factor Analysis. Psychometrika 30, 179–185. DOI: 10.1007/BF02289447.

Hu, L.-t. and Bentler, P. M. 1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling 6 (1), 1–55. DOI: 10.1080/10705519909540118.

Huber, P. J. 1981. Robust Statistics. Wiley series in probability and mathematical statistics. New York, NY: John Wiley & Sons.

Hussein, A. 2009. The use of Triangulation in Social Sciences Research: Journal of Comparative Social Work 4 (1), 106–117. DOI: 10.31265/jcsw.v4i1.48.

Ianculescu, M. and Alexandru, A. 2020. Microservices–A Catalyzer for Better Managing Healthcare Data Empowerment. Studies in Informatics and Control 29 (2), 231–242.

Ifenthaler, D. 2021. Learning analytics for school and system management. In OECD Digital Education Outlook 2021: Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots. Ed. by OECD. Paris: OECD Publishing, 161. DOI: 10.1787/589b283f-en.

Ifenthaler, D., Gibson, D., Prasse, D., Shimada, A., and Yamada, M. 2021. Putting learning back into learning analytics: actions for policy makers, researchers, and practitioners. Educational technology research and development: ETR & D 69 (4), 2131–2150. DOI: 10.1007/s11423-020-09909-8.

Ifenthaler, D. and Greiff, S. 2021. Leveraging Learning Analytics for Assessment and Feedback. In Online Learning Analytics. Ed. by Liebowitz, J. New York, NY: Auerbach Publications, 1–18.

Ifenthaler, D. and Schumacher, C. 2016. Student perceptions of privacy principles for learning analytics. Educ. Technol. Res. Dev. 64 (5), 923–938. DOI: 10.1007/s11423-016-9477-y.

Ifenthaler, D. and Yau, J. Y. 2020. Utilising learning analytics to support study success in higher education: a systematic review. ETR&D Educational technology research and development 68 (4), 1961–1990. DOI: 10.1007/s11423-020-09788-z.

International Educational Data Mining Society 2022. educationaldatamining.org. https://web.archive.org/web/20220204120140/https://educationaldatamining.org/. Accessed: 2022-2-6.

Jääskelä, P., Poikkeus, A.-M., Häkkinen, P., Vasalampi, K., Rasku-Puttonen, H., and Tolvanen, A. 2020. Students' agency profiles in relation to student-perceived teaching practices in university courses. Int. J. Educ. Res. 103, 101604. DOI: 10.1016/j.ijer.2020.101604.

Jääskelä, P., Poikkeus, A.-M., Vasalampi, K., Valleala, U. M., and Rasku-Puttonen, H. 2017. Assessing agency of university students: validation of the AUS Scale. Studies in Higher Education 42 (11), 1–19. DOI: 10.1080/03075079.2015.1130693.

Jaccard, J. and Jacoby, J. 2020. Theory Construction and Model-Building Skills, Second Edition: A Practical Guide for Social Scientists. second. New York, NY: The Guilford Press.

Jain, A. K., Murty, M. N., and Flynn, P. J. 1999. Data clustering: a review. ACM Comput. Surv. 31 (3), 264–323. DOI: 10.1145/331499.331504.

Jamieson, D. 2018. Animal Agency. The Harvard Review of Philosophy 25, 111–126. DOI: 10.5840/harvardreview201892518.

Jovanović, J., Saqr, M., Joksimović, S., and Gašević, D. 2021. Students matter the most in learning analytics: The effects of internal and instructional conditions in predicting academic success. Comput. Educ. 172, 104251. DOI: 10.1016/j.compedu.2021.104251.

Kangas, M., Vesterinen, O., Lipponen, L., Kopisto, K., Salo, L., and Krokfors, L. 2014. Students' agency in an out-of-classroom setting: Acting accountably in a gardening project. Learning, Culture and Social Interaction 3 (1), 34–42. DOI: 10.1016/j.lcsi.2013.12.001.

Kärkkäinen, T. and Äyrämö, S. 2004. Robust clustering methods for incomplete and erroneous data. WIT Transactions on Information and Communication Technologies 33.

Kärkkäinen, T. and Heikkola, E. 2004. Robust formulations for training multilayer perceptrons. Neural Comput. 16 (4), 837–862. DOI: 10.1162/089976604322860721.

Kelley, K. and Pornprasertmanit, S. 2016. Confidence intervals for population reliability coefficients: Evaluation of methods, recommendations, and software for composite measures. Psychol. Methods 21 (1), 69–92. DOI: 10.1037/a0040086.

Khan, A. I., Kaliteevskii, V., Shnai, I., and Chechurin, L. 2020. Analysis of Students' Performance in an Online Discussion Forum: A Social Network Approach. In Cyber-Physical Systems and Control CPS&C 2019. Ed. by Arseniev, D., Overmeyer, L., Kälviäinen, H., and Katalinić, B. Cham: Springer, 670–680. DOI: 10.1007/978-3-030-34983-7_66.

Kihlstrom, J. F. 2021. Ecological Validity and "Ecological Validity". Perspectives on psychological science: a journal of the Association for Psychological Science 16 (2), 466–471. DOI: 10.1177/1745691620966791.

Kim, Y. and Baylor, A. 2016. Research-Based Design of Pedagogical Agent Roles: a Review, Progress, and Recommendations. International Journal of Artificial Intelligence in Education 26 (1), 160–169. DOI: 10.1007/s40593-015-0055-y.

King, A. 1983. Agency, Achievement, and Self-Concept of Young Adolescent Art Students. Studies in Art Education 24 (3), 187–194. DOI: 10.1080/00393541.1983.11650342.

Kitto, K. and Knight, S. 2019. Practical ethics for building learning analytics. Br. J. Educ. Technol. 50 (6), 2855–2870. DOI: 10.1111/bjet.12868.

Klemenčič, M. 2015. What is student agency? An ontological exploration in the context of research on student engagement. In Student engagement in Europe: society, higher education and student governance. Ed. by Klemenčič, M., Bergan, S., and Primožič, R. Vol. 20. Council of Europe Higher Education Series. Strasbourg, France: Council of Europe Publishing, 11–29.

Knight, S., Shum, S. B., and Littleton, K. 2014. Epistemology, Assessment, Pedagogy: Where Learning Meets Analytics in the Middle Space. Learning Analytics 1 (2), 23–47. DOI: 10.18608/jla.2014.12.3.

Krippendorff, K. 2013. Content Analysis: An Introduction to Its Methodology. third. Thousand Oaks, CA: SAGE Publications.

Kruskal, W. H. and Wallis, W. A. 1952. Use of Ranks in One-Criterion Variance Analysis. J. Am. Stat. Assoc. 47 (260), 583–621. DOI: 10.1080/01621459.1952.10483441.

Lauritzen, S. L. 1996. Graphical Models. Oxford: Clarendon Press.

Lemay, D. J., Baek, C., and Doleck, T. 2021. Comparison of learning analytics and educational data mining: A topic modeling approach. Computers and Education: Artificial Intelligence 2, 100016. DOI: 10.1016/j.caeai.2021.100016.

Li, C.-H. 2016a. Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. Behav. Res. Methods 48 (3), 936–949. DOI: 10.3758/s13428-015-0619-7.

Li, C.-H. 2016b. The performance of ML, DWLS, and ULS estimation with robust corrections in structural equation models with ordinal variables. Psychol. Methods 21 (3), 369–387. DOI: 10.1037/met0000093.

Li, C.-H. 2021. Statistical estimation of structural equation models with a mixture of continuous and categorical observed variables. Behav. Res. Methods. DOI: 10.3758/s13428-021-01547-z.

Li, M. X., Nadj, M., Maedche, A., Ifenthaler, D., and Wöhler, J. 2021. Towards a physiological computing infrastructure for researching students' flow in remote learning. Technology Knowledge and Learning. DOI: 10.1007/s10758-021-09569-4.

Liddell, K., Simon, D. A., and Lucassen, A. 2021. Patient data ownership: who owns your health? J Law Biosci 8 (2), lsab023. DOI: 10.1093/jlb/lsab023.

Lim, S. and Jahng, S. 2019. Determining the number of factors using parallel analysis and its recent variants. Psychological Methods 24 (4), 452–467. DOI: 10.1037/met0000230.

Lincoln, Y. S. 1995. In search of students' voices. Theory Pract. 34 (2), 88–93. DOI: 10.1080/00405849509543664.

Loevinger, J. 1947. A systematic approach to the construction and evaluation of tests of ability. Psychol. Monogr. 61 (4), i.

Loevinger, J. 1948. The technic of homogeneous tests compared with some aspects of scale analysis and factor analysis. Psychol. Bull. 45 (6), 507–529. DOI: 10.1037/h0055827.

Lord, F. M. and Novick, M. R. 1968. Statistical Theories of Mental Test Scores. Reading, MA: Addison-Wesley.

Lundberg, S. M. and Lee, S.-I. 2017. A Unified Approach to Interpreting Model Predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems. NIPS'17. Long Beach, California, USA: Curran Associates Inc., 4768–4777.

MacQueen, J. 1967. Some methods for classification and analysis of multivariate observations. In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability. Vol. 1, 281–297.

Maguire, M. and Delahunt, B. 2017. 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).

Massara, G. P., Di Matteo, T., and Aste, T. 2016. Network Filtering for Big Data: Triangulated Maximally Filtered Graph. J Complex Netw 5 (2), 161–178. DOI: 10.1093/comnet/cnw015.

Masters, G. N. 2016. Partial Credit Model. In Handbook of Item Response Theory. Ed. by Linden, W. J. v. Boca Raton, FL: CRC Press, 109–126. DOI: 10.1201/9781315374512.

Masters, G. N. 1982. A rasch model for partial credit scoring. Psychometrika 47 (2), 149–174. DOI: 10.1007/BF02296272.

Mathrani, A., Susnjak, T., Ramaswami, G., and Barczak, A. 2021. Perspectives on the challenges of generalizability, transparency and ethics in predictive learning analytics. Computers and Education Open 2, 100060. DOI: 10.1016/j.caeo.2021.100060.

Maydeu-Olivares, A. 2017. Assessing the Size of Model Misfit in Structural Equation Models. Psychometrika 82 (3), 533–558. DOI: 10.1007/s11336-016-9552-7.

McDonald, R. P. 1981. The dimensionality of tests and items. Br. J. Math. Stat. Psychol. 34 (1), 100–117. DOI: 10.1111/j.2044-8317.1981.tb00621.x.

Mcintyre, A. 2006. Activist Research and Student Agency in Universities and Urban Communities. Urban Education 41 (6), 628–647. DOI: 10.1177/0042085906292510.

McNeish, D. and Wolf, M. G. 2021. Dynamic fit index cutoffs for confirmatory factor analysis models. Psychological Methods. DOI: 10.1037/met0000425.

Mikk, K. A., Sleeper, H. A., and Topol, E. J. 2017. The Pathway to Patient Data Ownership and Better Health. JAMA 318 (15), 1433–1434. DOI: 10.1001/jama.2017.12145.

Mikk, K. A., Sleeper, H. A., and Topol, E. J. 2018. Patient Data Ownership—Reply. JAMA 319 (9), 935–936. DOI: 10.1001/jama.2017.21688.

Miller, D. 1978. The role of multivariate "Q-techniques" in the study of organizations. Acad. Manage. Rev. 3 (3), 515–531. DOI: 10.5465/amr.1978.4305745.

Mislevy, R. J. 2019. On integrating psychometrics and learning analytics in complex assessments. In Data Analytics and Psychometrics. Ed. by Jiao, H., Lissitz, R. W., and Wie, A. van. The MARCES Book Series. Charlotte, NC: Information Age Publishing, 1–52.

Mokken, R. J. 1971. A theory and procedure of scale analysis: with applications in political research. PhD thesis. The Hague: University of Amsterdam.

Molenaar, I. W. 1991. A weighted Loevinger H-coefficient extending Mokken scaling to multicategory items. Kwantitatieve Methoden 12 (37), 97–117.

Molenaar, I. W. 1997. Nonparametric Models for Polytomous Responses. In Handbook of Modern Item Response Theory. Ed. by Linden, W. J. v. and Hambleton, R. K. New York, NY: Springer New York, 369–380. DOI: 10.1007/978-1-4757-2691-6_21.

Myers, M. D. 2020. Qualitative Research in Business and Management. third. Thousand Oaks, CA: SAGE Publications.

Namoun, A. and Alshanqiti, A. 2020. Predicting Student Performance Using Data Mining and Learning Analytics Techniques: A Systematic Literature Review. Applied sciences 11 (1), 237. DOI: 10.3390/app11010237.

Nations, U. 1948. Universal declaration of human rights. UN General Assembly resolution 217 A, 71–77.

Nembrini, S., König, I. R., and Wright, M. N. 2018. The revival of the Gini importance? Bioinformatics 34 (21), 3711–3718. DOI: 10.1093/bioinformatics/bty373.

Nguyen, A., Tuunanen, T., Gardner, L., and Sheridan, D. 2021. Design principles for learning analytics information systems in higher education. European Journal of Information Systems 30 (5), 541–568. DOI: 10.1080/0960085X.2020.1816144.

Nguyen, H., Harpstead, E., Wang, Y., and McLaren, B. M. 2018. Student Agency and Game-Based Learning: A Study Comparing Low and High Agency. In Artificial Intelligence in Education. Springer International Publishing, 338–351. DOI: 10.1007/978-3-319-93843-1_25.

NORAD 2021. NORAD, USNORTHCOM lead 3rd Global Information Dominance Experiment. U. S. Air Force. Accessed: 2022-2-15.

Nunnally, J. C. and Bernstein, I. H. 1994. Psychometric theory 3E. third. New York, NY: McGraw-Hill.

OECD 2019. OECD Skills Strategy 2019 Skills to Shape a Better Future: Skills to Shape a Better Future. Paris: OECD Publishing. DOI: 10.1787/9789264313835-en.

Onwuegbuzie, A. J. and Leech, N. L. 2005. On Becoming a Pragmatic Researcher: The Importance of Combining Quantitative and Qualitative Research Methodologies. Int. J. Soc. Res. Methodol. 8 (5), 375–387. DOI: 10.1080/13645570500402447.

Orne, M. T. 1962. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. The American psychologist 17 (11), 776–783. DOI: 10.1037/h0043424.

Ostrovsky, R., Rabani, Y., Schulman, L. J., and Swamy, C. 2006. The Effectiveness of Lloyd-Type Methods for the k-Means Problem. In 2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06). Berkeley, CA: IEEE, 165–176. DOI: 10.1109/FOCS.2006.75.

Ottey, S. D. 1996. Critical Pedagogical Theory and the Dance Educator. Arts Education Policy Review 98 (2), 31–39. DOI: 10.1080/10632913.1996.9935096.

Pakhira, M. K., Bandyopadhyay, S., and Maulik, U. 2004. Validity index for crisp and fuzzy clusters. Pattern Recognit. 37 (3), 487–501. DOI: 10.1016/j.patcog.2003.06.005.

Pargman, T. C. and McGrath, C. 2021. Mapping the Ethics of Learning Analytics in Higher Education: A Systematic Literature Review of Empirical Research. Journal of Learning Analytics 8 (2), 123–139. DOI: 10.18608/jla.2021.1.

Peeters, R. 2020. The agency of algorithms: Understanding human-algorithm interaction in administrative decision-making. Inf. polity 25 (4), 507–522. DOI: 10.3233/ip-200253.

Pons, P. and Latapy, M. 2006. Computing Communities in Large Networks Using Random Walks. J. Graph Algorithms Appl. 10 (2), 191–218. DOI: 10.7155/jgaa.00124.

Pressey, S. L. 1927. A machine for automatic teaching of drill material. School & Society 25, 549–552.

Pressey, S. L. 1932. A third and fourth contribution toward the coming" industrial revolution" in education. School & Society 36, 668–672.

Pressey, S. L. 1926. A simple apparatus which gives tests and scores-and teaches. Sch. & Soc. 23, 373–376.

Prieto, L. P., Sharma, K., Dillenbourg, P., and Jesús, M. 2016. Teaching analytics: towards automatic extraction of orchestration graphs using wearable sensors. In Proceedings of the Sixth International Conference on Learning Analytics & Knowledge. LAK '16. Edinburgh, United Kingdom: Association for Computing Machinery, 148–157. DOI: 10.1145/2883851.2883927.

Puranen-Impola, K. 2022. Kurinalaistamisen ja kategorisoinnin keino. Rangaistuskäytänteet seitsemässä suomalaisessa oppikoulussa 1905–1925. PhD thesis. University of Jyväskylä.

Qin, H. and Wang, G. 2022. Benefits, Challenges and Solutions of Artificial Intelligence Applied in Education. In 2022 11th International Conference on Educational and Information Technology (ICEIT), 62–66. DOI: 10.1109/ICEIT54416.2022.9690739.

Reichheld, F. F. 2003. The one number you need to grow. Harv. Bus. Rev. 81 (12), 46–54, 124.

Ribeiro, M. T., Singh, S., and Guestrin, C. 2016. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '16. San Francisco, California, USA: Association for Computing Machinery, 1135–1144. DOI: 10.1145/2939672.2939778. ⟨URL: https://doi.org/10.1145/2939672.2939778⟩.

Riessman, C. 2002. Narrative analysis. In The qualitative researcher's companion. Ed. by Huberman, A. M. and Miles, M. B. Thousand Oaks, CA: SAGE Publications, 216–270. DOI: 10.4135/9781412986274.

Riessman, C. K. 1993. Narrative Analysis. Thousand Oaks, CA: SAGE Publications.

Risjord, M., Moloney, M., and Dunbar, S. 2001. Methodological Triangulation in Nursing Research. Philos. Soc. Sci. 31 (1), 40–59. DOI: 10.1177/004839310103100103.

Romero, C. and Ventura, S. 2020. Educational data mining and learning analytics: An updated survey. WIREs Data Mining and Knowledge Discovery 10 (3). DOI: 10.1002/widm.1355.

Romero, C. and Ventura, S. 2017. Educational data science in massive open online courses. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 7 (1), e1187. DOI: 10.1002/widm.1187.

Rosé, C. P. 2018. Learning Analytics in the Learning Sciences. In International Handbook of the Learning Sciences. Ed. by Fischer, F., Hmelo-Silver, C. E., Goldman, S. R., and Reimann, P. New York: Routledge, 511–519.

Rousseeuw, P. J. and Hubert, M. 2018. Anomaly detection by robust statistics. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 8 (2), e1236. DOI: 10.1002/widm.1236.

Saarela, M. 2017. Automatic knowledge discovery from sparse and large-scale educational data: case Finland. PhD thesis. University of Jyväskylä.

Saarela, M., Hämäläinen, J., and Kärkkäinen, T. 2017. Feature Ranking of Large, Robust, and Weighted Clustering Result. In Advances in Knowledge Discovery and Data Mining. Springer International Publishing, 96–109. DOI: 10.1007/978-3-319-57454-7_8.

Saarela, M. and Kärkkäinen, T. 2017. Knowledge Discovery from the Programme for International Student Assessment. In Learning Analytics: Fundaments, Applications, and Trends. Ed. by Peña-Ayala, A. Studies in Systems, Decision and Control. Cham: Springer, 229–267. DOI: 10.1007/978-3-319-52977-6_8.

Sahin, M. and Ifenthaler, D. 2022. Interaction Preferences in Digital Learning Environments: Does Gender and Achievement Matter? In Orchestration of Learning Environments in the Digital World. Ed. by Ifenthaler, D., Isaías, P., and Sampson, D. G. Cham: Springer International Publishing, 231–244. DOI: 10.1007/978-3-030-90944-4_13.

Samuelsen, J., Chen, W., and Wasson, B. 2019. Integrating multiple data sources for learning analytics—review of literature. Research and Practice in Technology Enhanced Learning 14 (1). DOI: 10.1186/s41039-019-0105-4.

Sarikaya, A., Correll, M., Bartram, L., Tory, M., and Fisher, D. 2018. What Do We Talk About When We Talk About Dashboards? IEEE transactions on visualization and computer graphics. DOI: 10.1109/TVCG.2018.2864903.

Savalei, V. 2018. On the Computation of the RMSEA and CFI from the Mean-And-Variance Corrected Test Statistic with Nonnormal Data in SEM. Multivariate Behavioral Research 53 (3), 419–429. DOI: 10.1080/00273171.2018.1455142.

Scardamalia, M. and 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 (1), 37–68. DOI: 10.1207/s15327809jls0101_3.

Schauberger, G. and Mair, P. 2020. A regularization approach for the detection of differential item functioning in generalized partial credit models. Behav. Res. Methods 52 (1), 279–294. DOI: 10.3758/s13428-019-01224-2.

Scherer, S., Worsley, M., and Morency, L.-P. 2012. 1st international workshop on multimodal learning analytics: extended abstract. In Proceedings of the 14th ACM international conference on Multimodal interaction. ICMI '12. Santa Monica, California, USA: Association for Computing Machinery, 609–610. DOI: 10.1145/2388676.2388803.

Schlosser, M. 2019. Agency. In The Stanford Encyclopedia of Philosophy. Ed. by Zalta, E. N. Winter 2019. Metaphysics Research Lab, Stanford University.

Schön, D. and Ifenthaler, D. 2018. Prompting in pseudonymised learning analytics - implementing learner centric prompts in legacy systems with high privacy requirements. In Proceedings of the 10th International Conference on Computer Supported Education. Funchal, Madeira, Portugal: SCITEPRESS - Science and Technology Publications. DOI: 10.5220/0006780903830389.

Schön, D. A. 1983. The reflective practitioner: How professionals think in action. New York, NY: Basic Books.

Shapley, L. S. 1953. A Value for n-Person Games. In Contributions to the Theory of Games (AM-28), Volume II. Ed. by Kuhn, H. W. and Tucker, A. W. Princeton: Princeton University Press, 307–318. DOI: 10.1515/9781400881970-018.

El-Sheikh, A. A., Abonazel, M. R., and Gamil, N. 2017. A review of software packages for structural equation modeling: A comparative study. Applied Mathematics and Physics 5 (3), 85–94. DOI: 10.12691/amp-5-3-2.

Shi, D., Lee, T., and Maydeu-Olivares, A. 2019. Understanding the Model Size Effect on SEM Fit Indices. Educational and Psychological Measurement 79 (2), 310–334. DOI: 10.1177/0013164418783530.

Shi, D. and Maydeu-Olivares, A. 2020. The Effect of Estimation Methods on SEM Fit Indices. Educational and Psychological Measurement 80 (3), 421–445. DOI: 10.1177/0013164419885164.

Shortliffe, E. H., Davis, R., Axline, S. G., Buchanan, B. G., Green, C. C., and Cohen, S. N. 1975. Computer-based consultations in clinical therapeutics: explanation and rule acquisition capabilities of the MYCIN system. Comput. Biomed. Res. 8 (4), 303–320. DOI: 10.1016/0010-4809(75)90009-9.

Shum, S. B. and Crick, R. D. 2012. Learning dispositions and transferable competencies: pedagogy, modelling and learning analytics. In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge. LAK '12. Vancouver, British Columbia, Canada: Association for Computing Machinery, 92–101. DOI: 10.1145/2330601.2330629.

Siemens, G. 2013. Learning Analytics: The Emergence of a Discipline. Am. Behav. Sci. 57 (10), 1380–1400. DOI: 10.1177/0002764213498851.

Siemens, G. and Long, P. 2011. Penetrating the Fog: Analytics in Learning and Education. EDUCAUSE Review 46 (5), 30.

Sijtsma, K. and Ark, L. A. v. 2020. Measurement Models for Psychological Attributes. Boca Raton, FL: CRC Press. DOI: 10.1201/9780429112447.

Sijtsma, K. and Hemker, B. T. 2000. A Taxonomy of IRT Models for Ordering Persons and Items Using Simple Sum Scores. Journal of educational and behavioral statistics: a quarterly publication sponsored by the American Educational Research Association and the American Statistical Association 25 (4), 391–415. DOI: 10.3102/10769986025004391.

Sijtsma, K. and Molenaar, I. W. 2002. Introduction to Nonparametric Item Response Theory. Thousand Oaks, CA: SAGE Publications.

Sijtsma, K. and Pfadt, J. M. 2021. Part II: On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha: Discussing Lower Bounds and Correlated Errors. Psychometrika. DOI: 10.1007/s11336-021-09789-8.

Sitra 2020. Elinikäinen oppiminen Suomessa 2019 -kyselyn tulokset. Helsinki.

Slade, S. and Prinsloo, P. 2013. Learning Analytics: Ethical Issues and Dilemmas. Am. Behav. Sci. 57 (10), 1510–1529. DOI: 10.1177/0002764213479366.

Smith, B. J. 2000. Marginalized Youth, Delinquency, and Education: The Need for Critical-Interpretive Research. Urban Rev. 32 (3), 293–312. DOI: 10.1023/A:1005126412773.

Stenalt, M. H. and Lassesen, B. 2021. Does student agency benefit student learning? A systematic review of higher education research. Assessment & Evaluation in Higher Education, 1–17. DOI: 10.1080/02602938.2021.1967874.

Steward, H. 2009. Animal Agency. Inquiry 52 (3), 217–231. DOI: 10.1080/00201740902917119.

Stokes, D. E. 1997. Pasteur's Quadrant: Basic Science and Technological Innovation. Washington, DC: Brookings Institution Press.

Straat, J. H., Ark L, A. van der, and Sijtsma, K. 2016. Using Conditional Association to Identify Locally Independent Item Sets. Methodology 12 (4), 117–123. DOI: 10.1027/1614-2241/a000115.

Suchman, E. A. 1950. The scalogram board technique for scale analysis. In Measurement and prediction. Ed. by Stouffer, S. A. Studies in Social Psychology in World War II. Princeton, NJ: Princeton University Press, 91–121.

Taub, M., Sawyer, R., Smith, A., Rowe, J., Azevedo, R., and Lester, J. 2020. The agency effect: The impact of student agency on learning, emotions, and problem-solving behaviors in a game-based learning environment. Computers & Education 147, 103781. DOI: 10.1016/j.compedu.2019.103781.

Teasley, S. D. 2018. Learning analytics: where information science and the learning sciences meet. Information and Learning Sciences 120 (1/2), 59–73. DOI: 10.1108/ILS-06-2018-0045.

Tempelaar, D., Rienties, B., and Nguyen, Q. 2020. Subjective data, objective data and the role of bias in predictive modelling: Lessons from a dispositional learning analytics application. PLoS One 15 (6), e0233977. DOI: 10.1371/journal.pone.0233977.

Tempelaar, D. T., Rienties, B., and Nguyen, Q. 2017. Towards Actionable Learning Analytics Using Dispositions. IEEE Trans. Learn. Technol. 10 (1), 6–16. DOI: 10.1109/TLT.2017.2662679.

Thomson, W. 1884. Electrical Units of Measurement. In The Practical Applications of Electricity. A Series of Lectures Delivered at the Institution of Civil Engineers, Session 1882-83. Westminster: The Institution.

Thorndike, E. L. 1912. Education: a first book. New York City, NY: The Macmillan Company.

Timmerman, M. E. and Lorenzo-Seva, U. 2011. Dimensionality assessment of ordered polytomous items with parallel analysis. Psychological Methods 16 (2), 209–220. DOI: 10.1037/a0023353.

Tompkins, J. 1990. Pedagogy of the Distressed. College English 52 (6), 653–660. DOI: 10.2307/378032.

Tsai, Y.-S. and Gasevic, D. 2017. Learning analytics in higher education — challenges and policies: a review of eight learning analytics policies. In Proceedings of the Seventh International Learning Analytics & Knowledge Conference. LAK '17. Vancouver, British Columbia, Canada: Association for Computing Machinery, 233–242. DOI: 10.1145/3027385.3027400.

Tsai, Y.-S., Perrotta, C., and Gašević, D. 2020. Empowering learners with personalised learning approaches? Agency, equity and transparency in the context of learning analytics. Assessment & Evaluation in Higher Education 45 (4), 554–567. DOI: 10.1080/02602938.2019.1676396.

Tsai, Y.-S., Rates, D., Moreno-Marcos, P. M., Muñoz-Merino, P. J., Jivet, I., Scheffel, M., Drachsler, H., Delgado Kloos, C., and Gašević, D. 2020. Learning analytics in European higher education—Trends and barriers. Computers & education 155, 103933. DOI: 10.1016/j.compedu.2020.103933.

Tzimas, D. and Demetriadis, S. 2021. Ethical issues in learning analytics: a review of the field. Educ. Technol. Res. Dev. 69 (2), 1101–1133. DOI: 10.1007/s11423-021-09977-4.

Valtioneuvosto 2020. Osaaminen turvaa tulevaisuuden: Jatkuvan oppimisen parlamentaarisen uudistuksen linjaukset. Valtioneuvoston julkaisuja 2020:38. Helsinki: Valtioneuvosto.

Vaughn, M. 2020. What is student agency and why is it needed now more than ever? Theory Into Practice 59 (2), 109–118. DOI: 10.1080/00405841.2019.1702393.

Vaughn, M. 2021. Student Agency in the Classroom: Honoring Student Voice in the Curriculum. New York, NY: Teachers College Press.

Vessonen, E. 2021. Mielen mittarit. Tampere: Vastapaino.

Vieira, C., Parsons, P., and Byrd, V. 2018. Visual learning analytics of educational data: A systematic literature review and research agenda. Comput. Educ. 122, 119–135. DOI: 10.1016/j.compedu.2018.03.018.

Villaescusa-Navarro, F., Genel, S., Angles-Alcazar, D., Thiele, L., Dave, R., Narayanan, D., Nicola, A., Li, Y., Villanueva-Domingo, P., Wandelt, B., Spergel, D. N., Somerville, R. S., Matilla, J. M. Z., Mohammad, F. G., Hassan, S., Shao, H., Wadekar, D., Eickenberg, M., Wong, K. W. K., Contardo, G., Jo, Y., Moser, E., Lau, E. T., Valle, L. F. M. P., Perez, L. A., Nagai, D., Battaglia, N., and Vogelsberger, M. 2021. The CAMELS Multifield Dataset: Learning the Universe's Fundamental Parameters with Artificial Intelligence. https://arxiv.org/abs/2109.10915. arXiv: 2109.10915.

Vogler, G. P., Baker, L. A., Decker, S. N., Defries, J. C., and Huizinga, D. H. 1989. Cluster analytic classification of reading disability subtypes. Read. Writ. 1 (2), 163–177. DOI: 10.1007/BF00377469.

von Davier, A. A., Mislevy, R. J., and Hao, J. 2021a. Computational Psychometrics: New Methodologies for a New Generation of Digital Learning and Assessment. London: Springer Nature.

von Davier, A. A., Mislevy, R. J., and Hao, J. 2021b. Introduction to Computational Psychometrics: Towards a Principled Integration of Data Science and Machine Learning Techniques into Psychometrics. In Computational Psychometrics: New Methodologies for a New Generation of Digital Learning and Assessment: With Examples in R and Python. Ed. by von Davier, A. A., Mislevy, R. J., and Hao, J. Cham: Springer International Publishing, 1–6. DOI: 10.1007/978-3-030-74394-9_1.

Walker, C. M. 2011. What's the DIF? Why differential item functioning analyses are an important part of instrument development and validation. J. Psychoeduc. Assess. 29 (4), 364–376. DOI: 10.1177/0734282911406666.

Walsh, C. 1994. Engaging students in learning: Literacy, language, and knowledge production with Latino adolescents. In Adult Biliteracy in the United States. Language in Education: Theory and Practice 83. Ed. by Spener, D. McHenry, IL: Delta Systems, Inc., 211–237.

Wang, Y., Kung, L., and Byrd, T. A. 2018. Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. Technological forecasting and social change 126, 3–13. DOI: 10.1016/j.techfore.2015.12.019.

Whitelock-Wainwright, A., Gašević, D., Tejeiro, R., Tsai, Y.-S., and Bennett, K. 2019. The student expectations of learning analytics questionnaire. J. Comput. Assist. Learn. 35 (5), 633–666. DOI: 10.1111/jcal.12366.

Wilson, A., Watson, C., Thompson, T. L., Drew, V., and Doyle, S. 2017. Learning analytics: challenges and limitations. Teaching in Higher Education 22 (8), 991–1007. DOI: 10.1080/13562517.2017.1332026.

Wimsatt, W. C. 1981. Robustness, Reliability, and Overdetermination. In Scientific inquiry in the social sciences. Ed. by Brewer, M. and Collins, B. San Francisco,CA: Jossey-Bass, 123–162.

Worren, N. A., Moore, K., and Elliott, R. 2002. When theories become tools: Toward a framework for pragmatic validity. Hum. Relat. 55 (10), 1227–1250. DOI: 10.1177/a028082.

Xia, Y. and Yang, Y. 2019. RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. Behavior Research Methods 51 (1), 409–428. DOI: 10.3758/s13428-018-1055-2.

Yau, J. Y. and Ifenthaler, D. 2021. Utilizing learning analytics for teaching success. In EdMedia+ Innovate Learning, 330–338.

Zawacki-Richter, O., Marín, V. I., Bond, M., and Gouverneur, F. 2019. Systematic review of research on artificial intelligence applications in higher education – where are the educators? International Journal of Educational Technology in Higher Education 16 (1), 1–27. DOI: 10.1186/s41239-019-0171-0.

Ziegler, M. 2020. Psychological Test Adaptation and Development – A new beginning. Psychological Test Adaptation and Development, 1–2. DOI: 10.1027/2698-1866/a000001.

Zwick, W. R. and Velicer, W. F. 1986. Comparison of five rules for determining the number of components to retain. Psychological Bulletin 99 (3), 432–442. DOI: 10.1037/0033-2909.99.3.432.

Zygmont, C. and Smith, M. R. 2014. Robust factor analysis in the presence of normality violations, missing data, and outliers: Empirical questions and possible solutions. The Quantitative Methods for Psychology 10 (1), 40–55. DOI: 10.20982/tqmp.10.1.p040.

ORIGINAL PAPERS

PI

STUDENT AGENCY ANALYTICS: LEARNING ANALYTICS AS A TOOL FOR ANALYZING STUDENT AGENCY IN HIGHER EDUCATION

by

Jääskelä, P., Heilala, V., Häkkinen, P., and Kärkkäinen, T 2021

Behavior and Information Technology, 40(8)

This is an Accepted Manuscript of an article published by Taylor & Francis in Behaviour & Information Technology on 11 Feb 2020, available online: http://www.tandfonline.com/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 Jyvaskyla, Finland*
[b] *Faculty of Information Technology, University of Jyvaskyla, 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 analytics 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 sociocultural 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 sociocultural 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 studies 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; Ayllon 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.

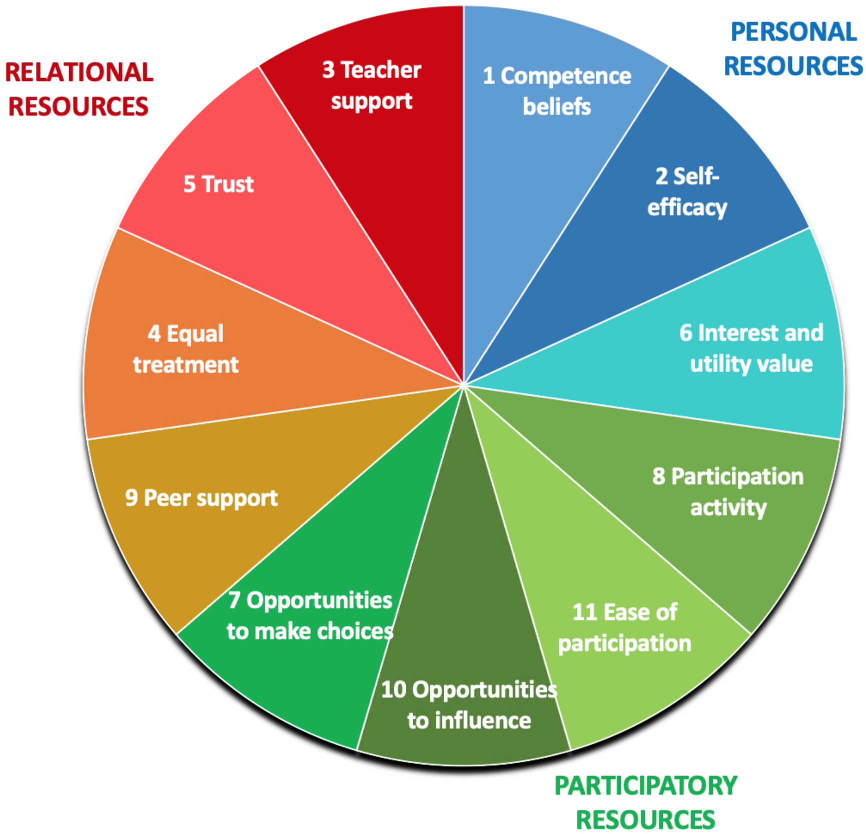


Figure 1: The dimensions of the Agency of University Student (AUS) Scale (Jääskelä et al., 2017b).

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 presented 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 topics 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: (χ^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 futher 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, nonparametric 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

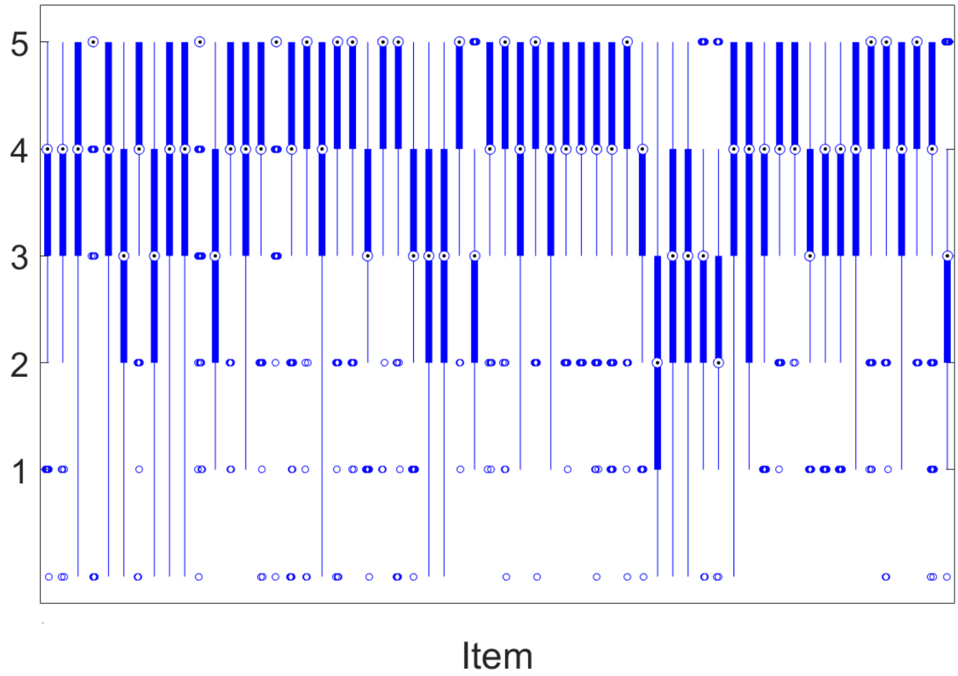


Figure 2: Boxplot of the reference dataset (missing value as zero).

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)

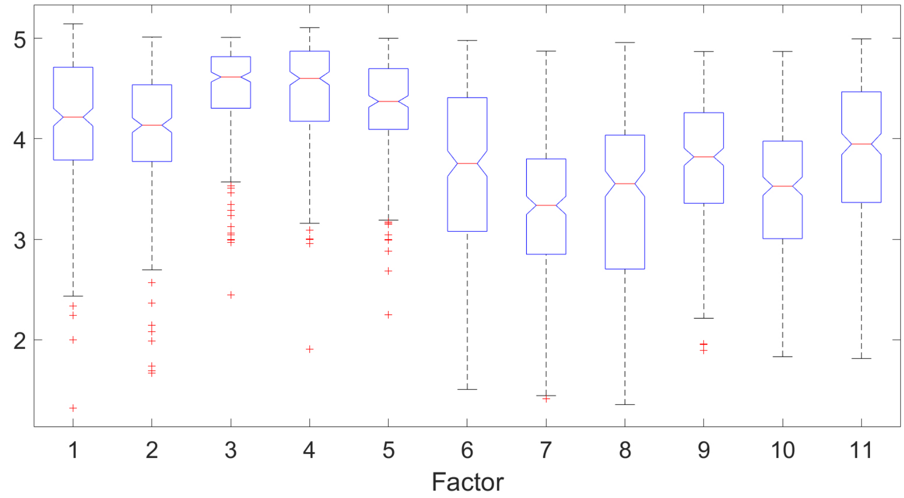


Figure 3: Boxplot of the agency factors.

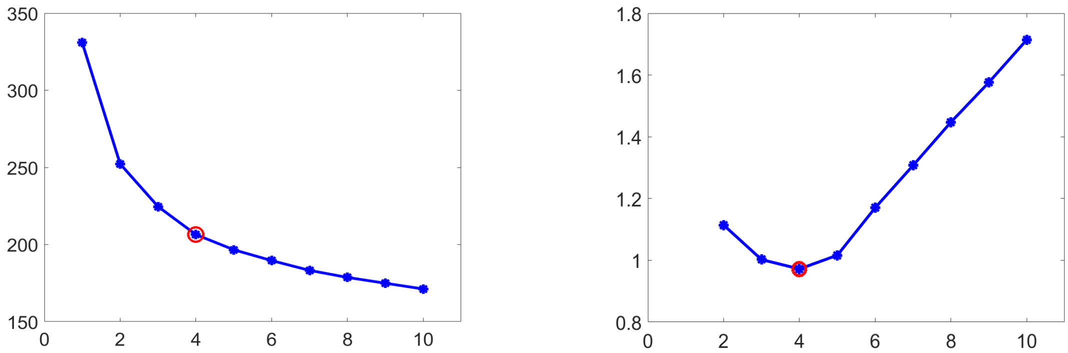


Figure 4: Clustering error (left) and the PBM CVI (right) for $K = 2 - 10$. Minimum of PBM ($K = 4$) marked.

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

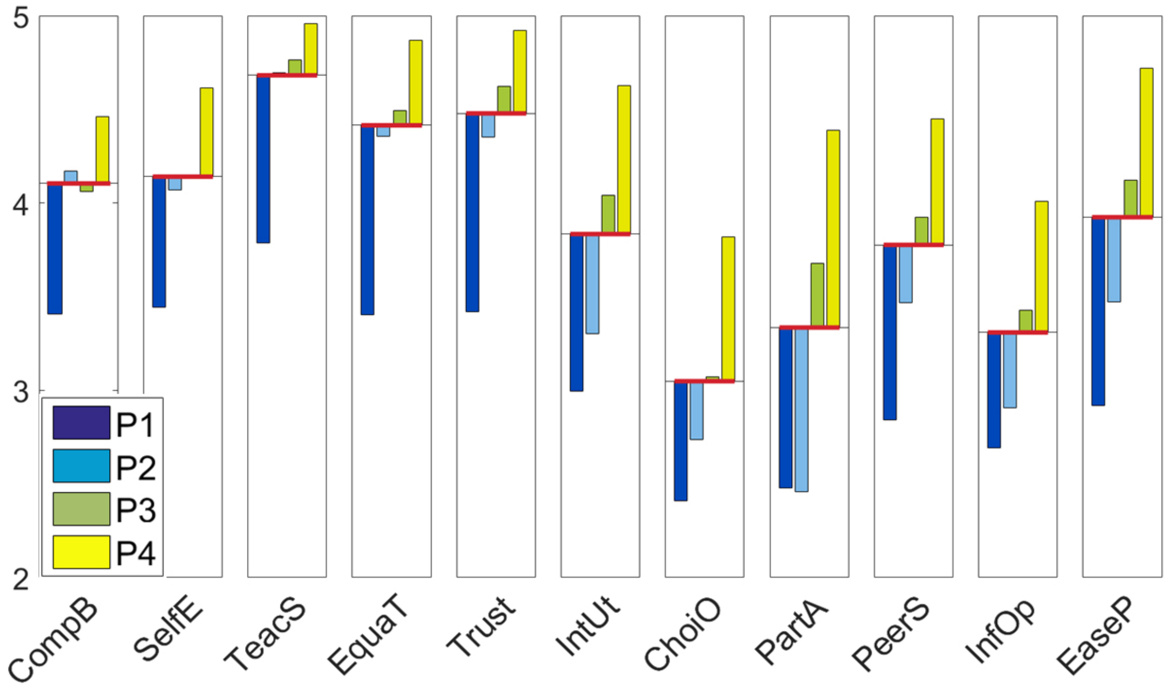


Figure 5: Deviations of the four agency profiles from the GAP in the reference dataset.

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 (SA^{4}S).

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 excecuted 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-Sneppe (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

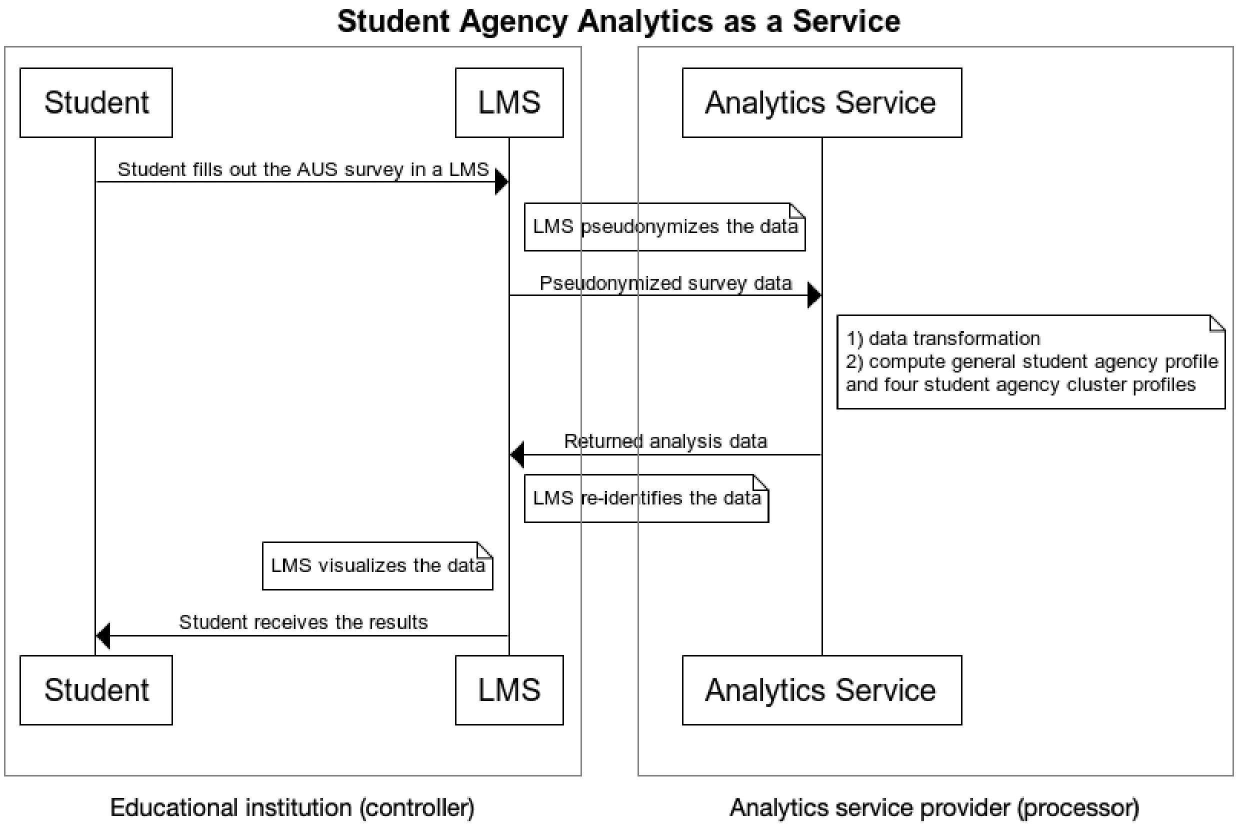


Figure 6: A sequence diagram presenting the flow of data between processor and controller.

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, $\chi^2(12, n = 92) = 27.9, p < .01$. 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.

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**
Agency profiles	**P1**	6	7	5	3	1
	P2	4	1	5	2	10
	P3	2	3	4	11	16
	P4	4	1	2	1	4

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

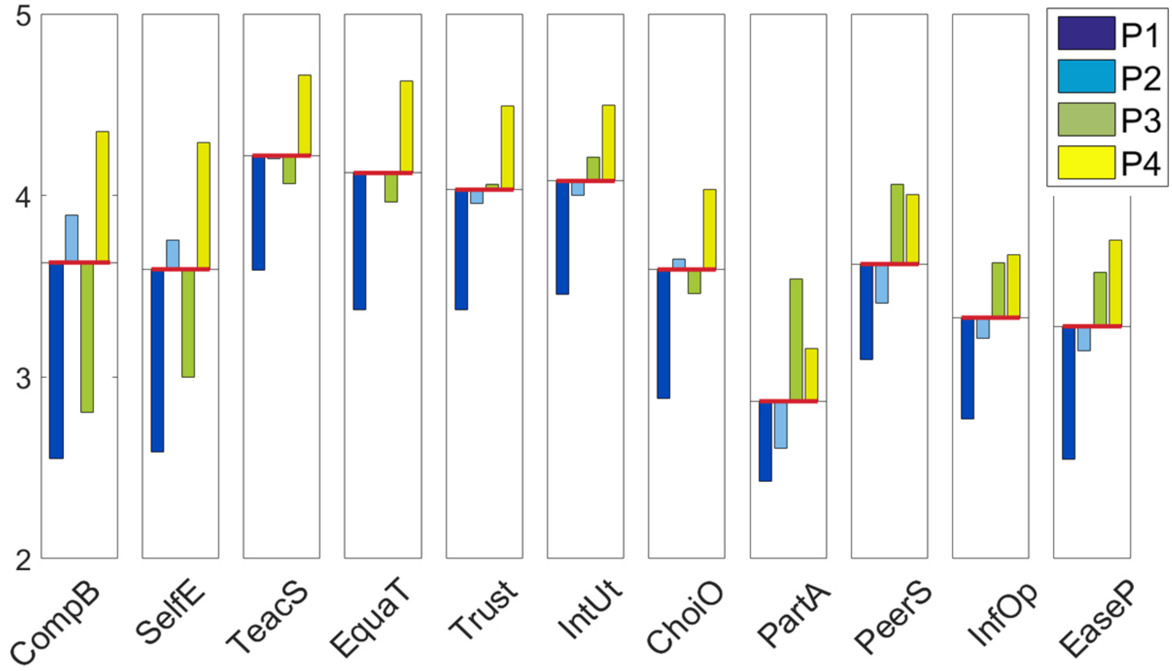


Figure 7: Deviations of the four agency profiles from the GAP in the Faculty of Information Technology.

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, $\chi^2(15, n = 64) = 16.3, p = 0.36$.

Agency profiles	Study groups: A	B	C	D	E	F
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.

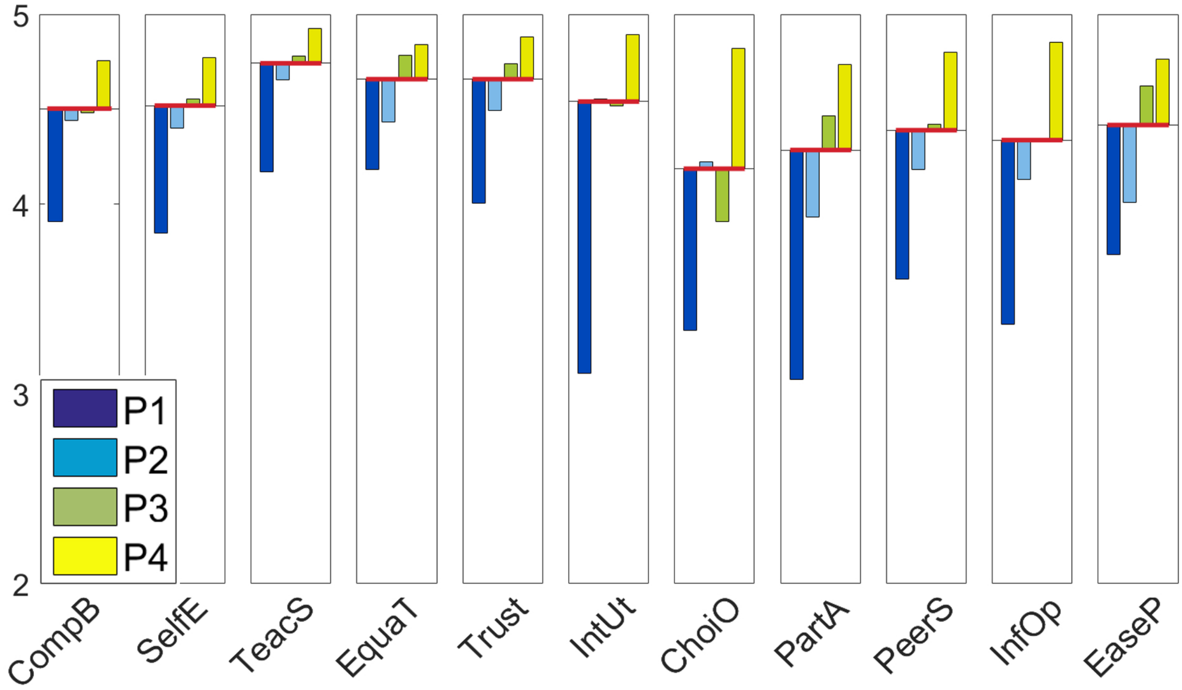


Figure 8: Deviations of the four agency profiles from the GAP in the Department of Teacher Education.

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 Pongsajapan, 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., Ambroise, C., Cocquerez, J.P., 2006. Feature selection in robust clustering based on laplace mixture. Pattern Recognition Letters 27, 627–635.

Damşa, 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 toweight 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-Sneppe, 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: Fundaments, 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.

PII

UNDERSTANDING THE STUDY EXPERIENCES OF STUDENTS IN LOW AGENCY PROFILE: TOWARDS A SMART EDUCATION APPROACH

by

Heilala V., Jääskelä P., Kärkkäinen T., and Saarela M 2020

Advances in Smart Technologies Applications and Case Studies. SmartICT 2019.
Lecture Notes in Electrical Engineering

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].

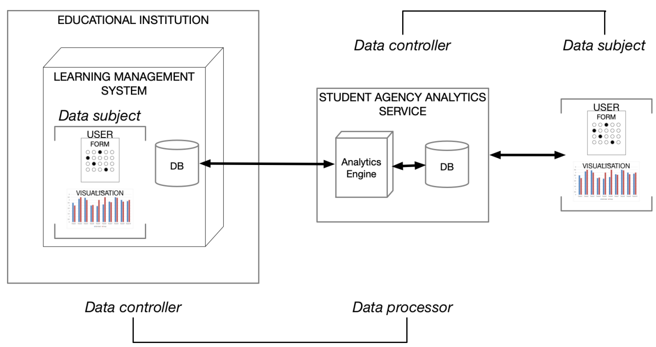


Fig. 1. Agency analytics service is either data controller or data processor depending on the use case.

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

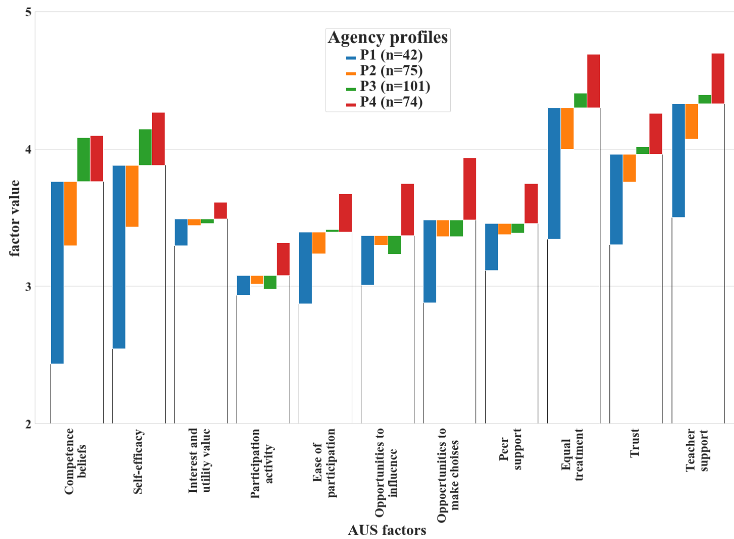


Fig. 2. General agency profile (GAP) and individual profile deviations (P1-P4). Factor values represent the original questionnaire Likert scale.

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

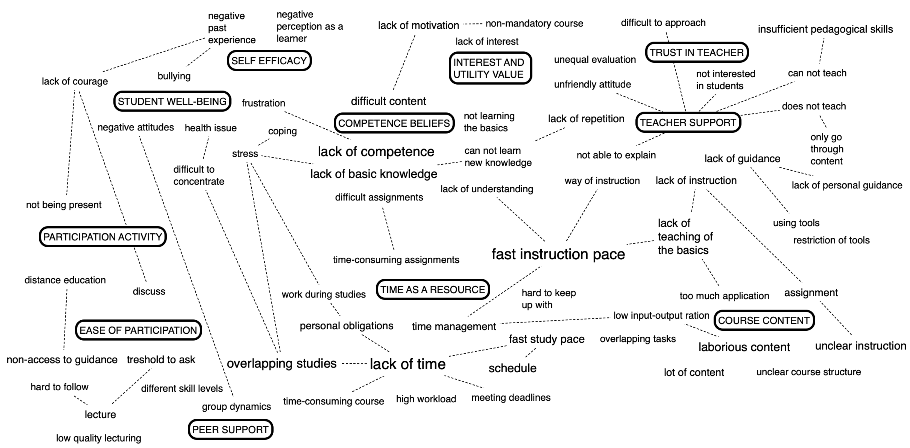


Fig. 3. The results of the thematic analysis: the thematic map and coding of the data in the low student agency profile (P1).

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

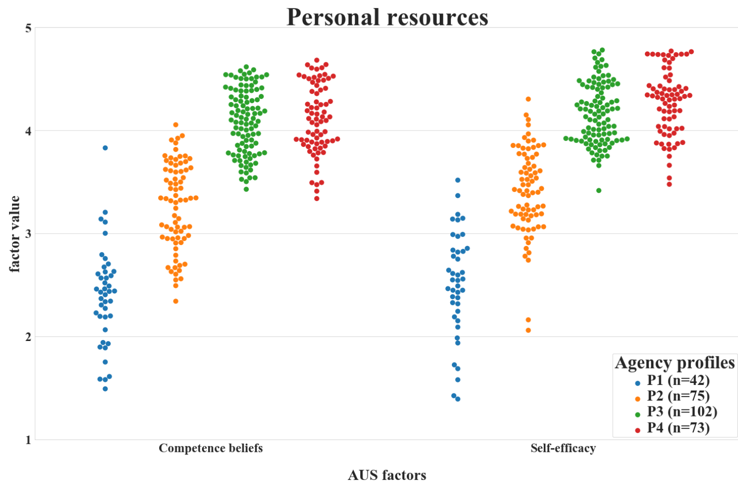


Fig. 4. The personal resources of student agency. Factor values represent the original questionnaire Likert scale.

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.

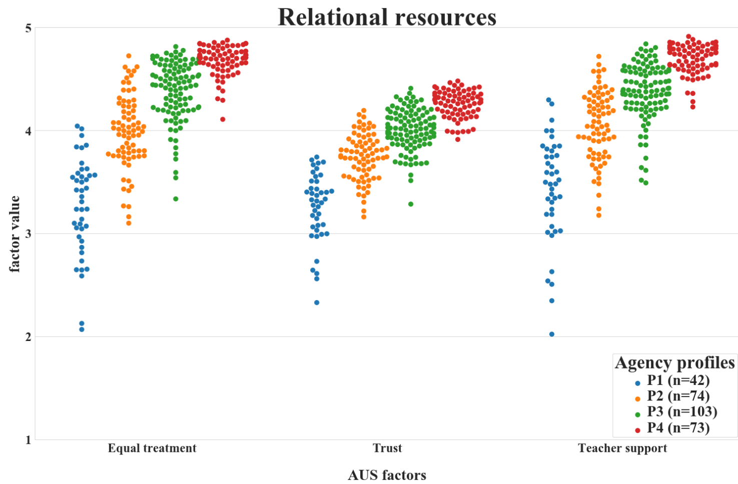


Fig. 5. The relational resources of student agency. Factor values represent the original questionnaire Likert scale.

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)

PIII

COURSE SATISFACTION IN ENGINEERING EDUCATION THROUGH THE LENS OF STUDENT AGENCY ANALYTICS

by

Heilala, V., Saarela, M., Jääskelä, P., and Kärkkäinen, T 2020

Proceedings of the 50th IEEE Frontiers in Education Conference (Conference proceedings: Frontiers in Education Conference)

This is an Accepted Manuscript of an article V. Heilala, M. Saarela, P. Jääskelä, and T. Kärkkäinen, "Course Satisfaction in Engineering Education Through the Lens of Student Agency Analytics," 2020 IEEE Frontiers in Education Conference (FIE), Uppsala, Sweden, 2020, pp. 1-9, DOI: https://doi.org/10.1109/FIE44824.2020.9274141.

Course Satisfaction in Engineering Education Through the Lens of Student Agency Analytics

Ville Heilala
Faculty of Information Technology
University of Jyväskylä
Jyväskylä, Finland
https://orcid.org/0000-0003-2068-2777

Mirka Saarela
Faculty of Information Technology
University of Jyväskylä
Jyväskylä, Finland
https://orcid.org/0000-0002-1559-154X

Päivikki Jääskelä
Finnish Institute for Educational Research
University of Jyväskylä
Jyväskylä, Finland
https://orcid.org/0000-0002-3794-0325

Tommi Kärkkäinen
Faculty of Information Technology
University of Jyväskylä
Jyväskylä, Finland
tommi.karkkainen@jyu.fi

***Abstract*—This Research Full Paper presents an examination of the relationships between course satisfaction and student agency resources in engineering education. Satisfaction experienced in learning is known to benefit the students in many ways. However, the varying significance of the different factors of course satisfaction is not entirely clear. We used a validated questionnaire instrument, exploratory statistics, and supervised machine learning to examine how the different factors of student agency affect course satisfaction among engineering students ($N = 293$). Teacher's support and trust for the teacher were identified as both important and critical factors concerning experienced course satisfaction. Participatory resources of agency and gender proved to be less important factors. The results provide convincing evidence about the possibility to identify the most important factors affecting course satisfaction.**

***Index Terms*—course satisfaction, student agency, exploratory statistics, supervised machine learning**

I. Introduction

Satisfaction experienced in learning and educational situations is beneficial for the students [1], and the importance of emotions in enhancing learning and achievement is recognized in the field of learning sciences [2]. From the viewpoint of learning activities in higher education, cognitive, motivational, social, and emotional aspects are tightly intertwined. Therefore, both experienced course satisfaction and affective experience in the form of active agency [3] are essential constructs in understanding and supporting students in higher education.

Previous research identifies numerous aspects which highlight the essential role of positive emotions in learning [1], [2], [4]. Previous research has also identified that student-perceived overall satisfaction is linked with learning in several ways (e.g., [5]–[10]). Furthermore, studies stress the central role of agency in high-order learning processes (e.g., [11]). However, a comprehensive perspective is needed to fully grasp the meaning of experienced educational satisfaction in a variety of complex learning processes. Moreover, the link between the overall course satisfaction and student-experienced agency has not been studied previously.

Recently developed Student Agency (AUS) Scale [12] provides a possibility for a multidimensional examination of the association between course satisfaction and student-experienced agency. Based on the multidimensional view on the construct, the AUS includes several dimensions relating to purposeful learning (e.g., efficacy, opportunities for participation, and instructor's support), which in the previous studies (e.g., [13]) have been examined separately and found to be related to course satisfaction.

The purpose of this paper is to explore the relationship between course satisfaction and student-experienced agency in engineering education. Exploratory analysis and supervised machine learning are used to achieve this study purpose. In terms of the course satisfaction, we apply the measurement of customer satisfaction via the Net Promoter Score (NPS) [14] used in business. On the part of student agency, we utilize student agency analytics developed in the previous stage of research [15]. Finally, we use exploratory analysis and different classifiers to assess how the student agency factors affect experienced course satisfaction. The results contribute to the operationalization of course satisfaction and the broader understanding of its underlying factors.

II. Theoretical background

A. Student agency

Student agency is a central concept, for example, in the *OECD Learning Compass 2030* [16], [17], which "is an evolving learning framework that sets out an aspirational vision for the future of education" [16, p. 2]. In the framework, student agency relates to identity, sense of belonging, motivation, hope, self-efficacy, growth mindset, and a sense of purpose [18, p. 4]. In engineering education, students' exertions of agency affect, for example, to the decisions to continue their degrees [19], [20]. Also, pressure towards student agency in higher education has turned the focus on student capabilities

[21]. The central role of student agency in the contemporary discourse about future education invites researchers to explore its meaning and relationships with other concepts.

In this paper, we utilize a multidimensional conceptualization of student agency in a higher education context (Fig. 1), and the validated Agency of University Students (AUS) Scale [12], [22] to examine students' agency experiences at the course level. In AUS Scale, student agency refers to a *student's experience of having access to and use of personal, relational, and participatory resources for purposeful, intentional, and meaningful action and learning*. Personal resources encompass a student's competence and self-efficacy beliefs, with the former referring, for example, to a student's sense of understanding the course contents and the latter to a broader self-confidence as a learner in the course. Relational resources include the aspects related to teacher – students (power) relations in the course, which manifests themselves as a student's sense of getting support from the teacher, of being treated equally, and of trust to the teacher in the course. Participatory resources cover the factors that maintain both personally meaningful and intentional and interactive action in the course. In line with Su [3], agency is seen as intrinsically intertwined with learning as an affective experience, cognition, and action in the courses and learning relations.

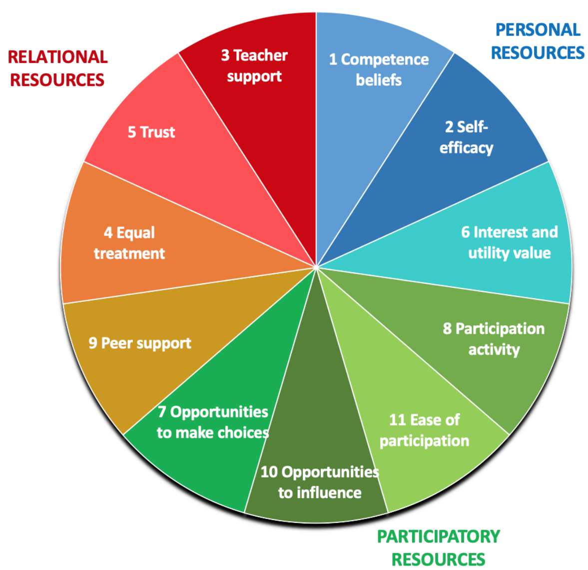


Fig. 1. The AUS model [12]

B. Course satisfaction in engineering education

In general, satisfaction towards a product or a service is a multifaceted phenomenon [23]. Student satisfaction is a continually shaped student's subjective evaluation of different experiences and outcomes relating to education [1]. Basically student satisfaction is ensued if the perceived performance meets or exceeds the student's expectations, and dissatisfaction will emerge in the opposite case [1]. Our study examines engineering students' satisfaction relating to a particular course they have attended. Thus, the concept of course satisfaction refers to the student satisfaction relating to a specific course.

Course satisfaction is an important component in successful learning. Studies suggest, for example, that course satisfaction has an effect on general academic satisfaction [5], [6]. In particular, it is linked to retention and academic locus of control [7], [8], as well as attrition [9] and approaches to learning [10]. Furthermore, satisfaction in academic life relates to the experienced life satisfaction in general [6].

In addition to learning-related effects, Browne et al. [24] found a moderate positive correlation between global satisfaction towards college and a willingness to recommend the college. Also, Mustafa et al. [25] found out that student satisfaction has a positive effect on students' willingness to promote the educational institution. Thus, it might be possible to categorize the experienced course satisfaction similarly to customer satisfaction (the willingness to promote a product or a service, c.f., [14]).

Several issues have been identified to contribute to the students' course satisfaction. Paechter et al. [13] found out that students' achievement goals, the instructor's support, expertise along with students' opportunities for self-regulated and collaborative learning, motivation, and the clarity of the course structure all contribute to course satisfaction and learning achievements. They also argue that competence beliefs are essential factors in course satisfaction. Komarraju et al. [26] found out that career self-efficacy, i.e., an individual's self-efficacy beliefs that one can complete the tasks and purposefully construct a career path, was one of the explaining factors of course and major subject satisfaction. McFarland and Hamilton [27] suggest, for example, that by enforcing appropriate course prerequisites, one could expect to find higher course satisfaction both in traditional and online instruction. In an online course context, Bolliger [28] reported that the course satisfaction of students is influenced by instructor variables, technical issues, and interactivity. Furthermore, Richardson et al. [29] report that quality of teaching, support for studying, and fair and clear course assessment correlated positively with the overall course satisfaction. Their results showed that teaching and support had the highest correlation with overall satisfaction.

Lynch et al. [30] identified eleven significant factors influencing engineering students' satisfaction that broadly concern interaction with the instructor, providing real-world connections, delivering meaningful content, advancing problem-solving and group work, and promoting student motivation. Similarly, González-Rogado et al. [31] found out that the usefulness of course content for future professionals, the methodology employed in the educational process, and the teamwork carried out throughout the course were related to course satisfaction in engineering education. The instructional design might also affect student satisfaction. For example, Kerr et al. [32] concluded that studies examining flipped learning in engineering education reported positive gains in student satisfaction.

Also, aspects that might have a biasing effect on assessing course satisfaction have been identified. For example, male students tend to give lower ratings as compared to females when evaluating teaching [33]. However, Leao et al. [34] found no difference in general course satisfaction between genders in the context of engineering education. Student-experienced course satisfaction can be influenced by the students' potential bias against female teachers, and teachers with non-native speaking backgrounds and instructors among students with high expectations of the course sometimes receive more favorable evaluations [35].

In general, student satisfaction seems to be a complex and multifaceted phenomenon with many influencing factors. Interestingly, many of the different elements affecting course satisfaction are also related to the dimensions of the student agency construct, for example, teacher support, trust for the teacher, and competence beliefs (see Fig. 1) [12]. Thus, the novel approach chosen for this study is to examine the relationship between student agency and course satisfaction.

III. Research questions

The present study aims to i) examine a way to quantify and categorize student-assessed course satisfaction and ii) explore the relationship of course satisfaction and student agency. Thus, more specifically we set the following research questions:

RQ1: Can course satisfaction be categorized similarly to the willingness to promote a product or a service?

RQ2: Is there differences in the dimensions of student agency between different course satisfaction experiences?

RQ3: What are the most important student agency factors contributing to experienced course satisfaction?

To answer the RQ1, we compare statistically the similarity and association between net promoter score (c.f., [14]) and course satisfaction (Section V-A). To answer the RQ2, we assess the differences in each dimension of student agency between the course satisfaction groups (Section V-B). To answer the RQ3, we use supervised machine learning methods to find out the most important features contributing to the classification accuracy (Section V-C).

IV. Data and methods

A. Research data

The research sample consists of questionnaire responses of engineering students ($N = 293$) in a higher education institution (ISCED Level 6) studying courses about basic IT skills and mathematics for engineers. The courses belonged to the engineering students' curricula and they were common to all engineering students regardless of their line of study. The courses consisted of lecturing and small group teaching. The data were collected using an online questionnaire, which was administered at the end of the courses before the final course grades were announced. Respondents' ages ranged from 18 to 52 years ($Mdn = 21, M = 23.6, SD = 5.8$) with 23% identified as female. The majority of the students were at the beginning of their studies, and their total amount of completed study credits (ETCS) ranged from 0 to 260 ($Mdn = 0, M = 21.0, SD = 46.4$).

B. Measures

1) Course satisfaction: The traditional approach of quantifying student satisfaction is to measure the student's overall satisfaction with an aggregate single-item measure [1]. Satisfaction scores are prone to a ceiling effect [36], which means that the respondents' scores cluster toward the high end of the scale [37]. Kleiss et al. [38] tested several scales for assessing medical patient satisfaction and found that an 11-point ordinal scale (range 0–10) approached the most normal distribution. As the use of a 0–10 scale is quite common, they claim the scale is also more familiar to respondents. However, it raises the question of what values or thresholds should be used to depict satisfaction and dissatisfaction. As there is no agreement of the optimal way of measuring course satisfaction [38], we chose 0–10 Likert-type item as a starting point. Furthermore, we adapt the categorization idea used in the Net Promoter Score™ (NPS) [14].

Net Promoter Score measures the customers' willingness to recommend a product or a service [14]. In the calculation of NPS, respondents are divided into three categories based on their answers on a 0-10 scale. Respondents answering 9 or 10 are categorized as promoters of a service or a product. Respondents answering 7 or 8 are passives, and respondents answering 6 or less are called as detractors. The actual net promoter score is the difference between the percentages of promoters and detractors. NPS has received wide criticism and, contrary to the original claims [14], studies suggest that it does not have a significant effect on business performance [39]. However, the approach allows us to examine the proposed categorization of the students into three satisfaction categories. Adapting the idea behind NPS, students were categorized as satisfied, neutrals, and dissatisfied. The course satisfaction category was then used in the exploratory analysis, and as the predicted variable in supervised learning.

We measured students' course satisfaction with two questions. First, respondents were asked to evaluate their overall satisfaction relating to the course ("How satisfied you were with the course? "). To compare the course satisfaction scale with the NPS, later in the questionnaire, we also asked how probably the respondent would recommend the course to their fellow students ("How probably would you recommend this course to a fellow student? "). Same as originally in the NPS [14], both items were measured using a Likert-type item in a 0 to 10 point scale. Value of 0 indicated an answer "no at all" and value 10 "very satisfied / very likely". To assess the applicability of the idea of NPS in assessing course satisfaction, we make a comparative analysis of the two scales. In a similar comparison, Laitinen [40] examined the library patrons' satisfaction and willingness to recommend a library service. He found that the two evaluation metrics converge; however, there was a statistically significant difference at the highest grades.

2) Student agency: Student agency was measured using the AUS scale, which consists of items measuring student agency at the course level. AUS scale captures three main domains of agency resources, and their respective eleven dimensions (Fig. 1): A. Personal resources (1. Competence beliefs, 2. Self-efficacy), B. Relational resources (3. Equal treatment, 4. Teacher support, 5. Trust), and 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). A student's responses to the AUS questionnaire will give us knowledge on the extent the student perceives to have personal resources, affordance for supportive relations, as well as opportunities for active participation and influencing the course. Also, students answered two open-ended questions about supportive and restrictive aspects experienced during the course.

C. Data analysis

The first step in analyzing student agency was to calculate the values of agency dimensions as factor values for each respondent [15]. We reverse-scored the inverted items using linear scaling and calculated the agency factors using the factor pattern matrix of the AUS factor model. The calculated agency factors were scaled to the original range $[1, 5]$ of the Likert scale. A Spearman's rank-order correlation was used to assess the similarity and association between students' answers to the course promoting scale and course satisfaction scale. Kruskal-Wallis H was used to compare the difference of medians within agency dimensions and how strongly the medians are separating the different satisfaction groups [41]–[43]. Mann-Whitney U was used to compare the pairwise difference of medians between each course satisfaction groups in all student agency dimensions [44]. In the exploratory analysis of the course satisfaction and student agency, the statistical analyses and visualizations of the associations between student agency and course satisfaction were executed in *R* 3.6.1 [45] using *ggplot2* and *ggpubr* packages.

We used supervised machine learning to analyze the feature importances, in other words, the important dimensions of student agency affecting course satisfaction. The supervised analysis (Section V-C) was performed using *Python* 3.7.1 and different classifiers implemented in the *scikit-learn* package. We trained one linear (logistic regression), and three non-linear (support vector machines with Gaussian kernel, random forest, and gradient boosting) classifiers to predict the binary course satisfaction score (i.e., the satisfied category). The support vector machine classifier was trained using *scikit-learn*'s *SVC* with a parameter search over the regularization parameter *C* and the width of the Gaussian kernel *gamma*. The *LogisticRegression* classifier was trained with a parameter search over *C* and the penalty *l*. The *RandomForestClassifier* was trained with a search over the number of maximum features and the *GradientBoostingClassifier* with a search over the learning rate and the maximum depth.

V. RESULTS

A. Comparing course satisfaction and willingness to recommend the course

Adapting the NPS method [14], we divided the respondents to three groups based on their course satisfaction score (Fig. 3). Students scoring their course satisfaction as 9 or 10 were classified as satisfied (c.f., promoters in NPS), students scoring 7 or 8 were classified as neutrals (c.f., passives), and students scoring their course satisfaction as 6 or below were classified as dissatisfied (c.f., detractors). The distribution of the satisfaction scores is similar to other studies assessing experienced satisfaction towards a service or a product (e.g., [40]). The majority of the students were classified as neutrals ($n = 136; 46\%$). The second-largest group was the satisfied students ($n = 88; 30\%$), and a quarter of the students were classified as dissatisfied ($n = 69; 24\%$) based on their answers to the course satisfaction item.

To validate the aforementioned approach, we examined the relationship between the course satisfaction item ("How satisfied you were with the course? ") and the course recommendation item ("How probably would you recommend this course to a fellow student? "). A total of 47% of the respondents scored equal scores on both scales, and the mean difference between scores was 0.81 in the 0–10 scale. The association between items was measured using the Spearman's rank correlation coefficient ($\rho = 0.71$), which indicated a strong positive monotonic association (Fig. 2). It is worth noting that the items still measure different constructs. However, students scored the same or close to the same score in both items. While there was deviating individual responses, the results indicate that both items behaved similarly.

We examined the relationship between the reported course satisfaction and gender, because the previous literature [33] has identified that gender might affect the reported educational satisfaction. We did not find any statistically significant difference between the reported course satisfaction and gender ($Mdn = 8$ for both). However, the result has to be interpreted carefully because data contained only 23% of the responses by female students.

To further validate the categorization of the groups, we examined the open-ended answers of the student agency questionnaire about the experienced support and restrictions in the course. Table I presents the count data of the occurrences of the mentioned support and restrictions of learning in each satisfaction group. If a student wrote an answer in the questionnaire to the question about supporting aspects in the course, it was counted as one occurrence of support. Similarly, if a student wrote an answer to the question about experienced restrictions in the course, it was counted as a restriction of learning. Responses containing only statements like "nothing" and "don't know" were removed. The results showed that 50% of the satisfied students reported restrictions, which was less compared to the students in other groups. Also, students in the satisfied group reported more likely only support in the course (25% of the satisfied students) comparing to the other

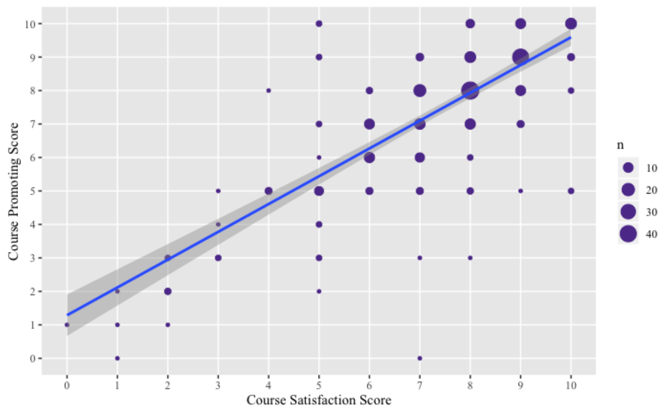


Fig. 2. Ordinal scatterplot and a regression line depicting the association of variables measuring students' assessed course satisfaction and willingness to recommend the course to a fellow student.

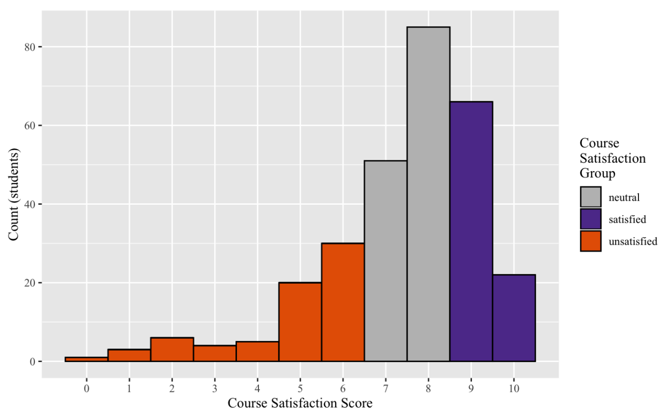


Fig. 3. Course satisfaction scores and division to neutral ($n = 136; 46\%$), satisfied ($n = 88; 30\%$), and dissatisfied ($n = 69; 24\%$) groups.

groups. Students classified as dissatisfied reported more often restrictions (75% of the dissatisfied students), and they were more likely to report only restrictive aspects (19%) comparing to students in other satisfaction groups. In general, satisfied students experienced more support, and dissatisfied students reported more likely restrictions. The result provides support for the categorization of the course satisfaction scale. However, further research is needed to get more insight into the student experiences and validation of the cut-off values.

TABLE I
OPEN-ENDED ANSWERS OF STUDENTS IN DIFFERENT SATISFACTION GROUPS

Answers reporting...	Dissatisfied	Neutral	Satisfied
support	58%	71%	68%
restrictions**	75%	60%	50%
only support***	1%	16%	25%
only restrictions*	19%	6%	7%
no answer	23%	24%	25%

Fisher's exact test for count data: $^{***}p < .001;^{**}p \approx .005;^{*}p \approx .011$

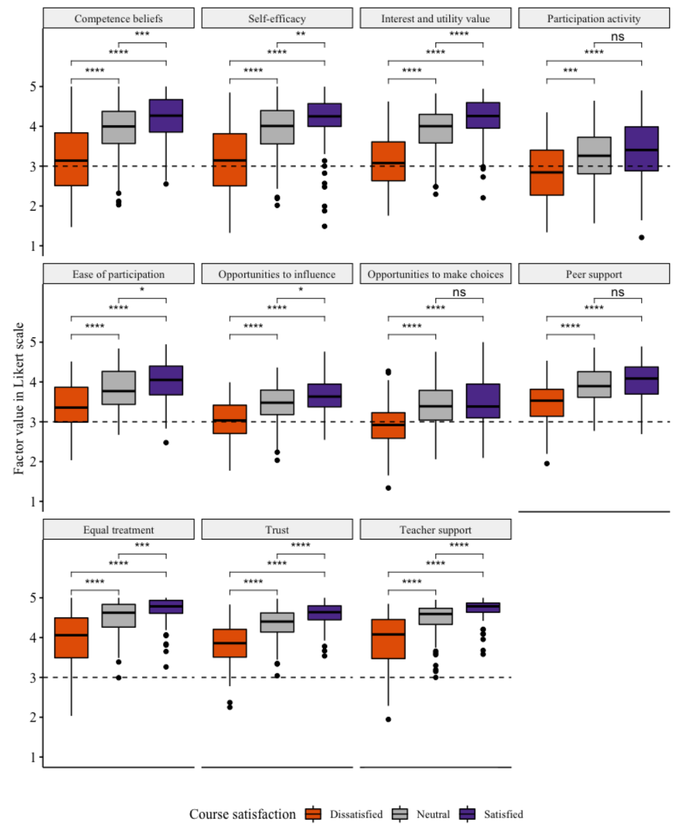


Fig. 4. Student agency in each course satisfaction category and pairwise statistical significance using Mann-Whitney U statistics.

B. Analysis of course satisfaction and student agency

We used the Kruskal-Wallis H test to determine if there were differences in the scores of the student agency dimensions between the three satisfaction category groups of students: the dissatisfied ($n = 69$), neutral ($n = 136$), and satisfied ($n = 88$). Distributions of student agency scores were similar for all course satisfaction groups, except in the dimensions of equal treatment, trust, and teacher support, as illustrated in the boxplots in Fig. 4. The medians of student agency scores were statistically significantly different between the course satisfaction groups in every dimension of student agency ($p < .001$). In general, students assessed the experienced resources of agency as lower in the lower satisfaction categories.

Kruskal-Wallis H can be used to evaluate the importance of different student agency dimensions in separating the course satisfaction categories [42], [43]. A high test statistic H indicates a strong separation of the medians, which in turn implies that the particular feature is more relevant in separating the different groups compared to features having lower test statistic score [42]. The order of the student agency dimensions concerning how strongly they separate the three satisfaction categories according to the H statistics is the following: teacher support ($H = 93$), trust for the teacher ($H = 90$), interest and utility value ($H = 82$), competence beliefs ($H = 63$), equal treatment ($H = 62$), self-efficacy

($H = 61$), opportunities to influence ($H = 52$), peer support ($H = 40$), ease of participation ($H = 39$), opportunities to make choices ($H = 37$), and participation activity ($H = 21$).

Subsequently, pairwise comparisons were performed using Mann-Whitney U statistics. The test can detect differences in the spread and shape of the distribution in addition to differences in medians, which can all be important features of the data [44]. Based on this post hoc test, the differences were statistically significant, as presented in Fig. 4, in all other cases except between satisfied and neutral categories in the dimension of participation activity, opportunities to make choices, and peer support.

In cases of equal treatment, trust for the teacher, and teacher support, the statistically significant differences might occur because of differences in the shape of the distributions. For example, satisfied categories in the dimensions mentioned above have small spread compared to other categories making them critical concerning course satisfaction: a slight decrease in the critical dimensions increases the chance of belonging to the lower satisfaction categories. In other dimensions of student agency where the difference was statistically significant, the difference can be characterized as a shift in location, which in turn can be described as a difference in medians [44]. In addition, we examined the relationship between gender and student agency dimensions and did not find any statistically significant differences.

C. Important factors contributing to course satisfaction

Finally, we predicted whether the students were satisfied with the course (i.e., belonged to the satisfied category). Traditional educational studies mostly use simple linear classifiers for predicting a categorical variable. However, many studies have shown that these are often outperformed by nonlinear classifiers (see, e.g., [46]–[48]). To test which method works best for our data, we employed four popular classifiers: The traditional linear logistic regression (LR) and three nonlinear classifiers, namely support vector machine (SVM) with Gaussian kernel, gradient boosting (GB), and random forest (RF), in comparison.

As input features, we utilized the eleven agency factors and gender. Gender was added as a control variable based on the previous literature (e.g., [33]). For all classifiers, we divided our data into training (80%) and test (20%) using a stratified split according to satisfaction category. We then used a five-fold cross-validation grid-search over the training data to determine the best parameters. As some classifiers are sensitive to unscaled data, we utilized min-max scaling to normalize the data into the range $[0, 1]$ (determining the scaling coefficients from the training set and applying them to the test set).

Similarly as in [49], we employed a pipeline on the training data chaining the different steps (preprocessing with and without scaling of data) and the five-fold grid-search over the different parameter settings together. The best combination (best preprocessing and best parameter settings) returned by the pipeline was then used to predict the dependent variable of the test set that was untouched the entire time during model training and metaparameter selection.

Table II summarizes for all classifiers the best preprocessing and best parameters from the cross-validated grid-search, and the performance on the test set. Figure 5 shows the receiver operating characteristic (ROC) curves [50] for the test sets for all classifiers with their best parameters. In Table II, the performance is summarized as area under the ROC curve (AUC) [51]. The AUC measures the area underneath the ROC curve and is equal to the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one. When comparing different classifiers on average, a higher AUC value indicates better classification performance. As can be seen from the table and figure, random forest provided the best performance.

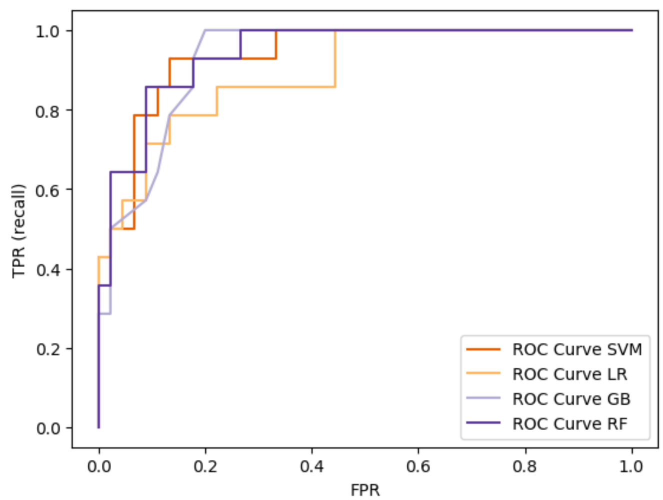


Fig. 5. ROC curves of the test set for the four classifiers predicting the satisfied category.

Figure 6 shows the feature importances for the best random forest model. Similar to results from Kruskall-Wallis H, the interest, trust, and teacher support student agency dimensions contributed the most, while the gender seemed less important when predicting the course satisfaction. Moreover, the dimensions of participatory resources in student agency were less important to predict the course satisfaction compared to the other resources of agency. Gender was the least important predictor of all features.

We also tried predicting the multinomial output (i.e., all three categories: dissatisfied, neutral, and satisfied) using the same input variables and classifiers as described above. However, as can be expected, the prediction performance for these multinominal classifiers was worse than for the binary ones. The random forest model performed again the best with a micro-averaged F1 score of 0.678, while support vector machines, logistic regression, and gradient boosting achieved micro-averaged F1 scores of less than 0.6. The feature importances of the best multinomial random forest classifier were very similar to the ones from the binary random forest model. Only the feature importances of teacher support and trust for

TABLE II
THE BEST PREPROCESSING AND PARAMETERS ON THE TRAINING SET AND PERFORMANCE (AUC) FOR THE TEST SET.

Classifier	**Best preprocessing**	**Best parameters**	**AUC**
Support vector machines	Min-max scaling	$'C' : 10, 'gamma' : 0.001, 'kernel' : rbf$	0.938
Logistic regression	Min-max scaling	$'C' : 100, 'penalty' : l2$	0.894
Gradient boosting	None	$'learning\_rate' : 0.01, 'max\_depth' : 1$	0.929
Random forest	None	$'max\_features' : 2$	0.943

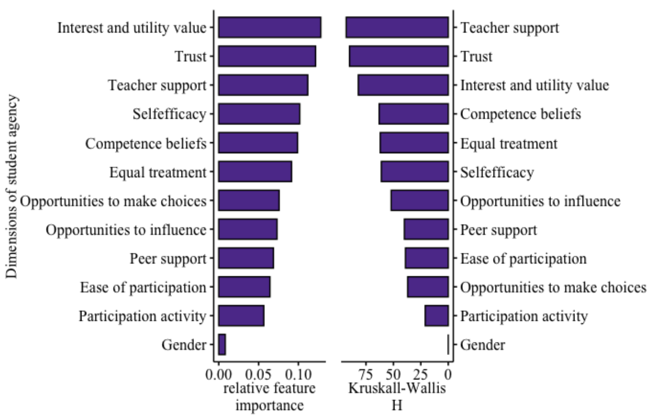


Fig. 6. Feature importances of the best random forest model predicting the satisfied category and the values of the Kruskall-Wallis H statistics in each dimensions of student agency depicting the separation of course satisfaction medians.

the teacher were reversed (i.e., teacher support was the second, and trust for the teacher was the third important variable).

VI. DISCUSSION

Positive experiences are fundamental in learning, and prior work has identified several issues having an impact on experienced satisfaction in learning situations. In engineering education, for example, Lynch et al. [30] found that interaction with the instructor, providing real-world connections, delivering meaningful content, advancing both problem-solving and group work, and promoting student motivation are among the factors influencing experienced course satisfaction. However, experienced satisfaction in learning is a complex phenomenon, and the significance of different factors affecting course satisfaction is not yet fully resolved. In this study, we examined how the various dimensions of student agency in engineering education contribute to student-reported course satisfaction.

With our first research question, we strived to find an adequate solution for measuring course satisfaction. Lacking the explicit guidance from the previous research literature for measuring course satisfaction, we devised an initial measuring scheme applied in this paper. Following Kleiss et al. [38], we opted to use a single Likert-type item on 0–10 scale. We then compared the answer scores between items measuring course satisfaction and willingness to recommend the course. Similarly to the previous studies [24], [25], we found out that the items had a positive correlation. Finally, we adapted the idea similar to NPS [14] to categorize the experienced course satisfaction into three categories (i.e., dissatisfied, neutral, satisfied) for further analysis. To answer the RQ1, we conclude that course satisfaction could be categorized similarly to customer satisfaction.

The second research question involved finding out the differences in the dimension of student agency between different satisfaction categories. Our findings indicate that dissatisfied students reported more often restrictive aspects, and satisfied students reported more often only supportive aspects in their learning. In general, the resources of student agency were experienced as lower in the lower satisfaction categories compared to more satisfied students in every dimension of student agency. To answer the RQ2, we conclude that there were significant differences in the dimensions of student agency between different course satisfaction categories. The results are consistent with a previous research, in which students with lower agency experiences reported a variety of restrictive aspects in their learning [52]. Contrary to the previous research suggesting that male students tend to give lower evaluations in an educational context [33] and similar to Leao et al. [34], we did not find any difference in satisfaction scores between the genders in the research sample.

The last research question aimed to find out what are the most important factors of student agency contributing to course satisfaction. By using training data, untouched test data, grid search, and different classifiers, we found out that random forest classifier was able to predict the satisfied category with a high AUC score. In general, all non-linear classifiers performed better than the linear logistic regression, which might be an indication of the complex and non-linear nature of course satisfaction. To answer the RQ3, we conclude that the three most important student agency dimensions of the model were interest and utility value, trust for the teacher, and teacher support. In terms of methodological triangulation [53], the same dimensions were identified in a slightly different order when using Kruskal-Wallis H statistics to quantify the separation between all three satisfaction categories. The results comply with the previous research that found teacher support to be one of the most influencing factors relating to satisfaction towards school and studying [6], [13]. The dimensions relating to participation (i.e., ease of participation, participation activity) were examples of the less important factors. Again, gender did not prove to have a noteworthy predictive power.

Most notably, our study highlights the differences in relevance among various factors and experiences affecting the perceived satisfaction in learning. Some factors (e.g., interest and utility value, self-efficacy) can be considered as *important*. However, some factors can be regarded as both important and *critical* (e.g., teacher support, trust for the teacher), which means that a small decrease in the critical factor can more likely cause a decline in experienced course satisfaction comparing to a non-critical factor. In the case of *less important* or unimportant factors, their predictive or separative power concerning course satisfaction is not significant (e.g., participation activity, gender).

A. Limitations and future work

Although student agency used in this study is a multidimensional construct, it is only one of the many possible constructs that could be used in examining the factors affecting course satisfaction. Future studies should aim to explore and analyze the important and especially the critical factors affecting course satisfaction in greater detail. The explorations should utilize a variety of different constructs from other points of view than student agency (e.g., approaches to learning [10], model of domain learning [54]).

Also, measuring and quantifying a complex phenomenon like course satisfaction is challenging. Our initial scheme for the measurement is a starting point and should be more thoroughly validated both quantitatively and qualitatively. Notably, the thresholds of the different satisfaction categories need to be carefully examined in future studies. Multinomial linear regression using the whole satisfaction scale would be one option in future research. Here the use of multinomial regression would have required more observations in the lower part of the satisfaction scale. Students' open-ended answers about support and restrictions in learning seem to yield relevant information about student agency, which should be investigated more in-depth using, for example, natural language processing techniques.

VII. Conclusion

In summary, we have examined the factors affecting course satisfaction in the context of engineering education by utilizing student agency analytics, exploratory statistics, and supervised machine learning. This study contributed to the measurement of course satisfaction and to the analysis of its underlying factors. The results provide evidence about the important, critical, and less important dimensions of student agency affecting experienced course satisfaction. From the dimensions of student agency, teacher support and trust for the teacher turned out to be both important and critical features, interest and utility value was important, and the dimensions relating to participation were less important with respect to the course satisfaction in engineering education. We expect the results to broaden the understanding of student agency and course satisfaction, and provide both educators and educational institutions capabilities to promote effective aspects of learning.

Practical implications of our study relate to the possibility of taking carefully into consideration especially the most important and critical factors affecting course satisfaction. Experienced satisfaction influences widely and positively students' learning [5], [6], and this might be worth to take into account throughout the educational system. A practical future application serving teachers could be, for example, a learning analytics system providing predictive information about course satisfaction in advance. Educational institutions also benefit from concentrating on the important aspects of course satisfaction as student satisfaction has a positive effect on students' willingness to promote their *alma mater* [25]. Finally, because of its relation to learning achievements [13], the course satisfaction is an important aspect in actionable learning analytics fulfilling the promises of "understanding and optimising learning and the environments in which it occurs" [55]. Finally, we hypothesize that if course satisfaction could be used as a proxy for learning outcomes and achievements, it might prove to be a valuable construct in various applications of learning analytics.

References

[1] K. M. Elliott and D. Shin, "Student satisfaction: An alternative approach to assessing this important concept," *Journal of Higher Education Policy and Management*, vol. 24, no. 2, pp. 197–209, 2002.

[2] R. Pekrun and L. Linnenbrink-Garcia, "Academic emotions and student engagement," in *Handbook of research on student engagement*, S. L. Christenson, A. L. Reschly, and C. Wylie, Eds. New York: Springer-Verlag, 2012, pp. 259–282.

[3] Y.-H. Su, "The constitution of agency in developing lifelong learning ability: The 'being'mode," *Higher Education*, vol. 62, no. 4, pp. 399–412, 2011.

[4] C. Kim and R. Pekrun, "Emotions and motivation in learning and performance," in *Handbook of Research on Educational Communications and Technology*, J. M. Spector, M. D. Merrill, J. Elen, and M. J. Bishop, Eds. New York, NY: Springer-Verlag New York, 2014, pp. 65–75.

[5] N. D. Aitken, "College student performance, satisfaction and retention: Specification and estimation of a structural model," *Journal of Higher Education*, vol. 53, no. 1, pp. 32–50, 1982.

[6] A. G. Danielsen, O. Samdal, J. Hetland, and B. Wold, "School-Related social support and students' perceived life satisfaction," *Journal of Educational Research*, vol. 102, no. 4, pp. 303–320, 2009.

[7] Y. Levy, "Comparing dropouts and persistence in e-learning courses," *Computers and Education*, vol. 48, no. 2, pp. 185–204, 2007.

[8] Y. Lee and J. Choi, "A structural equation model of predictors of online learning retention," *The Internet and Higher Education*, vol. 16, pp. 36–42, Jan. 2013.

[9] J. E. Edwards and L. K. Waters, "Involvement, ability, performance, and satisfaction as predictors of college attrition," *Educational and Psychological Measurement*, vol. 42, no. 4, pp. 1149–1152, 1982.

[10] M. Baeten, E. Kyndt, K. Struyven, and F. Dochy, "Using student-centred learning environments to stimulate deep approaches to learning: Factors encouraging or discouraging their effectiveness," *Educational Research Review*, vol. 5, no. 3, pp. 243–260, 2010.

[11] C. I. Damşa, P. A. Kirschner, J. E. B. Andriessen, G. Erkens, and P. H. M. Sins, "Shared epistemic agency: An empirical study of an emergent construct," *Journal of the Learning Sciences*, vol. 19, no. 2, pp. 143–186, 2010.

[12] P. Jääskelä, A.-M. Poikkeus, K. Vasalampi, U. M. Valleala, and H. Rasku-Puttonen, "Assessing agency of university students: validation of the aus scale," *Studies in Higher Education*, vol. 42, no. 11, pp. 2061–2079, 2017.

[13] M. Paechter, B. Maier, and D. Macher, "Students' expectations of, and experiences in e-learning: Their relation to learning achievements and course satisfaction," *Computers and Education*, vol. 54, no. 1, pp. 222–229, 2010.

[14] F. F. Reichheld, "The one number you need to grow," *Harvard Business Review*, vol. 81, no. 12, pp. 46–54, 124, 2003.

[15] P. Jääskelä, V. Heilala, T. Kärkkäinen, and P. Häkkinen, "Student agency analytics: learning analytics as a tool for analysing student agency in higher education," *Behaviour and Information Technology*, 2020.

[16] OECD, "Concept note: OECD learning compass 2030," 2019, accessed 2020-06-05. [Online]. Available: https://www.oecd.org/education/2030-project/teaching-and-learning/learning/learning-compass-2030/OECD_Learning_Compass_2030_concept_note.pdf

[17] J. Dunn, K. Salmela-Aro, V. Talreja, T. Abiko, N. Steinemann, C. Leadbeater, T. Bentley, A. Grayling, D. Kunin, N. Walt, A. Toutant, R. Allen, H. Haste, and P. Bishop, "Education 2030 - conceptual learning framework: Background papers," 2017, accessed 2020-06-05. [Online]. Available: https://www.oecd.org/education/2030-project/contact/Conceptual_learning_framework_Conceptual_papers.pdf

[18] OECD, "Conceptual learning framework: Student agency for 2030," 2019, accessed 2020-06-05. [Online]. Available: http://www.oecd.org/education/2030-project/teaching-and-learning/learning/student-agency/Student_Agency_for_2030_concept_note.pdf

[19] N. Ahmed, B. Kloot, and B. I. Collier-Reed, "Why students leave engineering and built environment programmes when they are academically eligible to continue," *European Journal of Engineering Education*, vol. 40, no. 2, pp. 128–144, Mar. 2015.

[20] M. M. Camacho, S. M. Lord, C. E. Brawner, and M. W. Ohland, "Climate in undergraduate engineering education from 1995 to 2009," in *2010 IEEE Frontiers in Education Conference (FIE)*, 2010, pp. T2H–1–T2H–6.

[21] M. Blackie, K. le Roux, and S. McKenna, "Possible futures for science and engineering education," *Higher Education*, vol. 71, no. 6, pp. 755–766, 2016.

[22] P. Jääskelä, K. Vasalampi, P. Häkkinen, A.-M. Poikkeus, and H. Rasku-Puttonen, "Students' agency experiences and perceived pedagogical quality in university courses," Paper presented at ECER 2017, 22 – 25 August, Copenhagen, Denmark, 2017.

[23] V. Zeithaml, L. Berry, and A. Parasuraman, "The nature and determination of customer expectation of service," *Journal of Marketing Sciene*, vol. 21, no. 1, pp. 1–12, 1993.

[24] B. A. Browne, D. O. Kaldenberg, W. G. Browne, and D. J. Brown, "Student as customer: Factors affecting satisfaction and assessments of institutional quality," *Journal of Marketing for Higher Education*, vol. 8, no. 3, pp. 1–14, 1998.

[25] Z. Mustafa, N. Basri, N. Z. Abidin, N. R. M. Suradi, Z. M. Ali, F. A. A. Shahabudin, and M. R. Ab Hamid, "Modeling of engineering student satisfaction," *Journal of Mathematics and Statistics*, vol. 8, no. 1, pp. 64–71, 2011.

[26] M. Komarraju, J. Swanson, and D. Nadler, "Increased career Self-Efficacy predicts college students' motivation, and course and major satisfaction," *Journal of Career Assessment*, vol. 22, no. 3, pp. 420–432, 2014.

[27] D. McFarland and D. Hamilton, "Factors affecting student performance and satisfaction: Online versus traditional course delivery," *Journal of Computer Information Systems*, vol. 46, no. 2, pp. 25–32, 2005.

[28] D. U. Bolliger, "Key factors for determining student satisfaction in online courses," *International Journal on E-learning*, vol. 3, no. 1, pp. 61–67, 2004.

[29] J. T. E. Richardson, J. B. Slater, and J. Wilson, "The national student survey: development, findings and implications," *Studies in Higher Education*, vol. 32, no. 5, pp. 557–580, 2007.

[30] P. Lynch, S. Sangelkar, G. Demeo, and B. Morkos, "The ICDM methodology improving undergraduate engineering student motivation, satisfaction, and performance," in *2017 IEEE Frontiers in Education Conference (FIE)*, 2017, pp. 1–7.

[31] A.-B. Gonzalez-Rogado, F. J. García-Peñalvo, M. Borham, S. Olmos-Migueláñez, and M. J. Rodríguez-Conde, "Key factors for determining student satisfaction in engineering: A regression study," *International Journal of Engineering Education*, vol. 30, no. 3, pp. 576–584, 2014.

[32] B. Kerr, "The flipped classroom in engineering education: A survey of the research," in *2015 International Conference on Interactive Collaborative Learning (ICL)*, 2015, pp. 815–818.

[33] C. N. Tatro, "Gender effects on student evaluations of faculty," *Journal of Research and Development in Education*, vol. 28, no. 3, pp. 169–173, 1995.

[34] C. P. Leão, F. Soares, A. Guedes, M. T. Sena Esteves, G. Alves, I. M. Brás Pereira, R. Hausmann, and C. A. Petry, "Is students' satisfaction in electrical engineering courses influenced by gender?" in *2017 International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, 2017, pp. 1362–1369.

[35] Y. Fan, L. J. Shepherd, E. Slavich, D. Waters, M. Stone, R. Abel, and E. L. Johnston, "Gender and cultural bias in student evaluations: Why representation matters," *PLoS One*, vol. 14, no. 2, p. e0209749, 2019.

[36] C. K. Ross, C. A. Steward, and J. M. Sinacore, "A comparative study of seven measures of patient satisfaction," *Med. Care*, vol. 33, no. 4, pp. 392–406, 1995.

[37] O. Garin, "Ceiling effect," in *Encyclopedia of Quality of Life and Well-Being Research*, A. Michalos, Ed. Dordrecht: Springer Netherlands, 2014.

[38] I. I. M. Kleiss, J. T. P. Kortlever, P. Karyampudi, D. Ring, L. E. Brown, L. M. Reichel, M. D. Driscoll, and G. A. Vagner, "A comparison of 4 Single-Question measures of patient satisfaction," *Journal of Clinical Outcomes Management*, vol. 27, no. 1, pp. 41–48, 2020.

[39] N. A. Morgan and L. L. Rego, "The value of different customer satisfaction and loyalty metrics in predicting business performance," *Marketing Science*, vol. 25, no. 5, pp. 426–439, 2006.

[40] M. A. Laitinen, "Net promoter score as indicator of library customers' perception," *Journal of Library Administration*, vol. 58, no. 4, pp. 394–406, 2018.

[41] A. Vargha and H. D. Delaney, "The Kruskal-Wallis test and stochastic homogeneity," *Journal of Educational and Behavioral Statistics*, vol. 23, no. 2, pp. 170–192, 1998.

[42] A. Cord, C. Ambroise, and J.-P. Cocquerez, "Feature selection in robust clustering based on laplace mixture," *Pattern Recognition Letters*, vol. 27, no. 6, pp. 627–635, 2006.

[43] M. Saarela, J. Hämäläinen, and T. Kärkkäinen, "Feature ranking of large, robust, and weighted clustering result," in *Advances in Knowledge Discovery and Data Mining*, ser. Lecture Notes in Computer Science, J. Kim, K. Shim, L. Cao, J. G. Lee, X. Lin, and Y. S. Moon, Eds., vol. 10234. Cham: Springer, 2017, pp. 96–109.

[44] A. Hart, "Mann-Whitney test is not just a test of medians: differences in spread can be important," *BMJ*, vol. 323, no. 7309, pp. 391–393, 2001.

[45] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2019. [Online]. Available: https://www.R-project.org/

[46] D. C. Gibson and D. Ifenthaler, "Preparing the next generation of education researchers for big data in higher education," in *Big Data and Learning Analytics in Higher Education: Current Theory and Practice*, B. Kei Daniel, Ed. Cham: Springer, 2017, pp. 29–42.

[47] R. Al-Shabandar, A. Hussain, A. Laws, R. Keight, J. Lunn, and N. Radi, "Machine learning approaches to predict learning outcomes in massive open online courses," in *2017 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2017, pp. 713–720.

[48] M. Saarela and T. Kärkkäinen, "Can we automate expert-based journal rankings? analysis of the finnish publication indicator," *Journal of Informetrics*, vol. 14, no. 2, p. 101008, 2020.

[49] M. Saarela, O.-P. Ryynänen, and S. Äyrämö, "Predicting hospital associated disability from imbalanced data using supervised learning," *Artificial Intelligence in Medicine*, vol. 95, pp. 88 – 95, 2019.

[50] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861 – 874, 2006.

[51] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recognition*, vol. 30, no. 7, pp. 1145–1159, 1997.

[52] V. Heilala, P. Jääskelä, T. Kärkkäinen, and M. Saarela, "Understanding the study experiences of students in low agency profile: Towards a smart education approach," in *Advances in Smart Technologies Applications and Case Studies*, A. El Moussati, K. Kpalma, M. Ghaouth Belkasmi, M. Saber, and S. Guégan, Eds. New York, NY: Springer, in press.

[53] N. K. Denzin, *The Research Act: A Theoretical Introduction to Sociological Methods*, 2nd ed. New York: McGraw-Hill, 1978.

[54] P. A. Alexander, "The development of expertise: The journey from acclimation to proficiency," *Educational Researcher*, vol. 32, no. 8, pp. 10–14, 2003.

[55] G. Conole, D. Gašević, P. Long, and G. Siemens, "Message from the LAK 2011 general & program chairs," in *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*. New York, NY: ACM, 2011.

PIV

EXPLAINABLE STUDENT AGENCY ANALYTICS

by

Saarela, M., Heilala, V., Jääskelä, P., Rantakaulio A., and Kärkkäinen, T 2021

IEEE Access

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received July 17, 2021, accepted September 26, 2021, date of publication September 29, 2021, date of current version October 13, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3116664

Explainable Student Agency Analytics

MIRKA SAARELA[1], VILLE HEILALA[1], PÄIVIKKI JÄÄSKELÄ[2], ANNE RANTAKAULIO[3], AND TOMMI KÄRKKÄINEN[1], (Senior Member, IEEE)

[1]Faculty of Information Technology, University of Jyväskylä, 40014 Jyväskylä, Finland
[2]Finnish Institute for Educational Research, University of Jyväskylä, 40014 Jyväskylä, Finland
[3]Department of Technology and Transport, Jyväskylä University of Applied Sciences, 40200 Jyväskylä, Finland

Corresponding author: Mirka Saarela (mirka.saarela@jyu.fi)

This work was supported by the Academy of Finland and related to the thematic research area Decision Analytics Utilizing Causal Models and Multiobjective Optimization (DEMO) (jyu.fi/demo), University of Jyväskylä, Finland, under Grant 311877.

ABSTRACT Several studies have shown that complex nonlinear learning analytics (LA) techniques outperform the traditional ones. However, the actual integration of these techniques in automatic LA systems remains rare because they are generally presumed to be opaque. At the same time, the current reviews on LA in higher education point out that LA should be more grounded to the learning science with actual linkage to teachers and pedagogical planning. In this study, we aim to address these two challenges. First, we discuss different techniques that open up the decision-making process of complex techniques and how they can be integrated in LA tools. More precisely, we present various global and local explainable techniques with an example of an automatic LA process that provides information about different resources that can support student agency in higher education institutes. Second, we exemplify these techniques and the LA process through recently collected student agency data in four courses of the same content taught by four different teachers. Altogether, we demonstrate how this process—which we call explainable student agency analytics—can contribute to teachers' pedagogical planning through the LA cycle.

INDEX TERMS Explainable artificial intelligence, decision making, higher education, student agency.

I. INTRODUCTION

The global COVID-19 and the related closures of educational institutions showed how significant it is for students to be able to rely on their own resources. In particular, to continue learning, the educational institutions' closures placed greater demands on students' autonomy and their capacity for independent learning, executive functioning, and self-monitoring [1]. It also showed that those students who lacked the resilience and engagement to learn on their own, in particular, were at risk of falling behind [1], [2]. In summary, COVID-19 and its consequences for students revealed the importance of being self-determined in learning and being able to adapt to situations involving rapid change.

Student agency equips students to manage such situations. It refers to students' holistic judgement of how they can affect and direct their learning in instructive settings, work effectively, and utilize the assets that are accessible in the learning environment [3], [4]. The importance of agency in education has been emphasized by policy-making informers, especially by the Organisation for Economic Co-operation and Development [5]. Agency is a basic need in any goal-oriented work, particularly in jobs that call for creativity and continuous development in work practices [6]. This means that graduates of higher education institutes, in particular, should be prepared to act as developers and change agents in their field. However, despite this need—especially in the COVID-19 context but also in general—and the particular emphasis on student agency by policy-making informers, student agency has received little explicit attention in educational practice in higher education so far.

Learning analytics (LA) refers to a research field that harnesses data on learners to understand, improve, and optimize learning [7]. The use of LA can, for example, predict academic success, improve quality assurance, and identify at-risk students [8]. Moreover, dashboards are often utilized to visualize learning processes and study pathways—not only to increase awareness but also to give personalized feedback

The associate editor coordinating the review of this manuscript and approving it for publication was Shuihua Wang.

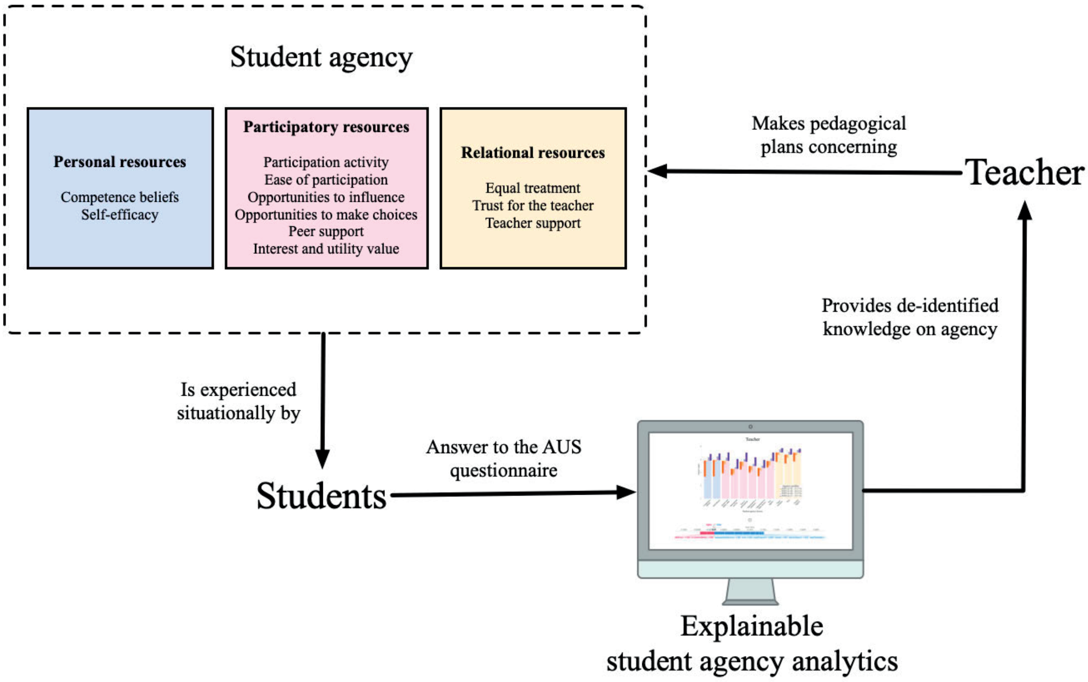


FIGURE 1. The XSAA process can be depicted as a loop, which starts when the teacher makes the initial pedagogical plans. At some point in the learning and teaching process, the students complete the AUS questionnaire, and the agency analytics is executed automatically. The teacher receives results, and can then adjust the pedagogical plans according to the students' experienced resources of agency.

to the learners. This kind of personalized feedback and consideration of the personal traits of learners can positively influence the learning process and outcomes. Since it is usually unfeasible for teachers to manually provide such individualized feedback to all students—especially for teachers in higher education settings who often have to instruct hundreds of students with different backgrounds—such automated feedback can offer significant support.

Jääskelä *et al.* [9] examined student agency as the theoretical framework for assessing and enhancing digital education at universities by making use of LA. Based on a factor and robust cluster analysis process, which is conducted to measure students' responses to a validated scale [3], [9], the students receive automated feedback on their individual agency profile. In addition, the teacher of a higher education course gets an aggregated overview of the different student agency profiles. The essence of this automated agency-based process—which is called student agency analytics (SAA)—is to provide actionable information for students on their learning efforts in relation to their perceived affordances in the course and for teachers on students' judgements of their situational agency to increase pedagogical knowledge.

In a recent review, Deeva *et al.* [10] classified automated feedback systems by their applied educational settings, the properties of their delivered automated feedback, and their design and evaluation approaches. They concluded that applied learning theories or educational frameworks had not been reported in most cases. Moreover, they urged the developers to use more data-based solutions and to be able to explain the reasons behind the automated system. Therefore, the purpose of the present article is to show how the integration of explainable artificial intelligence (XAI) techniques with the SAA process (see Figure 1) can support the transparency and data-based development of automated feedback systems in education. More precisely, we aim to integrate XAI techniques into the SAA process in the context of higher education. This procedure improves awareness of different stakeholders from such organizations on the learning arrangements, considers the complexity of the students' capacities and various contextual resources, and supports reflection.

Another reason why we aim to integrate XAI techniques within SAA is that explainability became a key issue in LA [11]. Relationships in educational data are often complex [8], [12], and several theoretic LA studies have shown that these relationship can be modeled better by complex models than by simple linear ones (e.g., [13]–[16]). However, in practice, these complex models are rarely used because they are reckoned to be inexplicable. XAI is an emerging research direction that can help the user or developer of complex models understand the model's behavior and provide human-understandable justifications for it [17], [18]. Thus, the integration of XAI techniques allow us to also use the better performing complex LA models in SAA and to explain them in such a way that even practitioners with no background in data analysis can easily understand them.

To demonstrate our explainable SAA process (XSAA), we provide the results from a study of four concurrently implemented courses on mathematics in an engineering education degree program. The content and curriculum of these mathematics courses are identical but they are taught independently by four different teachers. This means we built and explained our models not only by using the student-specific agency data but could also link them to the particular teaching approaches of the instructors. Such a setting is new and might help teachers to increase their awareness of the effects of their pedagogical planning and interventions.

The main contributions of this paper are twofold:

- We use XAI to produce explainability and actionability through dashboards. These dashboards not only show summaries of the raw student data (e.g., how active they were with the tasks or how long it took to solve a problem) but also—through nonlinear and universal machine learning models—explain the reasons for the students' actions, linking them to a well-defined body of pedagogical planning by the teacher.
- We discuss the usability of the results gained through XSAA at the teaching practice level; that is, how they may help teachers in reflecting and designing their curriculum and in developing agency-supportive practices in their teaching implementations.

The rest of the paper is organized as follows. Section II outlines the background at the basis of our contribution. First, we locate our research among the previous studies in the field of LA and XAI in higher education. Second, we summarize previous student agency LA studies. Section III provides a discussion of the need for explainable models, especially in LA. It also provides an overview of the different XAI techniques that we are using for our SAA dashboards. Section IV presents an example of an application of our explainable SAA in higher education (i.e., the data and our XSAA results from the four groups of students studying the same mathematics course taught by different teachers at a university of applied science). Finally, Section V presents the main findings and implications of our study.

II. BACKGROUND

A. LA AND XAI STUDIES IN HIGHER EDUCATION

Hundreds of primary studies depicting and analyzing the use of LA to improve educational actions in higher education institutes (HEI) have been published, and their impacts and outcomes have been summarized in many recent reviews (e.g., [19]–[21]). Their overall conclusions suggest that LA should be better grounded to learning science, its effectiveness should be assessed, and actual linkages to teachers and pedagogical planning should be emphasized.

For example, the review by Aldowah *et al.* [19], which included 402 articles from 2000 to 2017, presented many student-oriented characteristics such as "engagement," "achievement," "participation," "reflection," "motivation," and "satisfaction" to be approachable by using LA techniques. However, no linkage to the actual teaching activities was presented. In the combined review-meta-review by Du *et al.* [22] from 901 identified research papers from 2011 to 2017, the authors mentioned that instructors need to connect LA with learning science and use dashboards for student monitoring. Similarly, the knowledge gap between the theoretical frameworks of educational domain knowledge and the LA models was emphasized in the review by Cui *et al.* [23].

After multistage screening, the review by Sonderlund *et al.* [24] ended up with only 11 studies out of 689 that were found to evaluate the effectiveness of LA interventions, concluding that the lack of intervention studies where the educational institution (in practice, the instructor of a course in HEI) performs and evaluates systematic changes of its actions. Moreover, based on analyzing 252 papers published during 2012 to 2018, Viberg *et al.* [21] concluded that "the overall potential of LA is so far higher than the actual evidence, which poses a question of how we can facilitate the transfer of this potential into learning and teaching practice." Likewise, Ifenthaler and Yau [20] addressed the study success of HEI students through 46 primary studies, concluding that the lack of "rigorous, large-scale evidence of the effectiveness of LA in supporting study success." To this end, the review by Leitner *et al.* [25], which was based on 101 articles during 2011–2016, nominated teachers solely as a "side-product" of the research field.

Contrary to the huge amount of LA in HEI studies, studies dealing with XAI in HEI are extremely scarce. A Google Scholar and Scopus search in May 2021 identified only three studies of XAI in HEI [26]–[28]. Putnam and Conati [26] conducted experiments with nine university students testing whether the students would like to receive explanations for hints given in an intelligent tutoring system (ITS). They concluded that the majority of students would like explanations in the ITS, but the actual implementation of XAI was presented as future work. Likewise, Conati *et al.* [27] discussed only theoretically necessary considerations to make an ITS explainable for the benefit of learning. Alonso and Casalino [28] used XAI for a distance learning set. However, they did not provide any description of XAI techniques and solely used existing software (WEKA) to gather explanations for their prediction models. In sum, all three articles emphasized the need for XAI in automated feedback systems in HEIs, but none implemented and explained the underlying XAI techniques.

B. STUDENT AGENCY ANALYTICS IN A NUTSHELL

1) STUDENT AGENCY IN HIGHER EDUCATION

Agency has been under consideration in several disciplines and has been highlighted in various areas of life. In general, agency is one's capacity to act and cause change. However, different disciplines have their own and more detailed

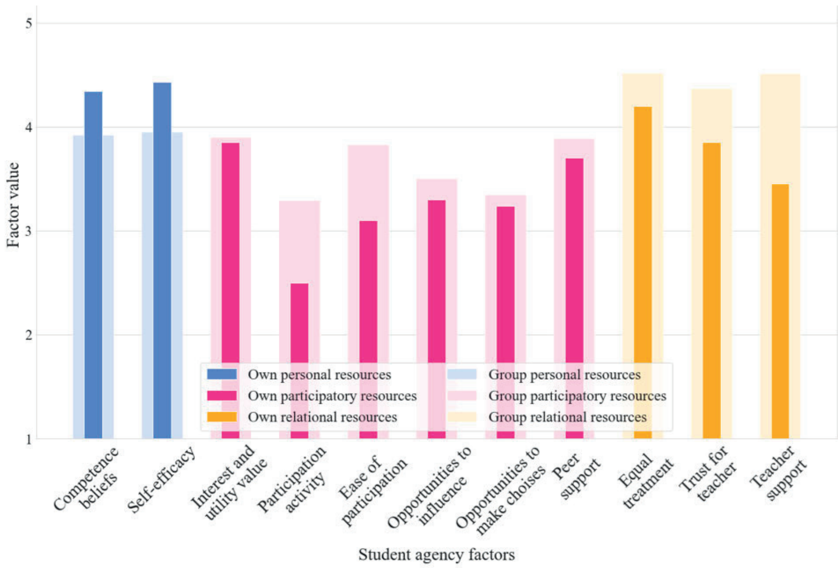


FIGURE 2. **Student agency analytics provides information about the inter-individual differences relating to resources of student agency. This figure shows a student's personal report that consists of his/her individual agency profile in comparison with the general agency profile of the group. A teacher's report consists of a general agency profile of the group combined with four prototypical agency profiles (as visualized in Figure 3).**

perspective on the meaning of agency. For example, in social cognitive theory, agency is understood as an individual's capability to engage in intentional, self-defined, and meaningful action [29]. Similarly, in social sciences, the concept of agency concerns an individual's capability to take intentional and self-defined (i.e., autonomous) action and is focused on the circumstances and structural factors that constitute frames for action (e.g., [30]). Contemporary educational discourse has emphasized the meaning of agency in lifelong learning [31] and in student-centered learning [32]. Within educational sciences, agency is seen as an integral part of learning, which manifests itself both as individuals' active action in knowledge construction (e.g., [33]) and a sense of being empowered in learning situations [34].

Our stance on student agency is based on the conceptualization made by Jääskelä *et al.* [3], who synthesized the previous literature on agency and defined student agency in higher education as "a student's experience of having access to or being empowered to act through personal, relational, and participatory resources, which allow him/her to engage in purposeful, intentional, and meaningful action and learning in study contexts." Student agency consists of three resource areas (see Figure 2). *Personal* agency resources consist of the dimensions of competence beliefs and self-efficacy. *Relational* resources refer to power relations in different educational settings, which include the experiences of equality among the students, trust for the teacher, and support from the teacher. *Participatory* resources of student agency involve dimensions relating to engaged and active participation in learning. Altogether, student agency is composed of 11 dimensions, and it is measured using a validated psychometric Agency of University Student (AUS) scale [3], [35].

2) STUDENT AGENCY ANALYTICS

Discerning different study experiences can be demanding in heterogeneous educational settings with a multitude of students. To address this challenge, we apply a LA process called student agency analytics, which utilizes robust statistics and psychometric information obtained using the AUS scale [9]. First, the students in a particular study group or course complete the AUS questionnaire. Second, the individual factor values of agency are calculated for each student using the factor pattern matrix, which enables the determination of the general agency profile of the whole study group. Third, unsupervised learning, specifically robust clustering, is used to provide prototypical agency profiles with four distinct groups based on cluster validation indices, as described in more detail in [9]. Kruskal-Wallis H and Mann-Whitney U tests can then be used for explaining the clustering results through the agency dimensions. Moreover, if the information on the quality of learning outcomes or course grades is available, it can be linked to the prototypical agency profiles using supervised learning.

The main representations obtained using SAA are the students' individual agency profiles (IAPs), the general agency profile (GAP) of a group (e.g., study group, course), and four distinct prototypical agency profiles (PAPs) within a group. IAP (Figure 2) represent the values of individual student's agency dimensions, which can be compared with the GAP.

IAP is a personal depiction, and it is aimed only at the student accompanied with general information about student agency. For the teacher, student agency analytics provide a general overview of the agentic resources of the students. To preserve students' privacy, teachers do not receive individual student profiles. Instead, their report consists of de-identified information about the GAP and PAPs. Both the GAP and PAPs are presented in the teacher report as a special combined bar graph (Figure 4).

3) TEACHER'S PERSPECTIVE

Teachers' actions and their pedagogical choices influence students' learning experiences (e.g., [36]–[40]). In terms of pedagogical planning, teachers would benefit from the analysis results concerning all their students. For instance, peer support can help students in higher education to develop self-regulation skills, decreasing or allowing better management of study-related exhaustion [41]. Thus, it would be worthwhile for the teacher to identify the different experiences of peer support to provide means and opportunities for students to actualized supportive collaboration. Students' prior knowledge can significantly influence student achievement [42]. Failing to consider students' prior knowledge might be manifested as a lack of competence beliefs and self-efficacy. In summary, becoming aware of students' agentic experiences could help teachers make better pedagogical plans and decisions.

From the teacher's perspective, SAA summarizes the inter-individual differences of learning experiences in a visually interpretable form. As a result, students' general assessment of their agency and four distinct student agency profiles are presented to the teacher. The process can be depicted as a loop (see Figure 1), which starts when the teacher makes the initial pedagogical plans. At some point in the learning and teaching process, the students complete the AUS questionnaire, and the agency analytics is automatically executed. The teacher receives results, which visually describe the GAP and the PAPs. The teacher can then adjust the pedagogical plans according to the students' experienced agency resources. In the following sections, we develop the SAA process toward explainable LA.

4) ETHICAL CONSIDERATIONS

A general prerequisite in LA should be the responsible use of educational data [43]. It is worth emphasizing that SAA aims not to evaluate or grade the students or their learning. Instead, the purpose is to identify and make visible different personal learning experiences through the concept of agency. Thus, it is essential to ensure the privacy of the students and teachers. The individual agency profile received by a student is personal and only for the student's use. Teachers or anyone else do not see the student's IAP unless they want to disclose the results, for example, to help study counseling. Generating aggregated results (GAP and PAPs) provide a means to present detailed but de-identified information for the teacher. Similarly, the teacher report depicting the aggregate results of a course is meant only for the teacher to use in personal pedagogical planning. The results should not be used to evaluate the individual teachers or their teaching.

III. TOWARD EXPLAINABLE LEARNING ANALYTICS

From a technical point of view, LA is about modeling students and learning. Its methods have roots in several different disciplines, such as statistics, education, psychology, and machine learning [44], [45]. While traditionally, statistical models were mainly used in LA to scaffold students and help teachers, the machine learning models have gained in importance in recent years [46]. This is mainly due to the challenge of modeling the increasingly rich, varied, and multimodal (such as eye tracking, physical movement, and face recognition for emotion detection) LA data [47], [48].

Often a trade-off occurs between the performance of a specific machine learning model and its explainability. For example, in supervised learning, the performance (i.e., the difference between the real outputs and the outputs of the model) is usually better for complex models with nonlinear combinations of inputs, but such models are harder or even impossible to understand. These kinds of models are also called "black boxes." On the contrary, simple linear methods are prone to perform worse, but they are easier to interpret and understand. One example of the latter is a linear regression model, where the coefficient of an input can be directly interpreted as the importance of that input.

Although they usually perform better, black boxes have several problems. One problem relates to assuring that such a model works as intended. If not even the designer of the model can explain the model's underlying logic and how it arrived at a result, it is impossible to verify that the model uses the right justifications for its decisions. In the worst case scenario, such black-box models may use questionable reasons for their decisions without anyone noticing them. This usually happens if they adopt bias in the training data. Bolukbasi *et al.* [49], for instance, showed that a model that was trained on a corpus of Google News text, learned the correct word embedding "man is to woman as king is to queen," but at the same time also learned the worrisome embedding "man is to woman as computer programmer is to homemaker."

Another example, discussed by Freitas [50], comes from the military: The military trained a classifier to distinguish pictures of enemy tanks from pictures of friendly tanks. This classifier was performing well on the training set but showed poor performance when it was used in the field. Later it was discovered that the pictures of enemy tanks in the training set were taken mostly on overcast days, while the pictures of other tanks were taken on fair weather days. It turned out that the classifier had learned this pattern from the training set and consequently mostly used background features to classify the tanks. Such examples prevent users from trusting a black box model. In fact, some studies have shown that even if they are proven to be more accurate than human forecasters, most people exhibit an inherent distrust of automated predictive models [51]. If the users do not trust a model or a prediction, they will not use or deploy it. Thus, the explainability of models is important, not only for developers but also for the end users, and all other parties involved.

XAI is a new research field. It refers to approaches attempting to make machine learning models more explainable and to address the above-mentioned issues. Several XAI review papers were recently published, indicating its importance and topicality [18], [52]–[56]. Generally, the explainability of a model refers to any approach that helps the user or developer understand the model behavior and its reasoning process [17]. While no definition of XAI is uniformly accepted, it can be conceptualized as the ability to provide human-understandable justifications explaining the way in which a model works so that observers can understand how and why it has delivered particular outcomes. For example, in the military classifier case discussed above [50], an explanation would have shown that the classifier used the background instead of the features of the tanks for classifying the photos. Thus, XAI can help to identify potential bias in the training data, ensure algorithmic fairness, and verify that the algorithms perform as intended [53].

As pointed out by Baker [11], explainability is also one of the biggest challenges in LA nowadays. Several LA studies have shown that complex models outperform the simpler ones. However, if an instructor does not understand such a complex LA model and if a development team cannot explain it, the LA model will probably never be employed in practice (ibid). Instead, only simple linear models that have been around for years continue to be used. This is a problem, because as argued for example in [12], relationships in educational data are often complex and cannot be modeled well enough with the simple models. If the better performing complex models could also be explained in such a way that even practitioners with no background in data analysis could easily understand them, they would probably be employed more often.

Conati *et al.* [27] argued that the explainability of models is also important for learners: For instance, if learners cannot comprehend the logic of an intelligent tutoring system, they are not motivated to follow the systems instructions and their trust in the system as a whole will decrease. Another reason the explainability of LA models has become increasingly important is that the new General Data Protection Regulation (GDPR) now includes a right to explanation and information [57], [58]. This means that if automatic profiling (e.g., in student analytics) is used, it is not only a desiderata but actually a requirement to be able to explain to a student why he/she was assigned to a particular profile.

In general, one can distinguish XAI methods that are intrinsic, meaning interpretable due to their simple structure, and post-hoc XAI methods, meaning methods applied after model training to explain the model's logic in retrospect. Moreover, one distinguishes between local and global explanations [59], [60]. While modular global explanations provide interpretation for the model as a whole, approaching it holistically, a local explanation provides interpretation for a specific observation (such as one particular student). Finally, explanation techniques can be model specific, meaning the explanation technique is specific to its model, or model agnostic, meaning the explanation technique can be applied to any model.

In this work, we use both intrinsic model-specific and post-hoc model-agnostic explanations as well as global and local explanations. Moreover, we want to explain not only the most important characteristics of the different agency profiles (global explanations) but also explain, for specific observations, why they were assigned to a particular group (local explanations). The latter are especially interesting for instructors who receive a report about their students' agency and can then see why a particular student was assigned to a particular agency group. Finally, as pointed out above, students have a right to information about individual decisions made by agency algorithms, and the local XAI techniques enable us to provide such information.

A. MULTINOMIAL LOGISTIC REGRESSION

Logistic regression is an example of a machine learning method that because of its linear structure is intrinsically explainable and offers model-specific modular global explanations. It is probably the most traditional technique to predict a categorical response variable (i.e., the class). If the class is dichotomous, a simple logistic regression can be used that employs a logistic function to measure the relationship between the class and the explanatory variables through estimating probabilities. If the class has more than two categories, multinomial logistic regression should be used. Multinomial logistic regression uses the softmax function (i.e., a generalization of the logistic function to multiple dimensions) to calculate the probabilities of each class category over all possible class categories. These calculated probabilities are then used for determining the class (i.e., the response variable category) for the given inputs.

Logistic regression is intrinsically explainable through its coefficients. The coefficient of a continuous explanatory variable can be explained as the estimated change in the natural log of the odds for the reference event for each unit increase in the predictor [61]. In general, the larger the absolute magnitude of a coefficient is, the more relevant the corresponding explanatory variable is for the classification. Moreover, the sign of the coefficient indicates whether the explanatory variable increases or decreases the probability of belonging to a certain class. Furthermore, if the logistic regression model is penalized with the l_1 norm, some of the feature coefficients shrink to exactly zero, which makes the model simpler and easier to explain [62]. However, although (multinomial) logistic regression generally meets the characteristics of an explainable model, Arrieta *et al.* [63] point out that it may also demand post-hoc explainability techniques, such as visualizations, particularly if the model is to be explained to non-expert audiences.

B. MULTILAYER PERCEPTRON

A multilayer perceptron (MLP) is an example of a machine learning technique that is also able to find and model complex nonlinear interactions in data and, thus, often outperforms linear techniques, such as the previous discussed logistic regression. It consists of an input layer, at least one hidden layer, and an output layer. Each layer consists of nodes, and except for the input nodes, all nodes are neurons with nonlinear activation functions. MLPs are fully connected, meaning that each node in one layer connects with a certain weight w_{ij} to every node in the succeeding layer. These weights on the nodes are automatically adjusted to construct the mathematical model that most accurately maps the input features (such as the agency dimensions of the students and the information in which course he/she was studying) to the output labels.

However, MLP models are generally regarded as black boxes and opaque. For example, even when techniques are used to identify the features that a particular MLP model assigned significant weights to, the relationships between those features and the classification can be weak because a small permutation in a seemingly unrelated aspect of the data can result in a significantly different weighting of features [64]. Moreover, different initial settings can result in the construction of different models [65].

C. RANDOM FOREST

Random forests, as well as other tree-based techniques, are one of the most popular nonlinear supervised machine learning methods nowadays [66]. They are ensemble learners based on decision trees, which are on the one hand, explainable and able to model nonlinear relationship in data, but on the other hand, generally low performing because they tend to overfit the training data. Through growing each tree in the ensemble (i.e., the forest) only on a bootstrap sample from the original data and by randomly using only a subset of the features for each node in each tree, random forest keep the main advantages of decision trees while at the same time overcoming their disadvantage. In other words, random forest are also explainable and able to model nonlinear relationship in data, but—through the bagging of many uncorrelated decision trees—surmount the overfitting and low-performance issue of decision trees. In fact, they perform so well that they are often the winner in machine learning competitions [66], [67]. Nevertheless, although the importance of a global model-specific feature is generally provided with the random forest implementation (for example, in Python, Gini measures the global importance of the input features), less attention has been paid o far to local explanations for random forest predictions [66].

D. LOCAL INTERPRETABLE MODEL-AGNOSTIC EXPLANATIONS

Local interpretable model-agnostic explanations (LIME) are a XAI tool developed by Ribeiro *et al.* [68]. LIME provides explanations, such as features and rules of features, that were important for predicting a specific observation (i.e., local explanations). It can be used for any prediction model, meaning it is model agnostic, because it does not even need know the actual ''black box'' prediction model f; it just uses its predictions. More specifically, it changes the model's inputs and then uses the model's outputs to make conclusions about the model. The main idea is that if the model prediction does significantly changes after the value of a feature is slightly adjusted, that feature may be an important predictor. Vice versa, if the prediction does not change, the changed feature may not be important at all.

It accomplishes this by taking the observation x for which the prediction should be explained and permuting its feature values. All of these permuted fake observations are weighted by their distance to x. Then, the black box model f is used to predict the permuted observations, and a new surrogate/explanation model (can be any explainable model, such as a linear model or decision tree) g is trained that reflects the original predictions as accurately as possible, while the complexity of this surrogate model is kept as low as possible. Then the explanations of the simple surrogate model (for example, the weights if g is a linear model) are used to explain the local behavior of $f(x)$.

Mathematically, this can be expressed as follows:

$$\xi = \arg\min_{g \in G} \mathcal{L}(f, g, \pi_x) + \Omega(g),$$

where π_x is the proximity measure to define locality around x, and $\Omega(g)$ is the complexity of g that should be kept low (for example, by minimizing the number of non-zero weights if g is a linear model).

The advantages of LIME are that it is relatively easy to use and understand. However, certain drawbacks are associated with it. One of these is the potential inconsistency between the surrogate model prediction $g(x)$ and the real model prediction $f(x)$. Another drawback is the lack of comparative values for the LIME values. SHAP, which will be discussed below, overcomes these drawbacks.

E. SHapley Additive exPlanations

Shapley values, introduced by Shapley [69], originate from cooperative game theory. They measure the fair payout that each player should receive based on his/her contribution to the total payout of the game. The payout for each player is proportional to his/her marginal contribution to the total payout. Similarly, when used as an explanation for a prediction, a Shapley value measures the contribution of an individual feature to the total prediction. This means a Shapley value is the average marginal contribution of a feature value across all possible coalitions of the features.

The fair contribution of feature i is obtained by taking the average of the contribution over the possible different permutations in which the coalition can be formed. Mathematically, this can be expressed as follows:

$$\phi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(N - |S| - 1)!}{N!}(v(S \cup \{i\}) - v(S)),$$

where N is the number of all features, S a subset of the N features, and $v(S)$ the prediction of the S features. When feature i joins the S features, its marginal contribution is $v(S \cup \{i\}) - v(S)$.

Shapley values come with four desirable properties: (i) efficiency, meaning that the sum of the Shapley values of all features equals the value of the total coalition; (ii) symmetry, meaning that all features have a fair chance to join the prediction; (iii) dummy, meaning if a feature contributes nothing to any coalition S, then the contribution of that feature is zero; and (iv) additivity, meaning that for any pair of predictions v, w: $\phi(v+w) = \phi(v)+\phi(w)$, where $(v+w)(S) = v(S)+w(S)$ for all S.

SHapley Additive exPlanations (SHAP) are a XAI tool developed by Lundberg and Lee [70] that uses these Shapley values to explain machine learning models. It includes the model-agnostic SHAP `KernelExplainer` that works universally for any prediction model. The `KernelExplainer` builds a weighted linear regression by using the given data, the predictions, and the function/model that predicts the predictions. It computes the feature importance values based on the Shapley values and the coefficients from a local linear regression. Besides the `KernelExplainer`, the SHAP tool also includes other explainers that have been optimized for specific models. One example is the `TreeExplainer`, which was optimized for tree-based prediction models [66]. According to Lundberg *et al.* [66], it is the only tool that enables the exact computation of optimal local explanations for tree-based models. The `TreeExplainer` can also be used as a global explanation method by averaging local explanations. For example, if this is done over all instances in a dataset, it results in a global measure of feature importance.

IV. APPLICATION OF EXPLAINABLE STUDENT AGENCY ANALYTICS

In this section, we present the results from an application of XSAA in higher education. All the analytics were performed in Python 3.8.2, using `LIME` and `SHAP` toolboxes.

Sample and Study Context: Four courses on mathematics (A1-A4) of first-year engineering students ($n = 141$) in a Finnish higher education institution (university of applied sciences, ISCED Level 6) were studied. Each course had a different responsible teacher but the same basic contents and learning goals. The teaching arrangements as a whole were mostly traditional: lectures and guided exercises in a classroom and additional homework. The courses consisted of instructional videos, automatic tests that guided the student depending on the answers, and a final test. In addition to class hours, teachers sent emails to the whole student group using the virtual learning environment. Personal messages between teachers and students were exchanged by email. In all the courses, mid-term feedback was collected, and depending on the results, some small modifications were made (for example, more time was allocated to topics the students found challenging). All the courses also had voluntary support classes guided by the teacher.

Different practices were also used between the courses. Attendance affected the evaluation in one course (A2). Two courses (A1 and A4) made continuous self-assessments; one based on homework and their model solutions (A1) and the other based on the results of automatic tests in the learning environment (A4). One course (A3) had extra support hours guided by a student assistant. In one course (A4), the students had the opportunity to get a small amount of personal guidance from the teacher if necessary. Moreover, this course (A4) made weekly applications on the topics practiced and had small teams.

Analysis Between Prototypes: Prototypical student agency profiles were created using clustering. The different prototypical agency profiles (PAP1–PAP4) and the general agency profile (GAP) are presented in Figure 3. GAP is the profile of all the analyzed students. All the agency dimensions maintain the order from the lowest profile PAP1 to the highest profile PAP4. In general, the relational resources of student agency (equal treatment, trust for the teacher, and teacher support) were experienced as the highest resource domain and > 4 in all profiles except in PAP1. Three of the participatory resources (participation activity, opportunities to make choices, and opportunities to influence) were generally experienced as lower than other resources in all the profiles. The rest of the participatory resources and the personal resources were experienced close to the factor value of 4 at the GAP level. PAP1 was particularly characterized by low personal resources.

Analysis Between Courses: The analysis between courses revealed differences in student agency between the four different course instances (A1–A4). Figure 4 presents the box plots of each student agency dimension in each of the course instances. There were statistically significant differences in all the dimensions based on the pairwise comparison using the Mann-Whitney U statistics. In particular, the student agency dimensions of trust for the teacher, teacher support, and opportunities to influence were experienced as lower in the A3 course instance comparing to other courses, and the difference was statistically significant.

We also examined if there were any dominant prototypical profiles present in each of the courses (Table 1). Based on the chi-square test of the contingency table, statistically significant differences were observed; $\chi^2(9, n = 141) = 30.1$, $p < .001$. More students were assigned to the higher agency profiles PAP3–PAP4 in the courses A1 and A4. In course A4, no students were observed in the low agency profile PAP1. In course A3, the majority of the students were in the profiles PAP1–PAP3, and only 5% were in the high agency profile. In A2, a somewhat equal quantity of students were assigned to each PAP.

Prediction Results: In comparison to earlier work, we not only created the student agency profiles here but also built

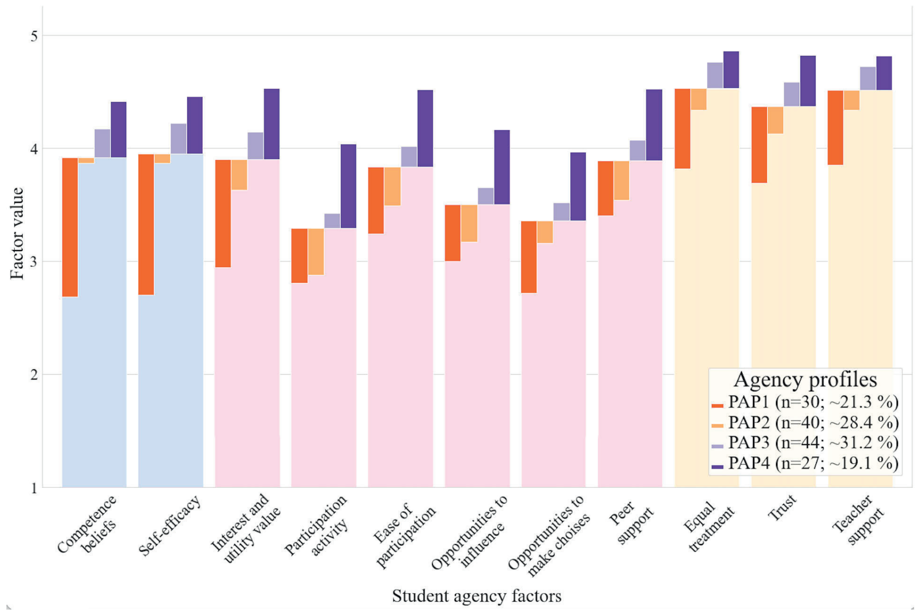


FIGURE 3. Student agency prototype profiles (PAP1–PAP4) and the general average profile of students (n = 141) studying in an engineering education program in a higher education institution.

TABLE 1. Students representing the different prototypical profiles PAP1–PAP4 in each course instance A1–A4, with row-wise percentages; $\chi^2(9, n = 141) = 30.1, p < .001$.

	PAP1	PAP2	PAP3	PAP4
A1	4 (13%)	4 (13%)	16 (53%)	6 (20%)
A2	10 (30%)	8 (24%)	7 (21%)	8 (24%)
A3	16 (28%)	17 (29%)	22 (38%)	3 (5%)
A4	0 (0%)	11 (37%)	9 (30%)	10 (33%)

models predicting these profiles. Using these models, their global model-specific explanations, and local model-agnostic LIME and SHAP explanations on top of them allows us to identify the most important characteristics explaining why certain students are assigned to certain profiles. To predict the multinomial class label (i.e., the agency profile), we used all 15 features: the 11 agency dimensions and the four course variables that were one-hot encoded into binary features.

To estimate and compare the models for the supervised task (i.e, predicting the student profile), we divided the data with a stratified split into a training (80%) and an independent test set (20%). Then, we used stratified fivefold cross-validation on the training set to estimate the best hyperparameters for the classifiers. We compared the multinomial logistic regression (MLR) with l_1, l_2, and elasticnet penalization, random forest, and MLP classification models to predict the agency profile. Table 2 summarizes the best model for each classifier as determined through the fivefold cross-validation on the training set and its performance on the independent test set. As shown in the table, the two nonlinear classifiers (random forest and MLP) outperformed the three linear classifiers. Overall, random forest was the best performing classifier

TABLE 2. Accuracy of the supervised models predicting the agency profile PAP1–PAP4 of the student.

Classifier	test set accuracy	train set mean (std)
MLR l_2	0.724	0.767 ($\pm$ 0.098)
MLR l_1	0.897	0.839 ($\pm$ 0.061)
MLR elastincnet	0.793	0.829 ($\pm$ 0.088)
Random forest	0.966	0.863 ($\pm$ 0.069)
MLP	0.897	0.875 ($\pm$ 0.103)

when comparing all classifiers, and multinomial logistic regression with l_1 penalization was the best linear classifier.

Global Explanations: Since random forest was the best classifier overall and the multinomial logistic regression with l_1 penalization the best linear classifier, we focused on these two models to explain the prediction results. Figure 8 shows the coefficients of the multinomial logistic regression with l_1 penalization predicting the highest agency profile PAP4. Figure 9 shows the coefficients of the multinomial logistic regression with l_1 penalization for all four agency profiles. The figures illustrate that overall, the agency dimensions seem more important for the prediction model than the course variables. However, being in a certain course can also increase or decrease the probability of belonging to a particular agency profile. For example, being in course A1 decreases the probability of belonging to agency profile PAP2 and increases the probability of belonging to agency profile PAP3 (see Figure 9).

Figure 10 shows the importance of the features of the random forest model predicting the agency profile. In comparison to the coefficients from the multinomial logistic

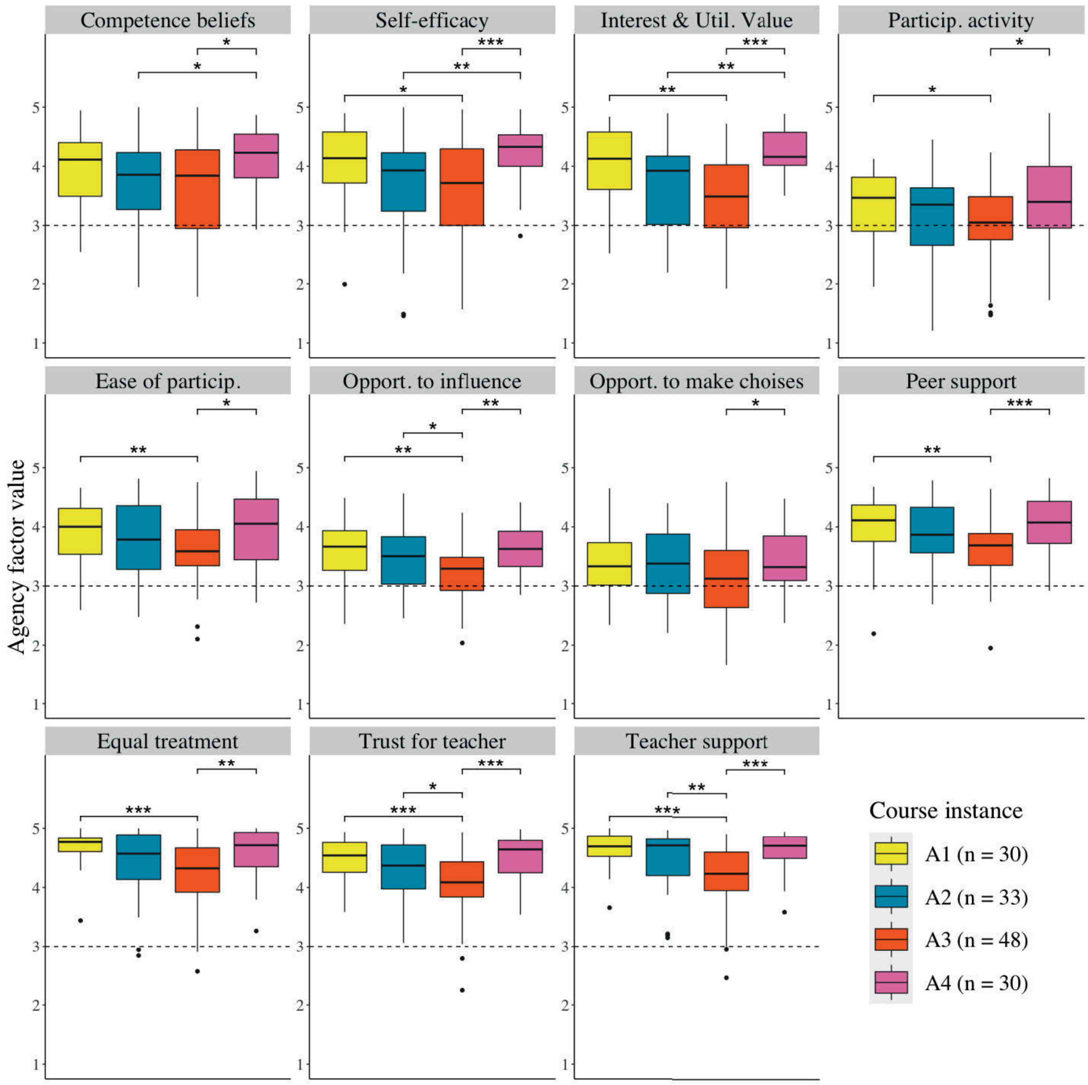


FIGURE 4. Student agency dimension in each course instance and pairwise statistical significance using Mann-Whitney U statistics. As usual, $\star$ corresponds to $p < 0.05$, $\star\star$ to $p < 0.01$, and $\star\star\star$ to $p < 0.001$.

regression, the feature importance levels of the random forest are always positive and do not encode which class a feature is indicative of. The random forest feature importance levels can tell us that a certain feature is important, but not whether it is indicative of a student having agency profile PAP1, PAP2, PAP3, or PAP4. Moreover, they provide no information in regard to whether a high feature value increases or decreases the probability for a certain class. They just summarize the importance of each feature for the whole model.

If we combine all the local SHAP values (the results of the individual local explanations are provided in the next section) for all the students, we can also get the global SHAP explanations for a model. This is shown in Figure 5 for the random forest classification model. As the figure shows, a student's competence belief was the most important feature for the model, especially when determining if he/she belongs to the lowest (PAP1) agency profile. This model-agnostic explanation is the same as that from the model-specific feature importance levels (see Figure 10, here the competence belief was also the most important feature) but more informative as it also shows which features are important for each profiles.

Local Explanations: As explained in Section III, local explanations enable us to explain why a certain student received his/her prediction and the contributions of the individual predictors. Global feature importance, as discussed above, only shows the results across the entire population,

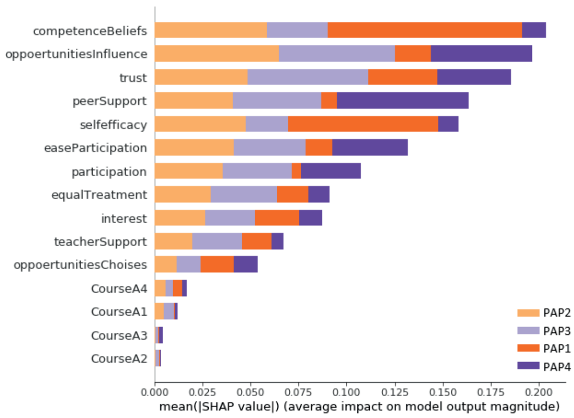


FIGURE 5. Global SHAP explanations for the random forest model. For competence beliefs, the mean absolute SHAP values are 0.1 for PAP1, 0.06 for PAP2, 0.03 for PAP3, and 0.01 for PAP4, making it altogether the most important global predictor for this model.

but not on each individual student. The local explanations, in contrast, enable us to pinpoint and contrast the impacts of the factors for particular students.

To explain the model predictions for particular students, we used the true positives with the highest probability for each agency profile; that is, those four students from the test set that the model correctly predicted to be PAP1, PAP2, PAP3, and PAP4, respectively, with the highest probability. Table 3 summarizes these local explanations for the random forest model. As we saw already in the global model-specific explanations (Figure 10), the opportunities to influence was one of the most important variables for the random forest model. However, from Table 3, we can also see for which profiles this variable was especially important (namely, agency profile PAP2, PAP4, and especially PAP3).

The LIME rules can also be presented visually. Figure 11 shows the LIME rule visualization for the PAP2 student who was predicted to be a PAP2 profile with the highest probability with the random forest model. For comparison, Figure 6 shows the SHAP local explanations for the same model and student. This plot provides a more comprehensive explanation overview of the prediction than the LIME rules.

More specifically, as Figure 6 shows, the model predicted an 88 percent chance that this student was a PAP2 student, whereas the base value (i.e., the prediction if nothing would be known about this student) for PAP2 was a 29 percent chance. The feature values causing increased predictions are in red, and their visual size shows the magnitude of the feature's effect. The biggest impact comes from the opportunities to influence, which is 3.16 for this student. The feature values decreasing the prediction are in blue. As can be seen in Figure 6, the fact that this student is in course A1 had a meaningful effect, decreasing the prediction. The model predicted some tiny probabilities that this student was a PAP1 or PAP3 student, but his/her competence beliefs are lower than for PAP3 and higher than for PAP1 students. If one subtracts the length of the blue bars from the length of the red bars, it equals the distance from the base value to the output. This means that the baseline plus the sum of individual effects add up to the prediction as discussed in Section III.

Local Explanations for the Student Needing the Most Support: The local explanations also enable us to locate the students needing support the most and to receive the explanations describing which factors could affect a change toward higher agency. Based on Table 1 and Figure 9, we can conclude that the students in course A3 needed the most support. Since profile PAP1 represents the lowest agency profile, we chose the student from the test set who was in course A3, and was predicted to have the lowest agency profile PAP1 with the highest probability, for the local explanations. Figure 7 shows the SHAP values explaining why this student was assigned to profile PAP1 with the highest probability. As Figure 7 illustrates, the base value of the prediction in the absence of any information on the independent variables is 0.2138. Knowing that the competence beliefs of this student are only 1.907 increased the prediction that this student is PAP1 by 0.222, and knowing that the self-efficacy value of this student is 1.878 increased the prediction for profile PAP1 by another 0.176 (see Table 4).

A. SUMMARY AND DISCUSSION OF RESULTS

Our results can be summarized from the application level and the methodological level. From the application level, we can conclude that the level of student agency was higher in the two courses, A1 and A4, where continuous task-driven self-assessment took place. No students were in the lowest agency profile PAP1 in the course A4, and the majority of the students in A1 and A4 were in the higher agency profiles PAP3 and PAP4. One reason for the students' generally high sense of agency in course A4 might be the personal guidance that the teacher offered in the course. Furthermore, a joint analysis of Figure 8, Figure 5, and Table 3 suggests that if the students found support from their peers and experienced opportunities to influence and participate in the course, they tended to have higher agency profiles.

From the teacher's perspective, the XSAA results could provide insight for pedagogical planning. For example, the students in course A1 seem to have received the proper amount of teacher's support and attention, as relational resources were scored high and those resources represented some of the most important resource areas for the second highest agency profile PAP3 (Figure 3, Figure 5, and Table 3). To foster student agency of the PAP2 and PAP3 students in A1, the teacher could provide low-threshold ways for participation because the participatory resources were considered important in the highest profile PAP4. In addition, suggestions to improve pedagogical planning could be made by analyzing the characteristics of the students in the lowest agency profile PAP1. The findings suggest that low self-efficacy and competence beliefs are important common nominators for students in PAP1 (Figure 3, Figure 5, and Figure 7). As there were many PAP1 students in course A3, these students might

TABLE 3. LIME rules explaining the true positive students for each profile from the test set with the highest probability with the random forest model. For each student, the rules are ordered by importance with the most important rule first.

Profile PAP1		Profile PAP2	
rule	importance	rule	importance
competenceBeliefs <= 3.44	0.241	participation <= 2.81	0.095
selfefficacy <= 3.41	0.169	3.03 < oppoertunitiesInfluence <= 3.43	0.08
trust <= 3.99	0.066	peerSupport <= 3.56	0.072
interest <= 3.33	0.042	3.99 < trust <= 4.44	0.065
equalTreatment <= 4.17	0.03	3.38 < easeParticipation <= 3.76	0.062
oppoertunitiesChoises <= 2.86	0.029	3.44 < competenceBeliefs <= 3.96	0.039
teacherSupport <= 4.17	0.029	3.41 < selfefficacy <= 4.03	0.032
oppoertunitiesInfluence <= 3.03	0.028	3.33 < interest <= 3.94	0.029
easeParticipation <= 3.38	0.021	4.17 < equalTreatment <= 4.61	0.026
CourseA4 <= 0.00	0.02	4.17 < teacherSupport <= 4.61	0.021
peerSupport <= 3.56	0.009	CourseA4 <= 0.00	-0.02
CourseA1 <= 0.00	-0.006	CourseA1 > 0.00	-0.018
CourseA2 <= 0.00	-0.005	2.86 < oppoertunitiesChoises <= 3.26	0.016
participation <= 2.81	0.001	CourseA2 <= 0.00	-0.007
0.00 < CourseA3 <= 1.00	0.0	CourseA3 <= 0.00	0.001
Profile PAP3		**Profile PAP4**	
rule	importance	rule	importance
3.43 < oppoertunitiesInfluence <= 3.79	0.078	peerSupport > 4.27	0.138
competenceBeliefs > 4.34	0.054	oppoertunitiesInfluence > 3.79	0.113
trust > 4.71	0.054	easeParticipation > 4.28	0.081
4.61 < equalTreatment <= 4.83	0.053	trust > 4.71	0.075
teacherSupport > 4.81	0.052	participation > 3.72	0.064
3.88 < peerSupport <= 4.27	0.045	competenceBeliefs > 4.34	0.03
interest > 4.35	0.044	interest > 4.35	0.028
3.76 < easeParticipation <= 4.28	0.031	oppoertunitiesChoises > 3.70	0.024
4.03 < selfefficacy <= 4.40	0.021	selfefficacy > 4.40	0.018
CourseA1 <= 0.00	-0.015	equalTreatment > 4.83	0.013
CourseA4 > 0.00	-0.014	CourseA4 > 0.00	0.011
2.81 < participation <= 3.34	0.012	teacherSupport > 4.81	0.005
oppoertunitiesChoises > 3.70	0.011	CourseA3 <= 0.00	0.005
CourseA3 <= 0.00	-0.01	CourseA1 <= 0.00	-0.001
CourseA2 <= 0.00	0.009	CourseA2 <= 0.00	-0.001

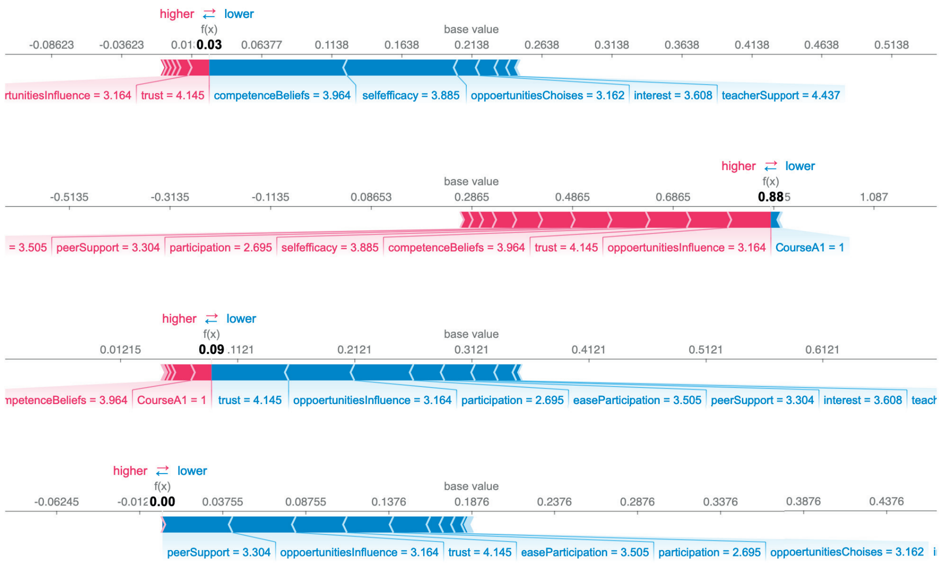


FIGURE 6. SHAP values explaining why the random forest model predicted an agency profile 2 student from the test set to be profile PAP2 and not profile 1, 3, or 4 (the bars are ordered by the profile number; i.e, the first bar predicts PAP1, the second PAP2, and so on). For each bar, the values explain how to get from the *base value* that would be predicted if no feature would be known to the current output for this particular profile 2 student. Feature values causing increased predictions are in red, and feature values decreasing the prediction are in blue. Their visual size shows the magnitude of the feature's effect.

benefit from more extensive encouragement as well as more attention and support in understanding the course contents (cf., [71]).

From the methodological level, our results showed that the complex nonlinear methods, especially the random forest, improved the accuracy of the predictive models. The

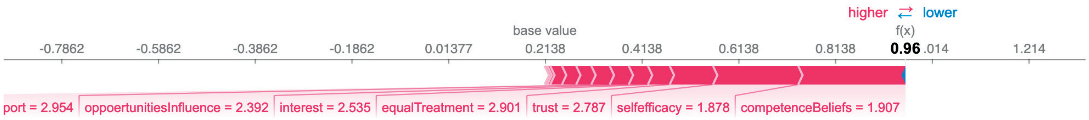


FIGURE 7. SHAP values explaining why the random forest model predicted an agency profile PAP1 student studying in course A3 from the test set to be profile PAP1 (true positive). The most important explanations are the low competence beliefs and self-efficacy values of this student.

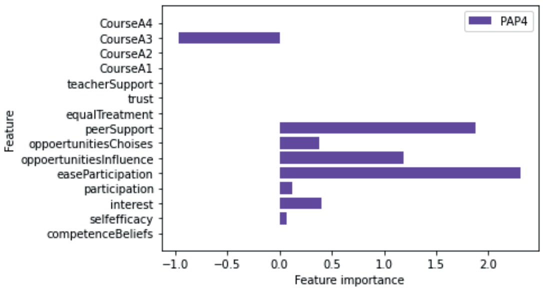


FIGURE 8. Coefficients of the multinomial logistic regression with l_1 penalization predicting the highest agency profile (PAP4). For seven features, the coefficient is zero, meaning they were irrelevant for this prediction model. A high value in all the picked features (except CourseA3) increases the probability that a student will be assigned to PAP4. However, if the student is in course A3, the probability that he/she will be assigned to PAP4 decreases.

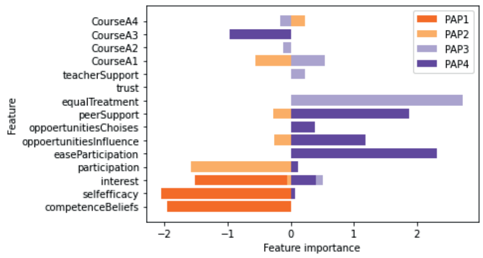


FIGURE 9. Coefficients of the multinomial logistic regression with l_1 penalization predicting agency profiles PAP1-PAP4. As a whole, the course features seem not as important as the agency dimensions but they are contributing. For example, if a student is in course A1, the probability that he/she will have the second highest agency (PAP3) increases.

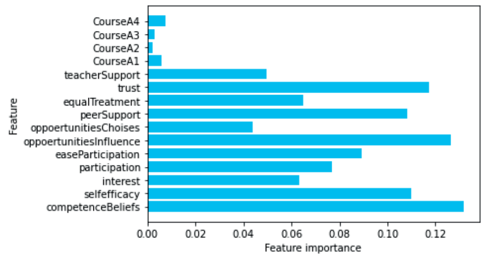


FIGURE 10. Feature importances of the random forest model predicting the agency profile. For the random forest, the agency dimensions are more important than the course features.

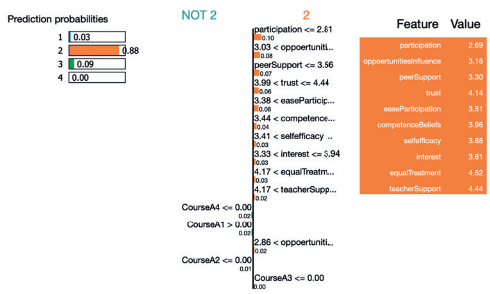


FIGURE 11. LIME rules explaining why the random forest model predicted an agency profile PAP2 student from the test set to be profile PAP2 (i.e., a true positive) with the highest probability. The most important local explanation why this student was assigned to this profile are his/her participation values.

traditional linear techniques performed worse but came with more informative global model-specific explanations. For example, while the global model-specific explanations from the random forest simply provided a ranking of the input features, the global model-specific explanations of the logistic regression with l_1 penalization also showed which feature was important for which class and which direction (i.e., whether it increased or decreased the probability for this class). Moreover, several features were dropped from the model, making it sparser and more interpretable.

Through recently developed model-agnostic XAI tools, we were able to also explain the better performing classifiers. LIME and SHAP can be used on top of any (complex) classifier to explain predictions for particular students (local explanations). These local explanations are very important, mainly for two reason. First, the GDPR now includes a right for explanation [57]. This means that if an automatic profiling is used in an LA tool, the student has a right to receive an explanation about his/her particular profiling.

Second, the local and global explanations can be different, and it is thus not enough to use the global explanations to explain why a particular student was assigned to a certain profile. For example, according to Figure 5, the most important agency dimensions for PAP2 (visual consideration of the lengths of the orange bars) are opportunities to influence, competence beliefs, and then trust for the teacher and self-efficacy. However, according to the LIME rules for that student in the test set who was assigned to PAP2 with the highest probability (Table 3), the order of importance concerning agency dimensions was participation activity, opportunities to influence, peer support, and then trust for the teacher.

In other words, the LIME rules (also those for the students that are representative for their PAP-profiles) do not always

TABLE 4. SHAP values (rounded to three decimals) for that student from the test set, who is in course A3 and was assigned to PAP1 with the highest probability by the random forest classifier. The table shows that the competence belief was the most important variable for the prediction.

variable	PAP1	PAP2	PAP3	PAP4
competenceBeliefs	0.222	-0.163	-0.042	-0.016
selfefficacy	0.176	-0.132	-0.03	-0.014
interest	0.04	-0.008	-0.024	-0.008
participation	0.006	0.034	-0.022	-0.017
easeParticipation	0.031	0.017	-0.025	-0.023
oppoertunitiesInfluence	0.039	0.036	-0.043	-0.032
oppoertunitiesChoises	0.029	-0.016	-0.008	-0.005
peerSupport	0.024	0.021	-0.013	-0.033
equalTreatment	0.045	-0.008	-0.031	-0.006
trust	0.089	-0.015	-0.049	-0.026
teacherSupport	0.038	-0.009	-0.025	-0.004
CourseA1	0.0	0.001	-0.002	0.0
CourseA2	-0.001	0.0	0.0	-0.0
CourseA3	0.0	0.0	0.001	-0.002
CourseA4	0.005	-0.005	0.001	-0.001

resemble the global explanations (Figure 5). For example, for the particular PAP2 student analyzed in Table 3, the value of participation activity was extremely low (2.69, see Figure 11), and the local surrogate model built by LIME to explain this prediction relied on this feature to a significant degree. This exemplifies the ''local fidelity'' of LIME: LIME explanations can be trusted only locally around the specific instance being explained. In contrast, the local SHAP explanations can—because of their additivity—be combined so that they can also be used to explain the global behavior of the model (Figure 5), being therefore more in line with the global model-specific explanations (Figure 10).

Naturally, our results are limited to the relatively small amount of data. Further data collection is required to increase the reliability of the observed connections between student agency and course implementations in higher education. In this paper, we have established the foundations for the use of XAI techniques in analyzing students' agency. Further work is required to examine, for example, the causal relationships of teaching practices and student agency.

V. CONCLUSION

Student agency is a key construct in the contemporary discourse about student-centered learning in higher education [3]–[5]. Jääskelä *et al.* [9] developed an LA process called student agency analytics (SAA), which utilizes a psychometric questionnaire instrument [35] and machine learning to provide information about the different resources of student agency. The recent literature on LA has highlighted the importance of explainability when utilizing complex models in education (e.g., [11], [14], [72]). In this study, we employed XAI techniques to derive more detailed information from student agency data. The purpose was to illustrate how the SAA process, combined with XAI techniques, could advance teachers' pedagogical awareness and reflection.

The purpose of the XAI techniques is to help to gain an understanding of how and why a model works. We used the multinomial logistic regression coefficients, feature importance levels of the random forest model, and combined SHAP values to explain the essential characteristics of the different agency profiles (global explanation). The prediction of the student profiles showed that the nonlinear techniques (especially random forest) modeled the data the best. The finding indicates that the relationships between the prototypical profiles of student agency and the teaching practices in higher education are relatively complex. Local explanations gave insight into why a student was assigned to a particular agency profile. Altogether, the XSAA results could be used to derive tentative explanations of the different experiences of student agency and to suggest ideas for pedagogical planning, as summarized in Section IV-A.

Educators at all levels of education need to take steps toward supporting student agency. To promote the educators' efforts, Moses *et al.* [4] called for connecting theory and practice and suggested increasing the research and practitioner-focused work about how teachers could support student agency. They emphasize that student agency ''is a practice-embedded construct that shapes the daily work of educators'' by involving them in reflecting the ways to create agentic spaces for students and making pedagogical decisions based on that reflection [4]. We see that this kind of teacher reflecting, pedagogical planning, and sharing of experiences of the agency-supporting practices among the colleagues could be facilitated using research-based tools and explainable SAA. These tools could help teachers to detect and understand the different experiences of student agency in their courses.

In summary, explainable models can provide more detailed and meaningful information about the different dimensions of student agency. By getting an overview of the different experiences of student agency in their courses, teachers could better meet the practical challenges of supporting student agency. Furthermore, higher education institutions could better adapt their capabilities to different learners' needs now and in the future. Thus, XSAA has the potential to contribute to teachers' pedagogical planning through the LA cycle.

REFERENCES

[1] A. Schleicher, ''The impact of COVID-19 on education insights from education at a glance 2020,'' Org. Econ. Co-Oper. Develop., Tech. Rep., 2020. [Online]. Available: https://www.oecd.org/education/the-impact-of-covid-19-on-education-insights-education-at-a-glance-2020.pdf

[2] M. Silverman, R. Sibbald, and S. Stranges, ''Ethics of COVID-19-related school closures,'' *Can. J. Public Health*, vol. 111, no. 4, pp. 462–465, Aug. 2020.

[3] P. Jääskelä, A.-M. Poikkeus, P. Häkkinen, K. Vasalampi, H. Rasku-Puttonen, and A. Tolvanen, ''Students' agency profiles in relation to student-perceived teaching practices in university courses,'' *Int. J. Educ. Res.*, vol. 103, Jan. 2020, Art. no. 101604.

[4] L. Moses, D. Rylak, T. Reader, C. Hertz, and M. Ogden, ''Educators' perspectives on supporting Student agency,'' *Theory Pract.*, vol. 59, no. 2, pp. 213–222, Apr. 2020.

[5] A. Schleicher, ''Concept note: OECD learning compass 2030,'' OECD, Paris, France, Tech. Rep., 2019.

[6] K. Collin, S. Lemmetty, S. Herranen, S. Paloniemi, T. Auvinen, and E. Riivari, ''Professional agency and creativity in information technology work,'' in *Agency at Work*. Cham, Switzerland: Springer, 2017, pp. 249–270.

[7] G. Siemens, "Learning analytics: The emergence of a discipline," *Amer. Behav. Scientist*, vol. 57, no. 10, pp. 1380–1400, 2013.
[8] J. Jovanović, M. Saqr, S. Joksimović, and D. Gašević, "Students matter the most in learning analytics: The effects of internal and instructional conditions in predicting academic success," *Comput. Educ.*, vol. 2021, Jun. 2021, Art. no. 104251.
[9] P. Jääskelä, V. Heilala, T. Kärkkäinen, and P. Häkkinen, "Student agency analytics: Learning analytics as a tool for analysing student agency in higher education," *Behav. Inf. Technol.*, vol. 40, pp. 1–19, Feb. 2020.
[10] G. Deeva, D. Bogdanova, E. Serral, M. Snoeck, and J. De Weerdt, "A review of automated feedback systems for learners: Classification framework, challenges and opportunities," *Comput. Educ.*, vol. 162, Mar. 2021, Art. no. 104094. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S036013152030292X
[11] R. S. Baker, "Challenges for the future of educational data mining: The baker learning analytics prizes," *J. Educ. Data Mining*, vol. 11, no. 1, pp. 1–17, 2019.
[12] V. Heilala, M. Saarela, P. Jaaskela, and T. Karkkainen, "Course satisfaction in engineering education through the lens of Student agency analytics," in *Proc. IEEE Frontiers Educ. Conf. (FIE)*, Oct. 2020, pp. 1–9.
[13] R. Al-Shabandar, A. Hussain, A. Laws, R. Keight, J. Lunn, and N. Radi, "Machine learning approaches to predict learning outcomes in massive open online courses," in *Proc. Int. Joint Conf. Neural Netw.*, May 2017, pp. 713–720.
[14] A. Dutt and M. A. Ismail, "Can we predict Student learning performance from LMS data? A classification approach," in *Proc. 3rd Int. Conf. Current Issues Educ. (ICCIE)*, 2019, pp. 24–29.
[15] A. P. Patil, K. Ganesan, and A. Kanavalli, "Effective deep learning model to predict Student grade point averages," in *Proc. IEEE Int. Conf. Comput. Intell. Comput. Res. (ICCIC)*, Dec. 2017, pp. 1–6.
[16] M. Stapel, Z. Zheng, and N. Pinkwart, "An ensemble method to predict student performance in an online math learning environment," presented at the 9th Int. Conf. Educ. Data Mining (EDM). Raleigh, NC, USA: Int. Educ. Data Mining Soc., Jun./Jul. 2016, pp. 231–238.
[17] U. Bhatt, A. Xiang, S. Sharma, A. Weller, A. Taly, Y. Jia, J. Ghosh, R. Puri, J. M. Moura, and P. Eckersley, "Explainable machine learning in deployment," in *Proc. Conf. Fairness, Accountability, Transparency*, 2020, pp. 648–657.
[18] E. Tjoa and C. Guan, "A survey on explainable artificial intelligence (XAI): Toward medical XAI," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Oct. 20, 2020, doi: 10.1109/TNNLS.2020.3027314.
[19] H. Aldowah, H. Al-Samarraie, and W. M. Fauzy, "Educational data mining and learning analytics for 21st century higher education: A review and synthesis," *Telematics Informat.*, vol. 37, pp. 13–49, Apr. 2019.
[20] D. Ifenthaler and J. Y.-K. Yau, "Utilising learning analytics to support study success in higher education: A systematic review," *Educ. Technol. Res. Develop.*, vol. 68, no. 4, pp. 1961–1990, Aug. 2020.
[21] O. Viberg, M. Hatakka, O. Bälter, and A. Mavroudi, "The current landscape of learning analytics in higher education," *Comput. Hum. Behav.*, vol. 89, pp. 98–110, Dec. 2018.
[22] X. Du, J. Yang, B. E. Shelton, J.-L. Hung, and M. Zhang, "A systematic meta-review and analysis of learning analytics research," *Behav. Inf. Technol.*, vol. 40, pp. 1–14, Sep. 2019.
[23] Y. Cui, F. Chen, A. Shiri, and Y. Fan, "Predictive analytic models of Student success in higher education: A review of methodology," *Inf. Learn. Sci.*, vol. 120, no. 3/4, pp. 208–227, Jan. 2019.
[24] A. L. Sønderlund, E. Hughes, and J. Smith, "The efficacy of learning analytics interventions in higher education: A systematic review," *Brit. J. Educ. Technol.*, vol. 31, p. 209, Nov. 2018.
[25] P. Leitner, M. Khalil, and M. Ebner, "Learning analytics in higher education—A literature review," in *Learning Analytics: Fundaments, Applications, and Trends: A View of the Current State of the Art to Enhance e-Learning*, A. Peña-Ayala, Ed. Cham, Switzerland: Springer, 2017, pp. 1–23.
[26] V. Putnam and C. Conati, "Exploring the need for explainable artificial intelligence (XAI) in intelligent tutoring systems (ITS)," in *Proc. IUI Workshops*, vol. 19, 2019, pp. 1–7.
[27] C. Conati, K. Porayska-Pomsta, and M. Mavrikis, "AI in education needs interpretable machine learning: Lessons from open learner modelling," 2018, *arXiv:1807.00154*. [Online]. Available: https://arxiv.org/abs/1807.00154
[28] J. M. Alonso and G. Casalino, "Explainable artificial intelligence for human-centric data analysis in virtual learning environments," in *Proc. Int. Workshop Higher Educ. Learn. Methodologies Technol.* Cham, Switzerland: Springer, 2019, pp. 125–138.
[29] A. Bandura, "Toward a psychology of human agency," *Perspect. Psychol. Sci.*, vol. 1, no. 2, pp. 164–180, Jun. 2006.
[30] M. S. Archer and M. S. Archer, *Structure, Agency and the Internal Conversation*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
[31] Y.-H. Su, "The constitution of agency in developing lifelong learning ability: The 'being' mode," *Higher Educ.*, vol. 62, no. 4, pp. 399–412, Oct. 2011.
[32] S. Hoidn and K. Reusser, "Foundations of student-centered learning and teaching," in *The Routledge International Handbook of Student-Centered Learning and Teaching in Higher Education*, S. Hoidn and M. Klemenčič, Eds. Abingdon-on-Thames, U.K.: Routledge, 2020, pp. 17–46.
[33] J. Martin, "Self-regulated learning, social cognitive theory, and agency," *Educ. Psychologist*, vol. 39, no. 2, pp. 135–145, Jun. 2004.
[34] L. Starkey, "Three dimensions of student-centred education: A framework for policy and practice," *Crit. Stud. Educ.*, vol. 60, pp. 1–16, Jan. 2017.
[35] P. Jääskelä, A.-M. Poikkeus, K. Vasalampi, U. M. Valleala, and H. Rasku-Puttonen, "Assessing agency of university students: Validation of the AUS scale," *Stud. Higher Educ.*, vol. 42, no. 11, pp. 1–19, 2017.
[36] J. Santos, A. S. Figueiredo, and M. Vieira, "Innovative pedagogical practices in higher education: An integrative literature review," *Nurse Educ. Today*, vol. 72, pp. 12–17, Jan. 2019.
[37] L.-M. Hero and E. Lindfors, "Students-learning experience in a multidisciplinary innovation project," *Educ. training*, vol. 61, no. 4, pp. 500–522, 2019.
[38] P. Garnjost and L. Lawter, "Undergraduates' satisfaction and perceptions of learning outcomes across teacher- and learner-focused pedagogies," *Int. J. Manage. Educ.*, vol. 17, no. 2, pp. 267–275, Jul. 2019.
[39] P. Koskinen, J. Lämsä, J. Maunuksela, R. Hämäläinen, and J. Viiri, "Primetime learning: Collaborative and technology-enhanced studying with genuine teacher presence," *Int. J. STEM Educ.*, vol. 5, no. 1, p. 20, Dec. 2018.
[40] M. J. M. Leenknecht, L. Wijnia, S. M. M. Loyens, and R. M. J. P. Rikers, "Need-supportive teaching in higher education: Configurations of autonomy support, structure, and involvement," *Teaching Teacher Educ.*, vol. 68, pp. 134–142, Nov. 2017.
[41] M. Räisänen, L. Postareff, and S. Lindblom-Ylänne, "Students' experiences of study-related exhaustion, regulation of learning, peer learning and peer support during university studies," *Eur. J. Psychol. Educ.*, vol. 2020, pp. 1–23, Nov. 2020.
[42] T. Hailikari, N. Katajavuori, and S. Lindblom-Ylanne, "The relevance of prior knowledge in learning and instructional design," *Amer. J. Pharmaceutical Educ.*, vol. 72, no. 5, p. 113, Sep. 2008.
[43] Y.-S. Tsai, D. Rates, P. M. Moreno-Marcos, P. J. Muñoz-Merino, I. Jivet, M. Scheffel, H. Drachsler, C. D. Kloos, and D. Gašević, "Learning analytics in European higher education—Trends and barriers," *Comput. Educ.*, vol. 155, Oct. 2020, Art. no. 103933.
[44] M. A. Chatti, A. L. Dyckhoff, U. Schroeder, and H. Thüs, "A reference model for learning analytics," *Int. J. Technol. Enhanced Learn.*, vol. 4, nos. 5–6, pp. 318–331, Jan. 2012.
[45] C. P. Rosé, "Learning analytics in the learning sciences," in *International Handbook of the Learning Sciences*. Oxfordshire, U.K.: Routledge, 2018, pp. 511–519.
[46] M. Saarela, *Automatic Knowledge Discovery From Sparse and Large-Scale Educational Data: Case Finland*, no. 262. Jyväskylä, Finland: Univ. Jyväskylä, 2017.
[47] P. Blikstein, "Multimodal learning analytics," in *Proc. 3rd Int. Conf. Learn. Anal. Knowl. (LAK)*, 2013, pp. 102–106.
[48] R. Luckin and M. Cukurova, "Designing educational technologies in the age of AI: A learning sciences-driven approach," *Brit. J. Educ. Technol.*, vol. 50, no. 6, pp. 2824–2838, Nov. 2019.
[49] T. Bolukbasi, K.-W. Chang, J. Y. Zou, V. Saligrama, and A. T. Kalai, "Man is to computer programmer as woman is to homemaker? Debiasing word embeddings," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 4349–4357.
[50] A. A. Freitas, "Comprehensible classification models: A position paper," *ACM SIGKDD Explor. Newslett.*, vol. 15, no. 1, pp. 1–10, 2013.
[51] B. J. Dietvorst, J. P. Simmons, and C. Massey, "Algorithm aversion: People erroneously avoid algorithms after seeing them err," *J. Exp. Psychol., Gen.*, vol. 144, no. 1, p. 114, 2015.
[52] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018.

[53] L. H. Gilpin, D. Bau, B. Z. Yuan, A. Bajwa, M. Specter, and L. Kagal, "Explaining explanations: An overview of interpretability of machine learning," in *Proc. IEEE 5th Int. Conf. Data Sci. Adv. Anal. (DSAA)*, Oct. 2018, pp. 80–89.
[54] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, "A survey of methods for explaining black box models," *ACM Comput. Surv.*, vol. 51, no. 5, pp. 1–42, 2018.
[55] S. Lapuschkin, S. Wäldchen, A. Binder, G. Montavon, W. Samek, and K.-R. Müller, "Unmasking clever Hans predictors and assessing what machines really learn," *Nature Commun.*, vol. 10, no. 1, pp. 1–8, Dec. 2019.
[56] S. T. Mueller, R. R. Hoffman, W. Clancey, A. Emrey, and G. Klein, "Explanation in human-AI systems: A literature meta-review, synopsis of key ideas and publications, and bibliography for explainable AI," 2019, *arXiv:1902.01876*. [Online]. Available: https://arxiv.org/abs/1902.01876
[57] B. Goodman and S. Flaxman, "European Union regulations on algorithmic decision-making and a 'right to explanation,'" *AI Mag.*, vol. 38, no. 3, pp. 50–57, 2017.
[58] S. Wachter, B. Mittelstadt, and L. Floridi, "Transparent, explainable, and accountable AI for robotics," *Sci. Robot.*, vol. 2, no. 6, May 2017, Art. no. eaan6080.
[59] C. Molnar, *Interpretable Machine Learning*. Victoria, BC, Canada: Lean Publishing, 2019.
[60] M. Saarela and T. Kärkkäinen, "Can we automate expert-based journal rankings? Analysis of the Finnish publication indicator," *J. Informetrics*, vol. 14, no. 2, May 2020, Art. no. 101008.
[61] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY, USA: Springer, 2009.
[62] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc., B Methodol.*, vol. 58, no. 1, pp. 267–288, 1996.
[63] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Inf. Fusion*, vol. 58, pp. 82–115, Jun. 2020.
[64] A. J. London, "Artificial intelligence and black-box medical decisions: Accuracy versus explainability," *Hastings Center Rep.*, vol. 49, no. 1, pp. 15–21, Jan. 2019.
[65] Z. C. Lipton, "The mythos of model interpretability," *Queue*, vol. 16, no. 3, pp. 31–57, 2018.
[66] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee, "From local explanations to global understanding with explainable ai for trees," *Nature Mach. Intell.*, vol. 2, no. 1, pp. 2522–5839, 2020.
[67] M. Saarela, O.-P. Ryynänen, and S. Äyrämö, "Predicting hospital associated disability from imbalanced data using supervised learning," *Artif. Intell. Med.*, vol. 95, pp. 88–95, Apr. 2019.
[68] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should i trust you?': Explaining the predictions of any classifier," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 1135–1144.
[69] L. S. Shapley, "A value for n-person games," *Contrib. Theory Games*, vol. 2, no. 28, pp. 307–317, 1953.
[70] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 4765–4774.
[71] V. Heilala, P. Jääskelä, T. Kärkkäinen, and M. Saarela, "Understanding the study experiences of students in low agency profile: Towards a smart education approach," in *Proc. Int. Conf. Smart Inf. Commun. Technol.* Cham, Switzerland: Springer, 2019, pp. 498–508.
[72] C. Romero and S. Ventura, "Educational data mining and learning analytics: An updated survey," *WIREs Data Mining Knowl. Discovery*, vol. 10, no. 3, p. e1355, May 2020.

MIRKA SAARELA received the degree in computer science from the University of Passau, Germany, and the Ph.D. degree in mathematical information technology from the University of Jyväskylä, Finland. She is currently a Postdoctoral Research Fellow with the Faculty of Information Technology, University of Jyväskylä. Her research interests combine machine learning, explainable artificial intelligence, cognitive computing, and education. She was awarded several grants for her research, amongst others, by the Ulla Tuomisen Foundation, the Savings Bank, and the Ella and Georg Ehrnroot Foundation.

VILLE HEILALA received the master's degree in education and the master's degree in computer science. He is currently pursuing the Ph.D. degree with the Faculty of Information Technology, University of Jyväskylä. He is currently a Researcher with the Faculty of Information Technology, University of Jyväskylä. His research interests include learning analytics and student agency.

PÄIVIKKI JÄÄSKELÄ received the Ph.D. degree in education. She is currently an Adjunct Professor. She is a Senior Researcher with the Finnish Institute for Educational Research, University of Jyväskylä, Finland. She is the main Developer of a validated scale with the Agency of University Students (AUS), that assesses university students' agency within the multidimensional framework which takes into account the interplay of individual, relational, and participatory aspects of agency. She has worked several years as a Responsible Researcher in various university-level projects on teacher development and the development of pedagogy of university studies. She has extensive experience as a Teacher in educational sciences and a Teacher Educator at the University of Jyväskylä. Her research interests include university student agency, learner-centered pedagogy, and teacher development in higher education.

ANNE RANTAKAULIO has been a Teacher in mathematics with the Department of Technology and Transport, JAMK University of Applied Sciences, since 1986. As the teaching tools have evolved and e-learning has increased, she has become particularly interested in what learning really is and how it can be best supported.

TOMMI KÄRKKÄINEN (Senior Member, IEEE) received the Ph.D. degree in mathematical information technology from the University of Jyväskylä (JYU), in 1995. Since 2002, he has been a Full Professor of mathematical information technology with the Faculty of Information Technology (FIT), JYU. He has led almost 50 different research and development projects and involved in supervising 60 Ph.D. students. He has published around 190 peer-reviewed articles. He has served in many administrative positions at FIT and JYU, where he is currently leading a Research Division and a Research Group on Human and Machine-Based Intelligence in Learning. His main research interests include data mining, machine learning, learning analytics, and nanotechnology. He received the Innovation Prize from JYU, in 2010.

PV

"SITTING AT THE STERN AND HOLDING THE RUDDER": TEACHERS' REFLECTION ON ACTION BASED ON STUDENT AGENCY ANALYTICS IN HIGHER EDUCATION

by

Heilala, V., Jääskelä, P., Saarela, M., Kuula., A-S., Eskola, A., and Kärkkäinen, T
forthcoming

In Leonid Chechurin (Ed.). Digital Teaching and Learning in Higher Education Developing and Disseminating Skills for Blended Learning, London: Palgrave Macmillan

Request a copy from the author.

PVI

THE FINNISH VERSION OF THE AFFINITY FOR TECHNOLOGY INTERACTION SCALE (ATI): PSYCHOMETRIC PROPERTIES AND AN EXAMINATION OF GENDER DIFFERENCES

by

Heilala, V., Kelly, R., Saarela, M., Jääskelä, P., and Kärkkäinen, T 2022

International Journal of Human—Computer Interaction

INTERNATIONAL JOURNAL OF HUMAN-COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2022.2049142

The Finnish Version of the Affinity for Technology Interaction (ATI) Scale: Psychometric Properties and an Examination of Gender Differences

Ville Heilala[a], Riitta Kelly[b], Mirka Saarela[a], Päivikki Jääskelä[c], and Tommi Kärkkäinen[a]

[a]Faculty of Information Technology, University of Jyväskylä, Jyvaskyla, Finland; [b]The Centre for Multilingual Academic Communication, University of Jyväskylä, Jyvaskyla, Finland; [c]Finnish Institute for Educational Research, University of Jyväskylä, Jyvaskyla, Finland

ABSTRACT

The pervasiveness of technical systems in our lives calls for a broad understanding of the interaction between humans and technology. Affinity for technology interaction (ATI) scale measures the tendency of a person to actively engage or to avoid interaction with technological systems, including both software and physical devices. This research presents a psychometric analysis of a Finnish version of the ATI scale. The data consisted of 796 responses of students in a Finnish university. The data were analyzed utilizing factor analysis and both nonparametric and parametric item response theory. The Finnish version of the ATI scale proved to be essentially unidimensional, showing high reliability estimates, and forming a strong Mokken scale. Hierarchical multiple regression analysis showed that men had a slightly higher affinity for technology than women when controlling for age and field of study; however the effect size was small.

1. Introduction

Urban legend or not, the famous quote from the 1950s allegedly attributed to Thomas J. Watson, a long-time chairperson and the CEO of International Business Machines, claimed that there would be market potential for only five electronic computers (IBM, 2007). The future turned out to be different, and today we live amid ubiquitous technical systems. The pervasiveness of different technological systems stretches out to many fields of life, including work (e.g., van Laar et al., 2017), education (e.g., Kim, Merrill, Xu, & Sellnow, 2021), sports (e.g., Cranmer et al., 2021), health (e.g., Tajudeen et al., 2021), culture (e.g., Kim & Lee, 2022), virtual life (e.g., Taufik et al., 2021), and even afterlife (Beaunoyer & Guitton, 2021). Thus, technological development poses several challenges which call for multidisciplinary, international and global research efforts (Stephanidis et al., 2019).

Scholars have identified mixed effects relating to how information and communication technologies affect our lives (Ali et al., 2020). Many effects are reciprocal and mediated by personality traits and other individual differences. For example, individual differences can affect self-disclosure in social media (Chen et al., 2015), usability assessment (Kortum & Oswald, 2018), gaming (Caci et al., 2019), online learning (Alabdullatif & Velázquez-Iturbide, 2020), and online privacy literacy and behavior (Sindermann et al., 2021). Furthermore, meta-reviews have documented gender differences relating to attitudes towards technology (Cai et al., 2017; Whitley, 1997): in general, men tended to have slightly more favorable attitudes towards technology. Information technology can facilitate personality research especially from the idiographic point of view (Matthews et al., 2021; Montag & Elhai, 2019) and, as Matthews et al. (2021) point out, socio-technological change might give rise to the evolution of the contemporary trait models in the future society. Thus, it is important to have valid constructs and psychometric instruments to discern individual differences and understand underlying phenomena relating to interactions between humans and technology.

A general concept for depicting the relationship between humans and technology is the person's affinity with technological systems and devices. Edison and Geissler (2003) consider affinity for technology as an attitude and a "positive affect towards technology (in general)." Franke et al. (2019) define affinity for technology interaction as a question of "whether users tend to actively approach interaction with technical systems or, rather, tend to avoid intensive interaction with new systems." In terms of technology interaction, it can be viewed as a key personal resource, and as such, it is of great importance considering the interaction between the user and technology.

One promising scale to assess human and technology interaction is the affinity for technology interaction (ATI) scale[1] developed by Franke et al. (2019). The scale was initially developed in English and German, and it is currently available as translations also in Italian, Spanish, Romanian,

CONTACT Ville Heilala ville@heila.la, ville.s.heilala@jyu.f Faculty of Information Technology, University of Jyväskylä, Jyvaskyla, Finland

Table 1. The Finnish translation of the ATI scale.

Item	Finnish translation
ATI1	Perehdyn mielelläni teknisiin järjestelmiin yksityiskohtaisesti.
ATI2	Pidän uusien teknisten järjestelmien toimintojen kokeilemisesta.
ATI3	Olen tekemisissä teknisten järjestelmien kanssa lähinnä siksi, koska minun täytyy.
ATI4	Kun kohtaan uuden teknisen järjestelmän, kokeilen sitä innokkaasti ja perusteellisesti.
ATI5	Käytän mielelläni aikaa uuteen tekniseen järjestelmään tutustumiseen.
ATI6	Minulle riittää, että tekninen järjestelmä toimii, mutta minulle on samantekevää miten tai miksi.
ATI7	Pyrin ymmärtämään, miten tekninen järjestelmä tarkalleen ottaen toimii.
ATI8	Minulle riittää, että tunnen teknisen järjestelmän perustoiminnot.
ATI9	Pyrin hyödyntämään teknisen järjestelmän kaikkia ominaisuuksia.

Note. ATI3, ATI6, and ATI8 are reversed-worded items. Translated Likert categories and introductory text in Appendix A.

Persian, and Dutch. However, to our knowledge, besides the original English and German versions, no published analyses of the psychometric properties of the scale exist for other languages. In terms of cross-cultural research, confirmation of translations of scales plays a crucial role, especially when the scales are constructed to measure some universal constructs or phenomena (Cha et al., 2007).

This research presents a psychometric analysis of the Finnish version of the ATI scale. We elaborate on the subject by examining the gender differences relating to ATI. In other words, we address the validity evidence concerning the scale's internal structure and it's ability to capture differences and similarities (AERA, APA, & NCME, 2014). We use a comprehensive analytical process utilizing methodological triangulation and multiple sources of information. By presenting a psychometric analysis of the Finnish version of the scale, we aim to provide added value to the original research-based model of using ATI scale in measuring individual differences in affinity for technology interaction.

2. Materials and methods

2.1. Affinity for technology interaction (ATI) scale (Franke *et al., 2019*)

The starting point of the definition of ATI mentioned earlier is a realization that affinity for technology interaction and need for cognition (NFC) are closely related; following Schmettow and Drees (2014), Franke et al. (2019) propose that the two "should be conceptualized in a close relationship". Relying on, for example, Cacioppo and Petty (1982), Cacioppo et al. (1996) and Fleischhauer et al. (2010), they note that NFC can be seen today as "the inter-individually varying, stable intrinsic motivation to engage in cognitively challenging tasks". Given that NFC can be applied in different psychological domains, they see developing ATI in line with the construct of NFC as useful.

The purpose of Franke et al. (2019) was both to develop and validate a new scale to be able to assess ATI. Their goal was to provide "a highly economical and reliable unidimensional scale that is suitable for differentiating between users across the whole range of the ATI trait," keeping in mind the focus that ATI has as a general interaction style in relation to technology (Franke et al., 2019). As a result, their ATI scale is a unidimensional 9-item scale having 3 reverse-worded (RW) items. All items are measured using a 6-point Likert scale. The shorter version of the scale (ATI-S) consisting of a subset of 4 items is currently available in German and in English (Wessel et al., 2019).

Franke et al. (2019) summarized the results of their validation process of ATI scale using multiple studies ($N > 1500$) as follows: first of all, the factor analyses indicated unidimensionality of the ATI scale. Secondly, their analysis revealed that reliability estimated using coefficient α ranged between good and excellent. Thirdly, when it comes to the need for cognition, geekism, technology enthusiasm, computer anxiety, control beliefs in dealing with technology, success in technical problem-solving and technical system learning, technical system usage, and the personality dimensions linked to Big Five, the expected relationships were supported by construct validity analysis. Fourth, when considering the ability of the ATI scale to differentiate between higher- and lower-ATI participants, item analysis and descriptive statistics showed that this was possible. Fifth, when taking into account analyses of demographic variables, the gender effect turned out to be large, the age effect small, and the educational background had no effect at all. Thus, the results showed that it could be possible to "discriminate between participants based on their differing tendency to actively engage in intensive (i.e., cognitively demanding) technology interaction" (Franke et al., 2019). The ATI scale has been used in varied contexts. These include studies on partially automated vehicles (e.g., Boelhouwer et al., 2020; Schartmüller et al., 2019), automated decision-making in health care (Schlicker et al., 2021), use of information technology among primary care physician trainees (Wensing et al., 2019), privacy concerns in users' acceptance of e-Health technologies (Schomakers et al., 2019) and activity tracker usage (e.g., Attig & Franke, 2019), as well as augmented reality (Kammler et al., 2019).

2.2. Translation protocol

The translation of the original scale to Finnish was conducted as a forward-backward translation utilizing a committee approach (Brislin et al., 1973, pp. 46–47). The translation process, in general, followed the protocol proposed by Sousa and Rojjanasrirat (2011), excluding the pilot testing (Table 1). Two independent professional translators conducted the forward translations from the original English version of the scale. The first and second authors constructed the initial Finnish version based on the two professional translations. While the translations were conducted using the original English version of the scale, we also used the original German version of the ATI scale for creating the Finnish translation. A native German speaker and a German language teacher evaluated the connotations of a few essential wordings between the initial Finnish version and the original German version of the scale to achieve consistency between both original versions. After a few minor refinements, the initial translated version was back-translated to English by two independent professional translators. The back-translated versions proved excellent similarity with the original English version of the scale. The exact word–by–word equivalences of the back-translated versions compared to the original scale version were 70% and 76%, and when considering synonyms 77% and 85% respectively. Thus, the final translated Finnish version (Table 1) was chosen to be used in the primary data collection. The translations of the introductory text and Likert categories are presented in the Appendix A.

2.3. Data collection and participants

We used a non-probabilistic convenience sample of students ($N = 796$) studying in a Finnish public multidisciplinary research university (ISCED 2011 level 6–8). The data were collected using an online questionnaire. The link to the questionnaire was sent through student email lists in six faculties or departments. The questionnaire contained a privacy statement complying with GDPR, and informed consent was obtained from the participants of the research. The participants had the opportunity to participate in a raffle to win one of 10 gift cards worth 20 euros each. Demographic information (i.e., age and gender) were asked, including information about the faculty where the respondent was studying. Gender was asked using a single-item open-ended question because "it allows respondents to define their own gender using whatever terminology they choose" (Cameron & Stinson, 2019).

Respondents' ages ranged from 17 to 73 years ($Mdn = 25, M = 27.6, SD = 8.9$). An open-ended form field was used to ask gender and 519 (65%) of the respondents identified themselves as women, 264 (33%) as men, 7 (1%) as nonbinary, and 6 (1%) that were unknown (i.e., preferred not to answer or the answer was uninterpretable) were coded as missing values. Respondents represented six university faculties or departments: information technology (28% of the total respondents, 47% women in the subsample), natural sciences (22%, 67% women), humanities and social sciences (14%, 74% women), sport and health sciences (14%, 76% women), education (11%, 92% women), business school (10%, 66% women), and other units (1%). There were eight responses with missing values relating to age or gender.Compared to the general population, the sample is limited by age and educational background as it consists of relatively young people pursuing university studies. The use of a convenience sample is exploratory in nature, and it limits the generalizability of the results with respect to the general population, which is discussed in more detail in the study limitations in Section 4.1.

2.4. Psychometric protocol

Our analysis process consisted of four main phases: (i) describing the data using common descriptive statistics, (ii) utilizing the non-parametric item response theory by conducting a Mokken scale analysis, (iii) conducting factor analysis based on a classical test theory, and (iv) utilizing the parametric item response theory by conducting a complemental analysis based on partial credit model. A similar approach excluding the parametric item response theory-based analysis has been conducted for the part of the original ATI scale data (c.f., Lezhnina & Kismihók, 2020). Furthermore, we applied the scale using hierarchical multiple regression analysis to examine the gender differences concerning the affinity for technology interaction. All analyses were conducted using *R* version 4.0.3 (R Core Team, 2020) and packages mentioned later in the methods section.

Our analytical process was as follows (Figure 1). We began with descriptive statistics and examined whether the data contained any peculiarities (e.g., excess skew or kurtosis, ceiling or flooring effects, categories without responses). Subsequently, we continued with Mokken scale analysis. It is a convenient nonparametric method making no assumptions about the data distribution while providing an initial assessment of the important scale properties. Mokken scale analysis addresses whether the total score of the scale can be used to order persons with respect to the measured construct. A scale forming a Mokken scale would be a promising candidate for further analysis.

Next, we used factor analyses for evaluating the structural properties of the scale in detail. The assessment of dimensionality is critical, and it "requires informed judgment that balances statistical information with conceptual plausibility and utility" (Fabrigar & Wegener, 2012, p. 148). The expected number of dimensions is naturally determined in the original scale validation research for an existing scale. However, a new translation and a new cultural context necessitate a new dimensionality assessment.

After the Mokken scale analysis and the factor analysis, we scrutinized the scale further by utilizing the parametric item response theory. As the ability to order persons in their latent variable is an important feature of a scale, we used the partial credit model (PCM) for the analysis because it is the least restrictive parametric IRT model still possessing a more accurate property of stochastic ordering of the latent variable by the total scale score (i.e., SOL by X_+) (Hemker et al., 1997; Ligtvoet, 2012; van der Ark, 2005). Furthermore, the PCM analysis provided information at the item level. For conciseness, the complemental analysis based on parametric item response theory using PCM is presented in Appendix C.

2.4.1. Descriptive statistics and multivariate outliers

The data were first examined using basic descriptive statistics. Nonnormality is usual in the case of real-world psychological and educational data (Cain et al., 2017). Mardia's test for multivariate skewness and kurtosis (Mardia, 1970; Mecklin & Mundfrom, 2007) was used to assess whether the data complies with the multivariate normal distribution

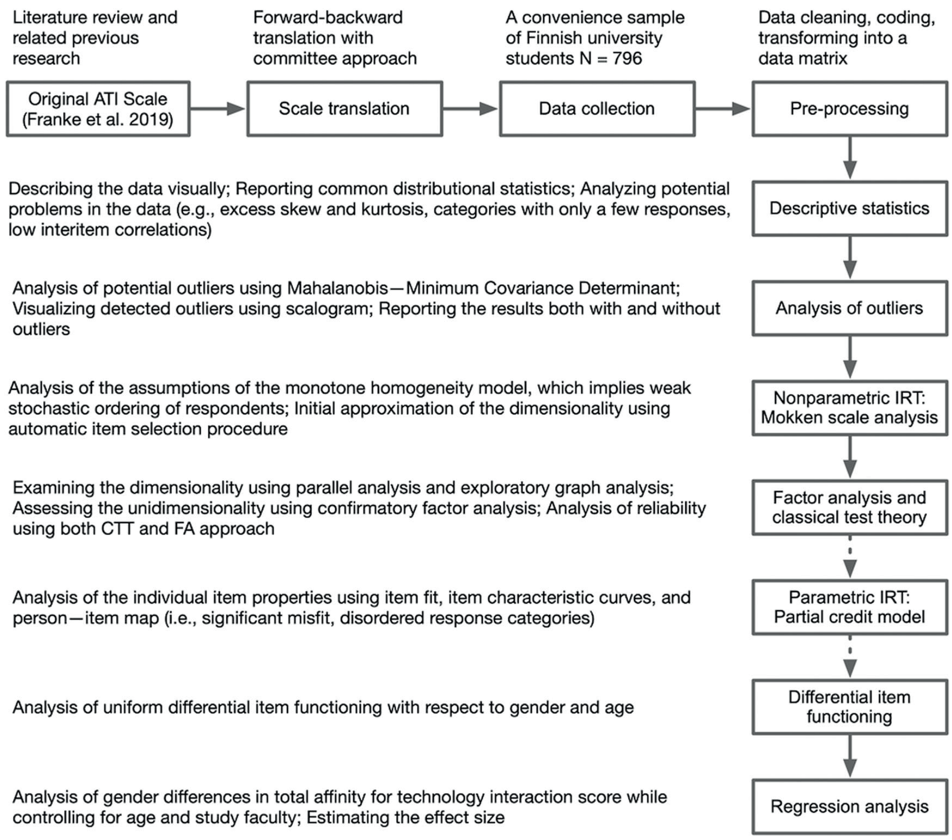


Figure 1. An overview of the analytics scheme.

(McDonald & Ho, 2002). A significant result in Mardia's test indicates that the data were not complying with multivariate normal distribution. Univariate skewness and kurtosis were assessed using Fisher's skewness (G_1) and kurtosis (G_2) (Cain et al., 2017; Joanes & Gill, 1998). A scalogram was used to describe the individual response patterns visually (e.g., Massof, 2004).

To detect possible multivariate outliers, we used Mahalanobis–Minimum Covariance Determinant with a breakdown point of 0.25 (MMCD75). As a robust version of the traditional Mahalanobis distance, MMCD75 was suggested to be efficient in detecting outlying values as well as having an acceptable false detection rate (Leys et al., 2018). We analyzed the data with and without the outliers. For transparency, both results are reported when the difference is deemed to be more than negligible or the result is otherwise crucial for assessing the effect of outliers in the data (e.g., measures of association). All possible outliers identified by the MMCD75 method were depicted visually using a scalogram.

2.4.2. Nonparametric item response theory

Models belonging to the field of nonparametric item response theory (NIRT) are data-driven exploratory models, which assume that the relationship between the latent variable and the item score is restricted only by order (Sijtsma, 2005). One such model is the monotone homogeneity model (MHM) for dichotomous (Mokken, 1971, 1997) and polytomous (Molenaar, 1997) data. MHM is also known as the nonparametric graded response model (np-GRM) (Sijtsma & van der Ark, 2020, p. 233) and it is the most general of all well-known IRT models for polytomous data (Hemker et al., 2001; van der Ark & Bergsma, 2010). The general MHM has some desirable psychometric properties in terms of total scale score. Thus, in this paper, we first start our analysis by assessing the applicability of MHM to the collected scale data.

MHM is based on three key assumptions: i) unidimensionality, which means that all items are measuring the same latent variable, ii) local independence, which means that the item scores depend only on the person's latent variable; and iii) latent monotonicity of the item step response functions (ISRFs), which means that the functions are nondecreasing concerning the latent variable (Sijtsma, 2005; Sijtsma & van der Ark, 2017; van der Ark, 2012). A more strict model, the double monotonicity model (DMM), also assumes invariant item ordering, which means that the scale items can be placed in order with respect to the latent variable (Sijtsma & van der Ark, 2017). Mokken scale analysis (MSA) (Sijtsma et al., 2011; Sijtsma & van der Ark, 2017)) is a set of tools, which can be used to analyze how dichotomous or polytomous scale data meet the assumptions of MHM and DMM. Scale identification in MSA involves examining the applicability of the assumptions in the data, in other words, assessing scalability, local independence, and invariant item ordering of the scale items in addition to the monotonicity of the ISRFs.

Scalability in MSA is based on the coefficient of homogeneity H (Loevinger, 1948; Mokken, 1971) also called as the scalability coefficient (Sijtsma & van der Ark, 2017). Existing scales can be evaluated directly using the inter-item coefficients H_{jk}, coefficients of the individual items H_j, and the overall coefficient H for the whole scale (Mokken, 1971, 1997). Higher H_j implies better item discrimination and values close to 0 do not discriminate well in terms of the latent variable (Sijtsma & van der Ark, 2017; Straat et al., 2014). Thus, a common approach to decide whether to include items to a scale is to define a threshold value c so that all $H_j > c$. The lowest threshold value traditionally used for considering the inclusion of an item to the scale is $H_j > 0.30$

for all items and the excluded items are considered as unscalable (Mokken, 1971; Sijtsma & van der Ark, 2017). For classifying complete scales, $0.30 \leq H < 0.40$ forms a weak Mokken scale, $0.40 \leq H < 0.50$ forms a medium scale, and $H \geq 0.50$ forms a strong scale (Mokken, 1971, p. 185).

Instead of relying on arbitrary threshold values, an automated item selection procedure (AISP) provides a way to examine the scale items' scalability and dimensionality. AISP is an iterative process, which aims to select items from the initial item bank so that (i) the selected item has a positive covariance with each of the already selected items, (ii) the item has $H_j \geq c$, and (iii) the selected item maximizes the overall H value of the scale with other selected items (Hemker et al., 1995). Instead of selecting a single threshold value c, one suggested approach is to run AISP for a sequence of thresholds (e.g., $c = \{0.05, 0.10, 0.15, \ldots, 0.60\}$) (Hemker et al., 1995; Sijtsma & van der Ark, 2017). The examinations of the sequential outcome pattern of AISP can reveal whether the data form one or more scales and whether some items turn out to be unscalable at a certain level of c (Hemker et al., 1995). Two different procedures have been proposed for AISP, Mokken's procedure (Mokken, 1971, p. 191) and a genetic algorithm (Straat et al., 2013). The two algorithms might yield different results for the same data (Sijtsma & van der Ark, 2017). The minimum sample size for using the AISP procedure depends on the item quality, but at least 250 to 500 responses are needed (Straat et al., 2014).

The scale items' local independence can be assessed using a procedure based on conditional association (CA). The procedure CA flags items as locally dependent and removes them one by one based on conditional covariances, indices $W^{(1)}$, $W^{(2)}$, and $W^{(3)}$, to identify a locally independent item set (Straat et al., 2016). An item or an item pair is flagged as locally dependent if $W > Q_3 + 1.5 * IQR$, where Q_3 is the third quartile, and IQR is the interquartile range of the empirical W distribution (Straat et al., 2016) (i.e., W is outside of Tukey's upper inner fence (Tukey, 1977, p. 44)). In this study, we utilized the procedure CA implemented in the *mokken* package (van der Ark, 2012).

Latent monotonicity means that the item step response function is a nondecreasing function with respect to the latent variable (van der Ark, 2012). In other words, the higher the person's ability on the latent variable, the higher the probability of scoring cases typical of the higher attribute level (Sijtsma & van der Ark, 2017). Manifest monotonicity—a property observed from the scale data—can be used to assess latent monotonicity using a procedure implemented in the *R* package *mokken* (van der Ark, 2007, 2012). The procedure combines respondents to rest score groups based on a selected minimum group size criterion *minsize*. Manifest monotonicity is assessed based on the probability of belonging to a higher rest score group with respect to a higher latent variable, and violations exceeding a minimum value *minvi* are considered relevant. For the data in this study, $minsize = N/10$ and $minvi = 0.30$ were used (van der Ark, 2007).

2.4.3. Factor analysis

The first step in factor analysis is to assess the dimensionality of the data and decide how many factors to retain. A suggested approach is to use multiple methods to assess the dimensionality of the data and compare their results (Lubbe, 2019). To assess the dimensionality and the number of factors to retain, we used parallel analysis (PA) and minimum average partials (MAP). The parallel analysis compares the structure in the collected data to a structure of randomly sampled data. The number of dimensions in the actual data exceeding the number of dimensions on the random data is retained. PA is often referred to as one of the most accurate and robust rules for determining the dimensionality of the data (Lubbe, 2019), and it performs well in a wide variety of scenarios (e.g., Golino et al., 2020). PA with PCA extraction (PA-PCA, a.k.a., Horn's PA (Horn, 1965)) using polychoric correlation has been suggested to be suitable for all types of data (Garrido et al., 2013). For PA-PCA, we used a non-parametric version of parallel analysis with column permutation (500 random data sets), polychoric correlation, and quantile thresholds 50% (median, PA-PCA-m) and 95% (PA-PCA-95) (Auerswald & Moshagen, 2019; Buja & Eyuboglu, 1992).

Another promising and recent approach for analyzing the dimensions of psychological constructs is exploratory graph analysis (EGA). EGA draws from the methods behind network psychometrics, which in turn aims to combine different latent variable models and network models (Epskamp et al., 2017, 2018). EGA utilizes partial correlations and the Gaussian graphical model with a clustering algorithm for a weighted network (i.e., Walktrap algorithm) (Golino et al., 2020). EGA is suggested to possess several advantages over more traditional methods. For example, the results of EGA can be interpreted visually instead of interpreting a factor loading matrix, and there is no need to make decisions about the factor rotation (Golino et al., 2020). Two different estimators have been suggested to be used in EGA (Golino et al., 2020): the graphical least absolute shrinkage and selection operator (GLASSO) (Friedman et al., 2008) and the triangulated maximally filtered graph (TMFG) (Massara et al., 2016). The advantage of the EGA-TMFG method is that it does not assume the data to be multivariate normal, and it is suggested to perform at its best with unidimensional data (Golino et al., 2020). Furthermore, total entropy fit index (TEFI) using Von Neuman's entropy can be used to evaluate the EGA model fit, and lower values of TEFI indicate lower disorder (i.e., better fit) (Golino et al., 2020).

Confirmatory factor analysis (CFA) covers the steps of model specification, estimation, and evaluation (Brown, 2015). The basis for the model specification in our research was the original (a priori) unidimensional model (i.e., Franke et al., 2019). The model estimation was conducted employing polychoric correlation (Holgado-Tello et al., 2010) and robust diagonally weighted least squares (DWLS) estimation with test statistics adjusted in terms of mean and variance (i.e., scale-shifted approach, a.k.a., WLSMV (El-Sheikh et al., 2017)), which is a suggested estimation method for ordinal data (Beauducel & Herzberg, 2006; DiStefano & Morgan, 2014; Foldnes & Grønneberg, 2021; Forero et al., 2009; Li, 2016a, 2016b, 2021).

To describe the goodness of fit of the CFA models, we used the standardized root mean squared residuals (SRMR) as an indicator of the absolute fit, the root mean square error of approximation (RMSEA) as an indicator of the parsimony corrected fit, and the comparative fit index (CFI) and Tucker—Lewis index (TLI) as indicators of the comparative fit. In general, SRMR and RMSEA values closer to zero and CFI and TLI values closer to one are considered as indicators of better fit of the model. Specifically, SRMR is relatively insensitive to different estimators and appropriate to use in the case of ordinal models (Shi & Maydeu-Olivares, 2020). Various suggestions for deriving cut off values and combinational rules for an acceptable model fit can be found in the literature (e.g., TLI or CFI > 0.95 and SRMR < 0.09 (Hu & Bentler, 1999), dynamic fit index (McNeish & Wolf, 2021)); however, no "golden rule" exists (e.g., Greiff & Heene, 2017; Shi et al., 2019).

Furthermore, we conducted a specification search using modification indices (i.e., Lagrange multipliers) to examine the localized areas of strain in the model. Complementing the global model fit assessments based on the goodness-of-fit measures, the modification indices based on the expected parameter change provide insights about the local misspecifications (e.g., Greiff & Heene, 2017). The use of modification indices is exploratory in nature, and the modifications should be based on underlying theoretical assumptions of the model (Brown, 2015, p. 106). CFA was conducted and the modifications were applied based on the expected parameter change and power analysis (Saris et al., 2009) using the *R* package *lavaan* (Rosseel, 2012).

We estimated the reliability of the scale from the classical test theory (CTT) and factor analysis points of views. Coefficient α (e.g., Cronbach, 1951) is based on the assumptions of CTT (Lord & Novick, 1968, p. 36–38). The underlying idea in reliability is replicability: the reliability $\rho_{XX'}$ of a test reflects the degree of linear correlation between two parallel tests having the same formal properties (Sijtsma & Pfadt, 2021). In essence, coefficient α is a lower bound to the reliability (Sijtsma, 2009; Sijtsma & Pfadt, 2021); however, under approximate unidimensionality it is close to reliability $\rho_{XX'}$ (Sijtsma & Pfadt, 2021). On the other hand, the reliability coefficient ω (e.g., McDonald, 1999, p. 88–90) is based on the concept of a factor analysis (FA) model. As suggested for categorical data, we estimated the reliability following FA approach by using categorical omega ω_c with bias-corrected and accelerated bootstrap confidence interval (Dunn et al., 2014; Kelley & Pornprasertmanit, 2016) implemented in *R* package *MBESS* (Kelley, 2007).

2.4.4. Differential item functioning

Differential item functioning (DIF) means that persons having the same level of ability in the latent variable respond differently to the item depending on the persons' characteristics. For example, if an item exhibits gender-based DIF, it means that men and women with the same ability with respect to the latent variable have different probabilities for response categories. A wide variety of methods have been proposed to detect DIF, but many of them suffer from fundamental issues (e.g., requiring a priori chosen anchor items) (Bechger & Maris, 2015; Yuan et al., 2021). For detecting uniform DIF, we used a recent method utilizing an approach based on the lasso principle, which does not require using anchor items (Schauberger & Mair, 2020). The DIF analysis was executed using *R* package *GPCMlasso* (Schauberger, 2019).

2.4.5. Total scale score

When measuring a latent variable using a psychometric scale, the total scale score X_+—usually calculated as the unweighted sum of all item scores—is assumed to be the proxy of the measured latent variable. Some IRT models have a property called stochastic ordering of the latent variable by X_+(SOL by X_+), which implies that a higher total score X_+ results in a higher expected latent variable value (Hemker et al., 1997). If the data comply with a model having the property of SOL by X_+, then the simple sum score X_+ can be used to order respondents in terms of the latent variable (Sijtsma & Hemker, 2000). SOL by X_+ holds for MHM with dichotomous data (Mokken, 1997; Sijtsma & Hemker, 2000). MHM with polytomous data does not imply SOL by X_+ (Hemker et al., 1997). However, a property called weak SOL was proposed to apply to MHM with polytomous data, which in turn is argued to justify the ordering of respondents on the latent variable using the total score X_+ (van der Ark & Bergsma, 2010). MHM for polytomous items does not imply complete person ordering, but it allows for pairwise person ordering (van der Ark et al., 2019). Even though MHM might not be completely satisfactory for the exact ordering of individuals, it can be used to order groups of people using statistics of central tendency (e.g., mean and median) as people with a higher X_+ have on average a higher ability on the latent variable compared to people with a lower X_+ (Zwitser & Maris, 2016). The corrected item-total correlation is used to indicate the coherence between an item and the other items in a scale, and it

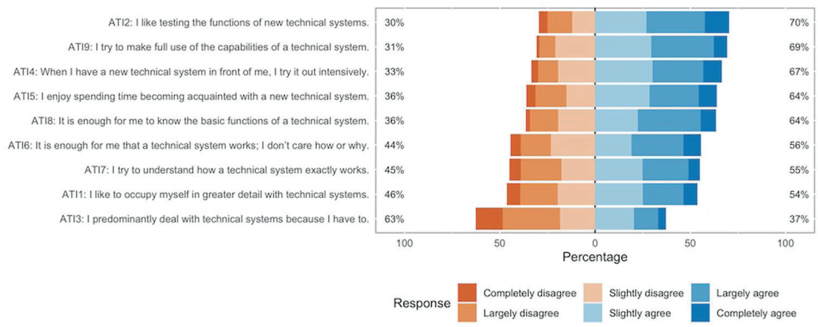


Figure 2. Likert responses without outliers. RW items ATI3, ATI6 and ATI8 are not reversed.

Table 2. Distributional properties of the items without outliers ordered by the mean.

Item	M	SD	Skew G_1	Kurtosis G_2	SE
ATI9	4.05	1.13	−0.40	−0.39	0.04
ATI2	4.04	1.37	−0.52	−0.57	0.05
ATI3R	4.00	1.40	−0.33	−0.86	0.05
ATI4	3.95	1.26	−0.35	−0.46	0.05
ATI5	3.83	1.36	−0.31	−0.77	0.05
ATI7	3.58	1.37	−0.13	−0.99	0.05
ATI1	3.56	1.38	−0.09	−0.92	0.05
ATI6R	3.25	1.39	0.17	−0.93	0.05
ATI8R	3.07	1.28	0.33	−0.84	0.05

The effect of including outliers was negligible.

is one of the best methods for item assessment when constructing tests (Zijlmans et al., 2018). The corrected item-total correlation is calculated by correlating the item score with the total scale score without that item. Items with a higher corrected item-total correlations are more desirable (DeVellis, 2017, p. 142).

2.5. Ethical considerations

The research was conducted following the guidelines of the World Medical Association (WMA) Declaration of Helsinki. According to the guidelines of The Finnish National Board on Research Integrity and the research institution where the research was conducted, ethical pre-evaluation or permission was not needed for executing the research. Participation in the research was voluntary. Research and privacy statement was prepared following the GDPR and national legislation. An informed consent was asked before respondents answered the questionnaire. Identifying information (i.e., email address) was asked to enable the voluntary gift card raffle. It was also possible to participate in the research entirely anonymously. The data were anonymized directly after data collection.

3. Results

3.1. Descriptive statistics

First, multivariate outliers were identified using MMCD75 (Leys et al., 2018), and 39 (4.9%) responses were identified as potential outliers. After excluding the outliers, there were $n = 757$ responses with similar demographic properties as the complete data. We analyzed the data with and without outliers. The results including outliers are interpreted in the text or presented using a marker †. It is worth noting that the distributional properties of the data represent the reponses of this particular convenience sample consisting of relatively young and educated people.

Figure 2 depicts the distributions of answer categories of each scale item without outliers. All answer categories received responses. The least amount of responses were in *Completely disagree* categories of items ATI8 ($n = 20$; 2.6%) and ATI9 ($n = 14$; 1.8%). Table 2 describes the distributional properties of the items without outliers. The effect of the outliers on the distributional properties was negligible. Mardia's tests for multivariate skewness and kurtosis of the items were significant, which indicated the data deviated from multivariate normal distribution. As expected, the reverse-worded items ATI3, ATI6, and ATI8 were negatively associated with all other items. After reverse-coding the reverse-worded items using linear scaling, all interitem correlations were positive.

Correlations between the items using polychoric correlation ranged between 0.37–0.83 (0.31–0.81†). The item ATI3R exhibited the weakest association with other items (0.37–0.49, 0.31–0.46†), however, not as weak as reported in Lezhnina and Kismihók (2020) (0.14–0.26). Heterogeneous interitem correlations could indicate that the items do not capture the latent variable equally well (DeVellis, 2017, p. 55).

We used a scalogram to depict the variation in the respondents' response patterns visually. Figure 3 shows all response patterns excluding the outliers. In the figure, the respondents were ordered according to the sum score X_+, and the scale items were ordered according to the item sum score. Thus, the colors depicting the amount of agreeableness would accumulate to the top right corner of the figure. Respectively, the colors depicting the amount of disagreeableness would accumulate to the lower left corner of the figure. Visual inspection showed that especially the item ATI3R exhibited a somewhat irregular response pattern. Appendix Figure D1 depicts the scalogram containing only the outliers identified by the MMCD75 procedure. Several dubious responses can be identified (e.g., extreme and low scorers, contradicting responses) exhibiting ambiguous response patterns. As described above, instead of subjective

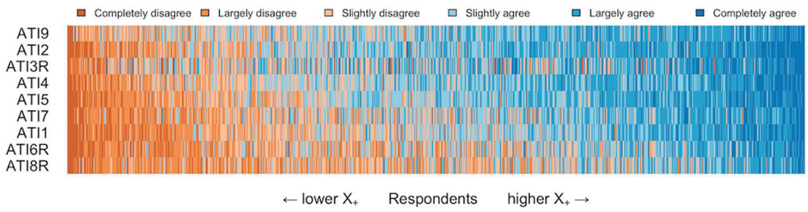


Figure 3. Scalogram without outliers ($n = 757$), respondents ordered by total score, and items ordered by total item sum.

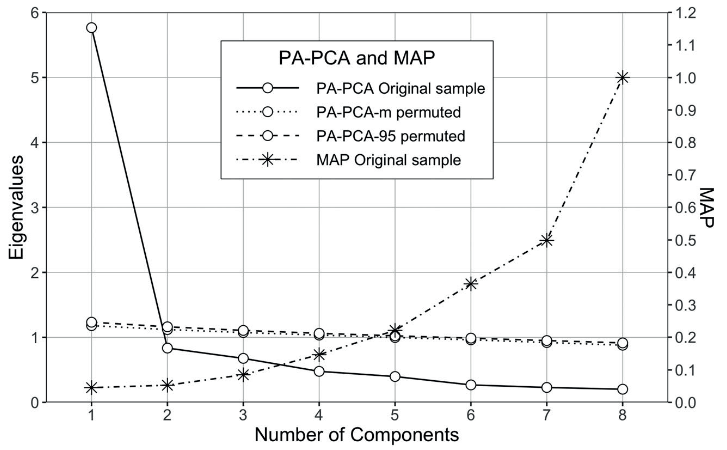


Figure 4. Both MAP and parallel analysis using PCA for the complete data without outliers supported structure with one component.

selection, we removed all outliers suggested by MMCD75 and report results with and without outliers.

3.2. Mokken scale analysis

We utilized non-parametric item response theory (Sijtsma, 2005), namely the monotone homogeneity model (Mokken, 1971), by applying Mokken scale analysis (Sijtsma & van der Ark, 2017) to the scale data. First, we examined the scalability of the scale items using the coefficient H (Appendix Table B1). After reverse coding the reverse-worded items, all scalability coefficients were positive. For individual items, the values were $0.43 < H_j < 0.66$ $(0.014 < SE < 0.027)$ exceeding the traditional cutoff value $c = 0.3$. For item pairs, the values were $0.35 < H_{jk} < 0.82$. The effect of including outliers was small. Local independence was assessed using conditional association procedure (Straat et al., 2016). Without outliers, all items were found to be locally independent. With outliers, $W^{(1)}$ flagged the item pair ATI7–ATI1 as locally dependent.

We examined the monotonicity of the ISRFs, and there were only six (five†) violations of manifest monotonicity (all in RW items) using $minvi = N/10$, $minvi = 0.03$. All violations of manifest monotonicity were non-significant at the level of $\alpha = .05$, which indicates that the assumption of monotonicity holds. However, the data with and without outliers showed significant violations of invariant item ordering. Thus, the data did not support the more strict assumption of the double monotonicity model.

Table 3. The results of the consecutive steps of the automatic item selection process (AISP).

Item	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55	.60
ATI1	1	1	1	1	1	1	1	1	1	1	1	1
ATI2	1	1	1	1	1	1	1	1	1	1	1	1
ATI3R	1	1	1	1	1	1	1	1;0†	0	0	0	0
ATI4	1	1	1	1	1	1	1	1	1	1	1	1
ATI5	1	1	1	1	1	1	1	1	1	1	1	1
ATI6R	1	1	1	1	1	1	1	1	1	1	1;2†	2
ATI7	1	1	1	1	1	1	1	1	1	1	1	1;2†
ATI8R	1	1	1	1	1	1	1	1	1	1;0†	0;2†	2;0†
ATI9	1	1	1	1	1	1	1	1	1	1	1	1

Note. Data with outliers is marked using †.

In addition, we used AISP to assess the dimensionality of the scale. When increasing the threshold consecutively using $c = \{0.05, 0.10, 0.15, \ldots, 0.60\}$, one would expect to find most or all items assigned to the same scale (Sijtsma & van der Ark, 2017). Both AISP algorithms (i.e., Mokken's method and the genetic algorithm) produced identical results for data without outliers. With outliers, the algorithms produced slightly different results, and Mokken's method is reported with outliers (Table 3). The AISP algorithms assigned all items to the same scale without outliers at the level of $c = 0.40$ and with outliers at the $c = 0.35$. According to AISP, the RW items were less scalable than other items. The outliers affected the scalability of RW items and ATI7; however, the effect was minimal. In summary, the results of the Mokken scale analysis supported the requirements of MHM (i.e., unidimensionality, monotonicity, and local independence). Furthermore, the Finnish

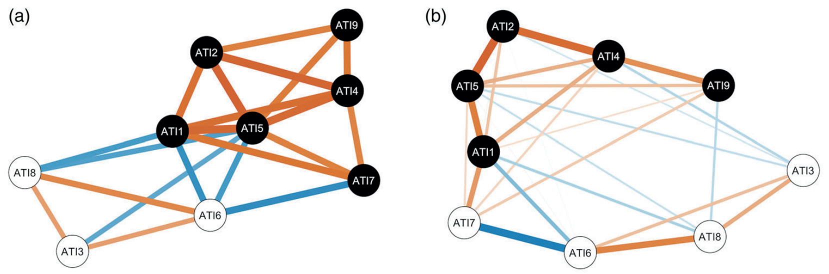


Figure 5. EGA for the complete data without outliers using TMFG (a) showed better fit (TEFI = −4.97) supporting two-dimensional structure. Edges represent the partial correlations between items.

version of the ATI scale formed a strong Mokken scale, which met the criteria of the monotone homogeneity model.

3.3. Factor analysis and classical test theory

3.3.1. Parallel analysis

Figure 4 shows the results of the parallel analysis. PA-PCA supported unidimensional structure because the eigenvalue of the first component in the scale data was higher and the value of the second component was lower than the corresponding values in the parallel simulated data. Two of the smallest values of MAP were 0.045 ($0.047^{\dagger}$) and 0.053 ($0.049^{\dagger}$). Also, the smallest MAP value suggested a unidimensional structure. The effect of including outliers in the parallel analysis was negligible.

3.3.2. Exploratory graph analysis

We applied EGA using both GLASSO and TMFG estimation for the data without outliers. The model using TMFG (Figure 5(a)) was chosen as the final model as TMFG was suggested to perform better in case of unidimensional data (Golino et al., 2020) and it also showed a smaller TEFI value (TEFI = −4.97) when compared to the GLASSO estimation (TEFI = −4.57) (Figure 5(b)). Both estimation methods showed similar two-dimensional structures, except that the GLASSO estimation assigned ATI7 to the same dimension with RW items. The structures were identical with the data including outliers, except that the GLASSO estimation assigned also ATI1 to the same dimension with the ATI7 and RW items (Figure 2(a,b)).

3.3.3. Confirmatory factor analysis

We conducted CFA for the original unidimensional a priori model (Model 1). The model without outliers showed a better fit than the model including outliers (Table 4). The post hoc exploratory power analysis of the modification indices suggested adding correlated errors between ATI6R–ATI7, ATI6R–ATI8R, ATI3R–ATI8R, ATI3R–ATI6R, and ATI1–ATI6R. The results reflect a local discrepancy in the model. In other words, the model does not adequately reproduce the relationships of the item pairs mentioned above. Adding correlated errors based on the modification indices can be justified if there is a reason to believe that some of the covariations are due to some common

Table 4. The CFA estimates for a priori model (Model 1) and modified model (Model 2) using the data without outliers and with outliers (†).

	χ^2	*df*	$p <$	χ^2/df	RMSEA [90% *CI*]	SRMR	CFI	TLI
Model 1	442.7	27	.001	16.4	0.143 [0.131–0.155]	0.053	0.971	0.961
Model $1^{\dagger}$	568.2	27	.001	21.0	0.159 [0.148–0.170]	0.061	0.958	0.944
Model 2	130.6	22	.001	5.9	0.081 [0.068–0.094]	0.024	0.992	0.988
Model $2^{\dagger}$	222.8	22	.001	10.1	0.107 [0.095–0.120]	0.032	0.984	0.974

exogenous cause instead of the latent variable (Brown, 2015, p. 157). The common cause for the discrepancy of the RW item pairs could be caused by a common method bias (DiStefano & Motl, 2006; Podsakoff et al., 2003; Woods, 2006) and systematic bias relating to item wording (Dalal & Carter, 2014, p. 117). For the other item pairs (i.e., ATI6R–ATI7 and ATI1–ATI6R), the cause for a local misfit could be their polar opposite wording as polar opposite items could affect on factorial structure (Zhang et al., 2016). Consequently, we applied the modification indices above, and the modified model (Model 2) resulted in an improved and sufficient fit both with and without outliers.

3.3.4. Reliability

From the FA point of view, the test score reliability using categorical omega showed high reliability, $\omega_c = 0.946$ ($0.927^{\dagger}$), 95% *CI* [0.926 ($0.915^{\dagger}$), 0.944 ($0.936^{\dagger}$)]. Also, in terms of CTT, the lower bound to the reliability of the ATI test score using coefficient α showed high reliability estimate, $\alpha = 0.915$ ($0.903^{\dagger}$), 95% *CI* [0.905 ($0.891^{\dagger}$), 0.924 ($0.913^{\dagger}$)]. The effect of outliers was minimal.

3.3.5. Differential item functioning

We examined the existence of a uniform DIF concerning age and gender using a regularization approach based on the lasso principle and PCM (Schauberger & Mair, 2020). Only item ATI3R exhibited uniform DIF based on gender, and none of the items showed a DIF concerning age. The results were the same for both data with and without outliers.

3.4. Total scale score

The corrected item-total correlations were adequate ranging between 0.51 (ATI3R) and 0.82 (ATI5). The RW items had the lowest corrected item-total correlations (0.51–0.67). Shapiro-Wilk normality test was significant ($W = 0.99$, $p <$

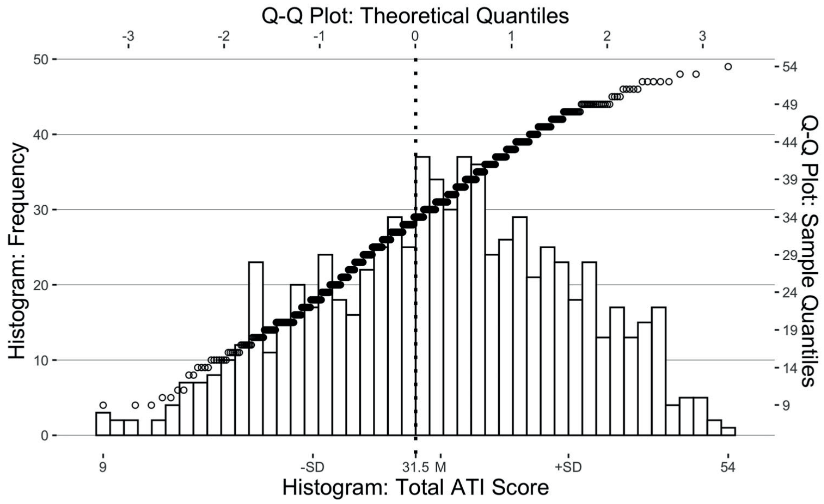


Figure 6. The histogram and the superimposed Q-Q plot of the total ATI scale score without outliers. The total ATI scale range is [9, 54], and the center of the scale is 31.5. The mean and the standard deviation in the figure represent the values from the complete data without outliers.

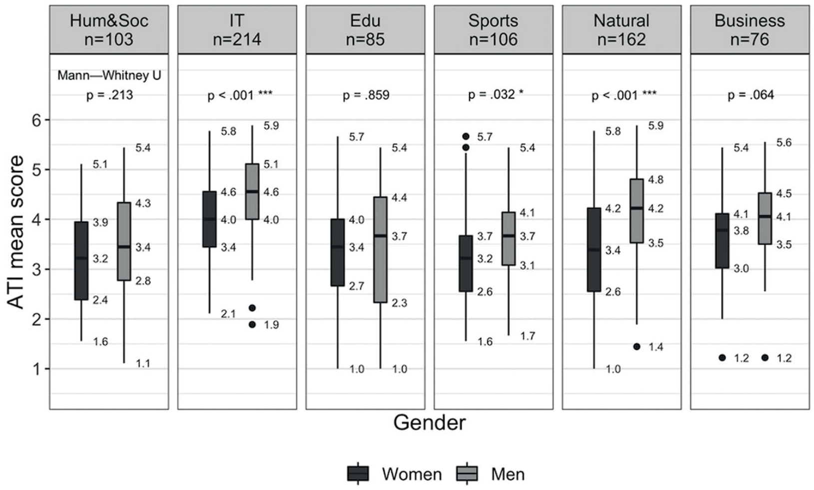


Figure 7. Difference in mean ATI scale score between genders in different field of studies.

.001), indicating that the total score was not normally distributed. However, the histogram and quantile-quantile plot of the total score (Figure 6) showed a shape of an approximately normal distribution. The sample mean ($M = 33.3$) was close to the center of the scale ($C = 31.5$). There were no ceiling or flooring effects present. The scale mean for all respondents without outliers was 3.7 ($SD = 1.0$), for all women 3.5 ($SD = 0.97$), for all men 4.1 ($SD = 0.99$), and for a very small sample of non-binary respondents 4.0 ($SD = 0.88$). Appendix Table B2 shows the descriptive statistics of the total score in different groups. Figure 7 shows the difference in mean ATI scale score between women and men by field of study in this sample. Men showed slightly higher score value than women, specifically and interestingly in the fields of information technology and natural sciences. The results describe the differences in this particular sample and convenience sampling limits the generalizability of the results.

We used hierarchical multiple regression analysis to determine if the addition of gender improved the prediction of the total ATI scale score over and above age and field of study alone. The difference in explained variance between the models with and without gender as an independent variable would indicate the effect of gender on ATI when controlling for age and field of study. After fitting the model, visual inspection of the plot of studentized residuals versus unstandardized predicted values and the quantile-quantile plot did not reveal heteroscedasticity or violations of normality in the full model.

The full regression model regressing the total ATI scale score on gender, age, and field of study was statistically significant, $R^2 = 0.18, R^2_{adj} = 0.17, F(9, 739) = 17.76, p < .001$. The full model results indicated that when controlling age and faculty, men had 4.6, 95% *CI* [3.2, 5.9] points higher total ATI scale score (0.51 on the Likert scale). When comparing to a nested model without gender variable, the addition of gender to the prediction of the total ATI scale score led to a statistically significant increase in the coefficient of determination, $\Delta R^2_{adj} = 0.05, F(2, 739) = 22.51, p < .001$. In other words, gender accounted for an additional 5% of the variance in the data and the relating effect size of gender on ATI score, $f^2 = 0.06$, 95% *CI* [0.05, 0.07], was small when using conventional criteria (Cohen, 1988, p. 413). The effect of outliers on the effect size was negligible.

In a meta-analysis by Cai et al. (2017), the overall weighted effect size relating to gender and attitudes toward technology (i.e., men having favorable attitudes towards technology) across all 87 reported studies was found to be small. On the other hand, when comparing group means between men and women using a quota sample, Franke et al. (2019) identified a large effect in ATI (i.e., men having a higher ATI score). Naturally, the findings relating to gender differences in our study can not be generalized due to sampling method and sample characteristics. However, the results are a one indication that the translated scale is able to capture differences between groups.

4. Discussion

Affinity for technology interaction (ATI) scale is a psychometric instrument used to quantify the tendency of a person to "actively approach interaction with technical systems or, rather, tend to avoid intensive interaction with new systems" (Franke et al., 2019). This research presented a psychometric analysis and properties of a Finnish version of the scale. The main aims of the analyses were to assess the evidence concerning the scale's internal structure (i.e., dimensionality and the functioning of the individual items). Furthermore, we examined its ability to capture differences and similarities in ATI among students in higher education by gender and field of study.

The Finnish translation was conducted using a professional forward-backward translation process with committee approach (Sousa & Rojjanasrirat, 2011). Data were collected using convenience sampling, and the respondents were university students from six different faculties in a Finnish university. Our comprehensive analysis involved factor analysis and analyses based on both parametric and non-parametric IRT. In addition, we analyzed the data in terms of outliers, and the results were reported both with and without outliers. In general, the outliers seemed to hinder the properties of the scale, but the effect was minimal.

Unidimensionality is a convenient feature of a psychometric scale and the original scale has been deemed unidimensional in previous studies (Franke et al., 2019; Lezhnina & Kismihók, 2020). For the translated version in this study, parallel analyses using traditional PA-PCA and MAP supported a unidimensional structure. On the other hand, a network psychometrics approach using EGA showed a two-dimensional structure where the RW items separated as a second dimension. Post hoc analysis of the unidimensional CFA model showed similar structural indications as EGA. As a result, it can be stated that RW items and both items ATI1 and ATI7 showed some discrepancy concerning unidimensionality. The discrepancy from the unidimensionality could be caused by a common method bias relating to RW items. It is well-known in the literature that the RW items in a scale can hinder the unidimensional structure (Boley et al., 2020; Suárez-Álvarez et al., 2018). Furthermore, they can affect response patterns (Baumgartner et al., 2018; Woods, 2006; Zhang et al., 2016) and mixed scales could also be less reliable and have more measurement error (Dalal & Carter, 2014; Schriesheim et al., 1991).

In general, however, the translated version showed at least moderate fit to the unidimensional model even though the cut-off values for the fit indices for ordinal CFA are not yet settled in the research literature. Essential unidimensionality means that the scale consists of minor dimensions still tapping the same latent variable (Slocum-Gori & Zumbo, 2011). Thus, the translated version could be considered as essentially unidimensional. From the both CTT and FA point of view, the translated scale showed excellent reliability estimates.

The Mokken scale analysis based on the non-parametric IRT showed that the translated scale's data fitted well to the MHM. In other words, the scale conformed with the requirements of unidimensionality, monotonicity, and local independence. Furthermore, the original ATI scale was found to support also invariant item ordering (Lezhnina & Kismihók, 2020). However, the Finnish version in our research did not support invariant item ordering indicating that the items do not have a specific ordering based on the item difficulty. The translated version formed a strong Mokken scale which means it supports at least the weak SOL by X_+. Thus, it is possible to form a composite total scale score that can be used to order persons on the latent variable (van der Ark & Bergsma, 2010).

The possibility to use the total scale score to differentiate persons concerning their latent variable (SOL by X_+) and the lack of uniform differential item functioning allowed us to advance our analysis by examining the gender differences relating to ATI. We used multiple regression to assess the gender difference in ATI while controlling for the known variables, age and the faculty the respondent was studying in. The results showed that men exhibited slightly more affinity towards technology interaction than women. Specifically, the difference in the sample among the IT students was interesting as the particular subsample was more balanced in terms of gender than other subsamples. The findings can be considered as indications of the differential validity of the scale. While the actual effect size relating to gender was small, a similar difference between genders has been identified in the research literature, and the effect size in this study was comparable to findings in a recent meta-review by Cai et al. (2017). Naturally, it is worth questioning if the difference is large enough to have any practical significance (c.f., Whitley, 1997). However, as the technical systems gain more and more traction and influence in our lives, even a small effect can become significant and meaningful over time. Technological agency would be an essential characteristic that enables and promotes equal participation in various fields of life. Thus, it is of the utmost importance to have valid and cross-cultural instruments to assess the different personal stances towards technology and technical systems.

Our study presented several contributions. Firstly, to our knowledge, this is the first comprehensively analyzed translation besides the original ATI scale versions (c.f., Franke et al., 2019) and the first Finnish psychometric scale for measuring affinity for technology interaction. Secondly, as the previous research has analyzed the ATI scale using factor analysis and non-parametric IRT methods (Franke et al., 2019; Lezhnina & Kismihók, 2020), our analysis utilized also parametric IRT methods using PCM. Lastly, we provided an estimate of the gender effect relating to the affinity for technology interaction.

4.1. Limitations and future research

There are limitations in our study that need to be considered. Using an online questionnaire as a medium could be a source of common method bias (Podsakoff et al., 2003). Data were collected from university students having secondary education backgrounds and who pursued at least a bachelor's degree in their studies. Also, the sample consisted of mostly relatively young people. The sample characteristics mentioned above limit the generalizability of the results with respect to the general population, which is the population of interest of the ATI scale. Specifically, the results concerning the gender differences in ATI can only be applied to the population used in this research (i.e., Finnish university students).

However, it is notable that the fairly large sample covered a broad range of fields of study, and thus it can be seen as sufficiently representing Finnish university students. The translated version of the scale and the results presented here can be useful in research (e.g., educational technology, social media use, technology adoption) targeting university students. Considering the solid results, simple language used in the scale, and the nature of the construct, one could assume that the internal structure of the translated scale in the general population could follow the results presented here. The promising results presented in this study should encourage researchers to conduct more extensive studies. Future research should complement the findings of this research by examining the properties of the translated scale version and gender differences among older people and people with more diverse educational backgrounds, using samples representing the general population, and using other mediums for provisioning the questionnaire.

In general, results concerning gender-based differences in affinity for technology interaction should be treated with caution. The threat of a stereotypical interpretation is important to consider as it can have detrimental effects in various situations (e.g., Barber, 2020; Cadaret et al., 2017; Doyle & Thompson, 2021). Thus, it is worth noting that research examining gender-based differences has the potential to advance unfounded and stereotypical beliefs if not conducted with rigor and interpreted with care. On the other hand, it is essential to examine the differences, for example, from the equity point of view. For that purpose, functional measurement instruments are valuable assets.

We aimed to conduct an accurate and thorough translation; however, in a cross-cultural research, the complete equivalence between different languages is challenging to achieve. Thus, the presented Finnish translation should be examined in different contexts in future research. Furthermore, we did not assess the relationship of ATI with other similar constructs as there is a limited resource of related and validated psychometric scales in Finnish. Thus, future studies should examine the relationships of ATI with similar constructs within a nomological network.

We used correlated errors to modify the CFA model, which can be problematic (Hermida, 2015). Instead, another approach would be to examine the effect of the reverse-worded items using a method factor (DiStefano & Motl, 2006). In general, the use of reverse-worded items in the first place is a controversial topic, and future studies should address the issue of how different types of reverse-worded items affect the factor structure (Zhang et al., 2016). Furthermore, the item ATI3R exhibiting lower qualities would need to be assessed critically and possibly revised as was also noted in a previous study (c.f., Lezhnina & Kismihók, 2020).

5. Conclusion

We analyzed the psychometric properties of a forward-backward translated Finnish version of the affinity for technology interaction (ATI) scale. The analysis utilized factor analysis, non-parametric IRT, and parametric IRT. To conclude, the Finnish version of the ATI scale showed solid psychometric properties. Furthermore, the scale proved to be essentially unidimensional, having high reliability estimates, and forming a strong Mokken scale. The scale also showed differential validity by identifying a gender difference with respect to the measured construct: men showed slightly more affinity towards technology among the respondents in the sample; however, the effect size was small.

Note

1. Affinity for Technology Interaction Scale https://ati-scale.org/

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Ville Heilala http://orcid.org/0000-0003-2068-2777
Riitta Kelly http://orcid.org/0000-0002-5285-7386
Mirka Saarela http://orcid.org/0000-0002-1559-154X
Päivikki Jääskelä http://orcid.org/0000-0002-3794-0325
Tommi Kärkkäinen http://orcid.org/0000-0003-0327-1167

References

Adams, R. J., Wu, M. L., & Wilson, M. (2012). The rasch rating model and the disordered threshold controversy. *Educational and Psychological Measurement*, *72*(4), 547–573. https://doi.org/10.1177/0013164411432166

Aghekyan, R. (2020). Validation of the SIEVEA instrument using the rasch analysis. *International Journal of Educational Research*, *103*, 101619. https://doi.org/10.1016/j.ijer.2020.101619

Alabdullatif, H., & Velázquez-Iturbide, J. Á. (2020). Personality traits and intention to continue using massive open online courses (ICM) in Spain: The mediating role of motivations. *International Journal of Human–Computer Interaction*, *36*(20), 1953–1967. https://doi.org/10.1080/10447318.2020.1805873

Ali, M. A., Alam, K., Taylor, B., & Rafiq, S. (2020). Does digital inclusion affect quality of life? Evidence from Australian household panel data. *Telematics and Informatics*, *51*, 101405. https://doi.org/10.1016/j.tele.2020.101405

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Attig, C., & Franke, T. (2019). I track, therefore I walk – exploring the motivational costs of wearing activity trackers in actual users. *International Journal of Human-Computer Studies*, *127*, 211–224. https://doi.org/10.1016/j.ijhcs.2018.04.007

Auerswald, M., & Moshagen, M. (2019). How to determine the number of factors to retain in exploratory factor analysis: A comparison of extraction methods under realistic conditions. *Psychological Methods*, *24*(4), 468–491. https://doi.org/10.1037/met0000200

Barber, S. J. (2020). The applied implications of Age-Based stereotype threat for older adults. *Journal of Applied Research in Memory and Cognition*, *9*(3), 274–285. https://doi.org/10.1016/j.jarmac.2020.05.002

Baumgartner, H., Weijters, B., & Pieters, R. (2018). Misresponse to survey questions: A conceptual framework and empirical test of the effects of reversals, negations, and polar opposite core concepts. *Journal of Marketing Research*, *55*(6), 869–883. https://doi.org/10.1177/0022243718811848

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*(2), 186–203. https://doi.org/10.1207/s15328007sem1302_2

Beaunoyer, E., & Guitton, M. J. (2021). Cyberthanathology: Death and beyond in the digital age. *Computers in Human Behavior*, *122*, 106849. https://doi.org/10.1016/j.chb.2021.106849

Bechger, T. M., & Maris, G. (2015). A statistical test for differential item pair functioning. *Psychometrika*, *80*(2), 317–340. https://doi.org/10.1007/s11336-014-9408-y

Boelhouwer, A., Beukel, A. P. V. D., van der Voort, M. C., Verwey, W. B., & Martens, M. H. (2020). Supporting drivers of partially automated cars through an adaptive digital In-Car tutor. *Information*, *11*(4), 185. https://doi.org/10.3390/info11040185

Boley, B. B., Bynum Boley, B., Jordan, E., & Woosnam, K. M. (2020). Reversed polarity items in tourism scales: Best practice or dimensional pitfall? *Current Issues in Tourism*, *24*(4), 1–13. https://doi.org/10.1080/13683500.2020.1774517

Bond, T., Yan, Z., & Heene, M. (2020). *Applying the Rasch model: Fundamental measurement in the human sciences* (4th ed.). Routledge.

Brislin, R. W., Lonner, W. J., & Thorndike, R. M. (1973). *Cross-cultural research methods*. John Wiley & Sons.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford Publications.

Buja, A., & Eyuboglu, N. (1992). Remarks on parallel analysis. *Multivariate Behavioral Research*, *27*(4), 509–540. https://doi.org/10.1207/s15327906mbr2704_2

Caci, B., Scrima, F., Tabacchi, M. E., & Cardaci, M. (2019). The reciprocal influences among motivation, personality traits, and game habits for playing pokémon GO. *International Journal of Human–Computer Interaction*, *35*(14), 1303–1311. https://doi.org/10.1080/10447318.2018.1519167

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*(1), 116–131. https://doi.org/10.1037/0022-3514.42.1.116

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin*, *119*(2), 197–253. https://doi.org/10.1037/0033-2909.119.2.197

Cadaret, M. C., Hartung, P. J., Subich, L. M., & Weigold, I. K. (2017). Stereotype threat as a barrier to women entering engineering careers. *Journal of Vocational Behavior*, *99*, 40–51. https://doi.org/10.1016/j.jvb.2016.12.002

Cai, Z., Fan, X., & Du, J. (2017). Gender and attitudes toward technology use: A meta-analysis. *Computers & Education*, *105*, 1–13. https://doi.org/10.1016/j.compedu.2016.11.003

Cain, M. K., Zhang, Z., & Yuan, K.-H. (2017). Univariate and multivariate skewness and kurtosis for measuring nonnormality: Prevalence, influence and estimation. *Behavior Research Methods*, *49*(5), 1716–1735. https://doi.org/10.3758/s13428-016-0814-1

Cameron, J. J., & Stinson, D. A. (2019). Gender (mis)measurement: Guidelines for respecting gender diversity in psychological research. *Social and Personality Psychology Compass*, *13*(11), e12506. https://doi.org/10.1111/spc3.12506

Cha, E.-S., Kim, K. H., & Erlen, J. A. (2007). Translation of scales in cross-cultural research: issues and techniques. *Journal of Advanced Nursing*, *58*(4), 386–395. https://doi.org/10.1111/j.1365-2648.2007.04242.x

Chen, J. V., Widjaja, A. E., & Yen, D. C. (2015). Need for affiliation, need for popularity, Self-Esteem, and the moderating effect of big five personality traits affecting individuals' Self-Disclosure on facebook. *International Journal of Human-Computer Interaction*, *31*(11), 815–831. https://doi.org/10.1080/10447318.2015.1067479

Chen, L.-M., & Jin, K.-Y. (2020). Development and validation of the willingness to intervene in bullying scale. *The British Journal of Educational Psychology*, *90*(Suppl 1), 224–239. https://doi.org/10.1111/bjep.12319

Christensen, K. B., Makransky, G., & Horton, M. (2017). Critical values for yen's q3: Identification of local dependence in the rasch model using residual correlations. *Applied Psychological Measurement*, *41*(3), 178–194.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum.

Cranmer, E. E., Han, D.-I. D., Gisbergen, M. V., & Jung, T. (2021). Esports matrix: Structuring the esports research agenda. *Computers in Human Behavior*, *117*, 106671. https://doi.org/10.1016/j.chb.2020.106671

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334. https://doi.org/10.1007/BF02310555

Dalal, D. K., & Carter, N. T. (2014). Negatively worded items negatively impact survey research. In C. E. Lance & R. J. Vandenberg (Eds.), *More statistical and methodological myths and urban legends* (pp. 112–132). Routledge.

DeVellis, R. F. (2017). *Scale development: Theory and applications* (4th ed.). SAGE Publications.

DiStefano, C., & Morgan, G. B. (2014). A comparison of diagonal weighted least squares robust estimation techniques for ordinal data. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*(3), 425–438. https://doi.org/10.1080/10705511.2014.915373

DiStefano, C., & Motl, R. W. (2006). Further investigating method effects associated with negatively worded items on Self-Report surveys. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*(3), 440–464. https://doi.org/10.1207/s15328007sem1303_6

Doyle, R. A., & Thompson, A. E. (2021). The influence of stereotype threat and implicit theories of emotion on gender differences in emotional intelligence. *The Journal of Men's Studies*, *29*(2), 131–155. https://doi.org/10.1177/1060826520920878

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: a practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology (London, England : 1953)*, *105*(3), 399–412. https://doi.org/10.1111/bjop.12046

Edison, S. W., & Geissler, G. L. (2003). Measuring attitudes towards general technology: Antecedents, hypotheses and scale development. *Journal of Targeting, Measurement and Analysis for Marketing*, *12*(2), 137–156. https://doi.org/10.1057/palgrave.jt.5740104

El-Sheikh, A. A., Abonazel, M. R., & Gamil, N. (2017). A review of software packages for structural equation modeling: A comparative study. *Applied Mathematics and Physics*, *5*(3), 85–94.

Epskamp, S., Maris, G. K. J., Waldorp, L. J., & Borsboom, D. (2018). Network psychometrics. In P. Irwing, T. Booth, & D. J. Hughes

(Eds.), *The Wiley handbook of psychometric testing: A multidisciplinary reference on survey, scale and test development* (pp. 953–986). Wiley.

Epskamp, S., Rhemtulla, M., & Borsboom, D. (2017). Generalized network psychometrics: Combining network and latent variable models. *Psychometrika*, *82*(4), 904–927. https://doi.org/10.1007/s11336-017-9557-x

Fabrigar, L. R., & Wegener, D. T. (2012). *Exploratory factor analysis*. Oxford University Press.

Fleischhauer, M., Enge, S., Brocke, B., Ullrich, J., Strobel, A., & Strobel, A. (2010). Same or different? Clarifying the relationship of need for cognition to personality and intelligence. *Personality & Social Psychology Bulletin*, *36*(1), 82–96. https://doi.org/10.1177/0146167209351886

Foldnes, N., & Grønneberg, S. (2021). The sensitivity of structural equation modeling with ordinal data to underlying non-normality and observed distributional forms. *Psychological Methods*, 1–11. Advance online publication. https://doi.org/10.1037/met0000385

Forero, C. G., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: A monte carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(4), 625–641. https://doi.org/10.1080/10705510903203573

Franke, T., Attig, C., & Wessel, D. (2019). A personal resource for technology interaction: Development and validation of the affinity for technology interaction (ATI) scale. *International Journal of Human–Computer Interaction*, *35*(6), 456–467. https://doi.org/10.1080/10447318.2018.1456150

Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics (Oxford, England)*, *9*(3), 432–441. https://doi.org/10.1093/biostatistics/kxm045

Garrido, L. E., Abad, F. J., & Ponsoda, V. (2013). A new look at Horn's parallel analysis with ordinal variables. *Psychological Methods*, *18*(4), 454–474. https://doi.org/10.1037/a0030005

Golino, H., Moulder, R. G., Shi, D., Christensen, A. P., Garrido, L. E., & Nieto, M. D. (2020). Entropy fit indices: New fit measures for assessing the structure and dimensionality of multiple latent variables. *Multivariate Behavioral Research*, 56(6), 874–902. https://doi.org/10.1080/00273171.2020.1779642

Golino, H., Shi, D., Christensen, A. P., Garrido, L. E., Nieto, M. D., Sadana, R., Thiyagarajan, J. A., & Martinez-Molina, A. (2020). Investigating the performance of exploratory graph analysis and traditional techniques to identify the number of latent factors: A simulation and tutorial. *Psychological Methods*, *25*(3), 292–320. https://doi.org/10.1037/met0000255

Greiff, S., & Heene, M. (2017). Why psychological assessment needs to start worrying about model fit. *European Journal of Psychological Assessment*, *33*(5), 313–317. https://doi.org/10.1027/1015-5759/a000450

Hemker, B. T., Andries van der Ark, L., & Sijtsma, K. (2001). On measurement properties of continuation ratio models. *Psychometrika*, *66*(4), 487–506. https://doi.org/10.1007/BF02296191

Hemker, B. T., Sijtsma, K., & Molenaar, I. W. (1995). Selection of unidimensional scales from a multidimensional item bank in the polytomous mokken I RT model. *Applied Psychological Measurement*, *19*(4), 337–352. https://doi.org/10.1177/014662169501900404

Hemker, B. T., Sijtsma, K., Molenaar, I. W., & Junker, B. W. (1997). Stochastic ordering using the latent trait and the sum score in polytomous IRT models. *Psychometrika*, *62*(3), 331–347. https://doi.org/10.1007/BF02294555

Hermida, R. (2015). The problem of allowing correlated errors in structural equation modeling: concerns and considerations. *Computational Methods in Social Sciences*, *3*(1), 5–17.

Holgado-Tello, F. P., Chacón-Moscoso, S., Barbero-García, I., & Vila-Abad, E. (2010). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality & Quantity*, *44*(1), 153–166. https://doi.org/10.1007/s11135-008-9190-y

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185. https://doi.org/10.1007/BF02289447

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. [Database] https://doi.org/10.1080/10705519909540118

IBM (2007). *Frequently asked questions*. https://web.archive.org/web/20210509015346/https://www.ibm.com/ibm/history/documents/pdf/faq.pdf

Joanes, D. N., & Gill, C. A. (1998). Comparing measures of sample skewness and kurtosis. *Journal of the Royal Statistical Society: Series D (the Statistician)*, *47*(1), 183–189. https://doi.org/10.1111/1467-9884.00122

Kammler, F., Brinker, J., Vogel, J., Hmaid, T., Thomas, O. (2019). How do we support technical tasks in the age of augmented reality? Some evidence from prototyping in mechanical engineering. In *Proceedings of the International Conference on Information Systems (ICIS) 2019* (pp. 1–17). AIS Electronic Library. https://aisel.aisnet.org/icis2019/future_of_work/future_work/1/

Kelley, K. (2007). Methods for the behavioral, educational, and social sciences: An R package. *Behavior Research Methods*, *39*(4), 979–984. https://doi.org/10.3758/bf03192993

Kelley, K., & Pornprasertmanit, S. (2016). Confidence intervals for population reliability coefficients: Evaluation of methods, recommendations, and software for composite measures. *Psychological Methods*, *21*(1), 69–92. https://doi.org/10.1037/a0040086

Kim, Y., & Lee, H. (2022). Falling in Love with Virtual Reality Art: A New Perspective on 3D Immersive Virtual Reality for Future Sustaining Art Consumption. *International Journal of Human–Computer Interaction*, *38*(4), 371–382. https://doi.org/10.1080/10447318.2021.1944534

Kim, J., Merrill Jr, K., Xu, K., & Sellnow, D. D. (2021). I like my relational machine teacher: An AI instructor's communication styles and social presence in online education. *International Journal of Human-Computer Interaction*, *37*(18), 1760–1770. http://doi.org/10.1080/10447318.2021.1908671

Kortum, P., & Oswald, F. L. (2018). The impact of personality on the subjective assessment of usability. *International Journal of Human–Computer Interaction*, *34*(2), 177–186. https://doi.org/10.1080/10447318.2017.1336317

Leys, C., Klein, O., Dominicy, Y., & Ley, C. (2018). Detecting multivariate outliers: Use a robust variant of the mahalanobis distance. *Journal of Experimental Social Psychology*, *74*, 150–156. https://doi.org/10.1016/j.jesp.2017.09.011

Lezhnina, O., & Kismihók, G. (2020). A multi-method psychometric assessment of the affinity for technology interaction (ATI) scale. *Computers in Human Behavior Reports*, *1*, 100004. https://doi.org/10.1016/j.chbr.2020.100004

Li, C.-H. (2016a). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, *48*(3), 936–949. https://doi.org/10.3758/s13428-015-0619-7

Li, C.-H. (2016b). The performance of ML, DWLS, and ULS estimation with robust corrections in structural equation models with ordinal variables. *Psychological Methods*, *21*(3), 369–387. https://doi.org/10.1037/met0000093

Li, C.-H. (2021). Statistical estimation of structural equation models with a mixture of continuous and categorical observed variables. *Behavior Research Methods*, *53*(5), 2191–2213. March). https://doi.org/10.3758/s13428-021-01547-z

Ligtvoet, R. (2012). An isotonic partial credit model for ordering subjects on the basis of their sum scores. *Psychometrika*, *77*(3), 479–494. https://doi.org/10.1007/s11336-012-9272-6

Loevinger, J. (1948). The technic of homogeneous tests compared with some aspects of scale analysis and factor analysis. *Psychological Bulletin*, *45*(6), 507–529. https://doi.org/10.1037/h0055827

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Addison-Wesley.

Lubbe, D. (2019). Parallel analysis with categorical variables: Impact of category probability proportions on dimensionality assessment accuracy. *Psychological Methods*, *24*(3), 339–351. https://doi.org/10.1037/met0000171

Mair, P., & Hatzinger, R. (2007a). CML based estimation of extended Rasch models with the eRm package in R. *Psychology Science Quarterly*, *49*(1), 26–43.

Mair, P., & Hatzinger, R. (2007b). Extended Rasch modeling: The eRm package for the application of IRT models in R. *Journal of Statistical Software*, *20*(9), 1–20. https://doi.org/10.18637/jss.v020.i09

Makransky, G., Lilleholt, L., & Aaby, A. (2017). Development and validation of the multimodal presence scale for virtual reality environments: A confirmatory factor analysis and item response theory approach. *Computers in Human Behavior*, *72*, 276–285. https://doi.org/10.1016/j.chb.2017.02.066

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, *57*(3), 519–530. https://doi.org/10.1093/biomet/57.3.519

Massara, G. P., Di Matteo, T., & Aste, T. (2016). Network filtering for big data: Triangulated maximally filtered graph. *Journal of Complex Networks*, *5*(2), 161–178. https://doi.org/10.1093/comnet/cnw015

Massof, R. W. (2004). Likert and guttman scaling of visual function rating scale questionnaires. *Ophthalmic Epidemiology*, *11*(5), 381–399. https://doi.org/10.1080/09286580490888771

Masters, G. N. (1982). A rasch model for partial credit scoring. *Psychometrika*, *47*(2), 149–174. https://doi.org/10.1007/BF02296272

Masters, G. N. (2016). The partial credit model. In W. J. van der Linden (Ed.), *Handbook of item response theory* (pp. 109–126). CRC Press.

Matthews, G., Hancock, P. A., Lin, J., Panganiban, A. R., Reinerman-Jones, L. E., Szalma, J. L., & Wohleber, R. W. (2021). Evolution and revolution: Personality research for the coming world of robots, artificial intelligence, and autonomous systems. *Personality and Individual Differences*, *169*, 109969. https://doi.org/10.1016/j.paid.2020.109969

McDonald, R. P. (1999). *Test theory: A unified treatment*. Lawrence Erlbaum Associates, Inc.

McDonald, R. P., & Ho, M.-H R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods*, *7*(1), 64–82. https://doi.org/10.1037/1082-989x.7.1.64

McNeish, D., & Wolf, M. G. (2021). Dynamic fit index cutoffs for confirmatory factor analysis models. *Psychological Methods*, 1–28. https://doi.org/10.1037/met0000425

Mecklin, C. J., & Mundfrom, D. J. (2007). An appraisal and bibliography of tests for multivariate normality. *International Statistical Review*, *72*(1), 123–138. https://doi.org/10.1111/j.1751-5823.2004.tb00228.x

Mokken, R. J. (1971). *A theory and procedure of scale analysis: With applications in political research*. Walter de Gruyter.

Mokken, R. J. (1997). Nonparametric models for dichotomous responses. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 351–367). Springer New York.

Molenaar, I. W. (1997). Nonparametric models for polytomous responses. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 369–380). Springer New York.

Montag, C., & Elhai, J. D. (2019). A new agenda for personality psychology in the digital age? *Personality and Individual Differences*, *147*, 128–134. https://doi.org/10.1016/j.paid.2019.03.045

Müller, M. (2020a). *iarm: Item analysis in Rasch models* (R package version 0.4.2) [Computer software manual]. https://CRAN.R-project.org/package=iarm

Müller, M. (2020b). Item fit statistics for Rasch analysis: Can we trust them? *Journal of Statistical Distributions and Applications*, *7*(1), 1–12. https://doi.org/10.1186/s40488-020-00108-7

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *The Journal of Applied Psychology*, *88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

R Core Team (2020). *R: A language and environment for statistical computing [Computer software manual]*. R Core Team. https://www.R-project.org/

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. http://www.jstatsoft.org/v48/i02/ https://doi.org/10.18637/jss.v048.i02

Saris, W. E., Satorra, A., & van der Veld, W. M. (2009). Testing structural equation models or detection of misspecifications? *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(4), 561–582. https://doi.org/10.1080/10705510903203433

Schartmüller, C., Wintersberger, P., Frison, A.-K., & Riener, A. (2019, June). Type-o-Steer: Reimagining the steering wheel for productive non-driving related tasks in conditionally automated vehicles [Paper presentation]. *2019 IEEE Intelligent Vehicles Symposium (IV)*, Paris, France (pp. 1699–1706). https://doi.org/10.1109/IVS.2019.8814088

Schauberger, G. (2019). *Gpcmlasso: Differential item functioning in generalized partial credit models* (R package version 0.1–4) [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=GPCMlasso

Schauberger, G., & Mair, P. (2020). A regularization approach for the detection of differential item functioning in generalized partial credit models. *Behavior Research Methods*, *52*(1), 279–294. https://doi.org/10.3758/s13428-019-01224-2

Schlicker, N., Langer, M., Ötting, S. K., Baum, K., König, C. J., & Wallach, D. (2021). What to expect from opening up 'black boxes'? Comparing perceptions of justice between human and automated agents. *Computers in Human Behavior*, *122*, 106837. https://doi.org/10.1016/j.chb.2021.106837

Schmettow, M., & Drees, M. (2014). What drives the geeks? Linking computer enthusiasm to achievement goals. In *Proceedings of the 28th International BCS Human Computer Interaction Conference (HCI 2014)* (pp. 189–194). BCS Learning & Development. https://doi.org/10.14236/ewic/HCI2014.37

Schomakers, E.-M., Lidynia, C., & Ziefle, M. (2019). *Listen to my heart? How privacy concerns shape users' acceptance of e-health technologies* [Paper presentation]. 2019 International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob) (pp. 306–311). https://doi.org/10.1109/WiMOB.2019.8923448

Schriesheim, C. A., Eisenbach, R. J., & Hill, K. D. (1991). The effect of negation and polar opposite item reversals on questionnaire reliability and validity: An experimental investigation. *Educational and Psychological Measurement*, *51*(1), 67–78. https://doi.org/10.1177/0013164491511005

Shi, D., Lee, T., & Maydeu-Olivares, A. (2019). Understanding the model size effect on SEM fit indices. *Educational and Psychological Measurement*, *79*(2), 310–334. https://doi.org/10.1177/0013164418783530

Shi, D., & Maydeu-Olivares, A. (2020). The effect of estimation methods on SEM fit indices. *Educational and Psychological Measurement*, *80*(3), 421–445. https://doi.org/10.1177/0013164419885164

Sijtsma, K. (2005). *Nonparametric item response theory models*. John Wiley & Sons, Ltd.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's Alpha. *Psychometrika*, *74*(1), 107–120. https://doi.org/10.1007/s11336-008-9101-0

Sijtsma, K., & Hemker, B. T. (2000). A taxonomy of IRT models for ordering persons and items using simple sum scores. *Journal of Educational and Behavioral Statistics*, *25*(4), 391–415. https://doi.org/10.2307/1165222

Sijtsma, K., Meijer, R. R., & Andries van der Ark, L. A. (2011). Mokken scale analysis as time goes by: An update for scaling practitioners. *Personality and Individual Differences*, *50*(1), 31–37. https://doi.org/10.1016/j.paid.2010.08.016

Sijtsma, K., & Pfadt, J. M. (2021). Part II: On the use, the misuse, and the very limited usefulness of cronbach's alpha: Discussing lower bounds and correlated errors. *Psychometrika*, *86*(4), 843–860. https://doi.org/10.1007/s11336-021-09789-8

Sijtsma, K., & van der Ark, L. A. (2017). February). A tutorial on how to do a mokken scale analysis on your test and questionnaire data. *The British Journal of Mathematical and Statistical Psychology*, *70*(1), 137–158. https://doi.org/10.1111/bmsp.12078

Sijtsma, K., & van der Ark, L. A. (2020). *Measurement models for psychological attributes*. CRC Press.

Sindermann, C., Schmitt, H. S., Kargl, F., Herbert, C., & Montag, C. (2021). Online privacy literacy and online privacy behavior – the role of crystallized intelligence and personality. *International Journal of Human–Computer Interaction*, *37*(15), 1–12.

Slocum-Gori, S. L., & Zumbo, B. D. (2011). Assessing the unidimensionality of psychological scales: Using multiple criteria from factor analysis. *Social Indicators Research*, *102*(3), 443–461. https://doi.org/10.1007/s11205-010-9682-8

Sousa, V. D., & Rojjanasrirat, W. (2011). Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: A clear and user-friendly guideline. *Journal of Evaluation in Clinical Practice*, *17*(2), 268–274. https://doi.org/10.1111/j.1365-2753.2010.01434.x

Stephanidis, C., Salvendy, G., Antona, M., Chen, J. Y. C., Dong, J., Duffy, V. G., Fang, X., Fidopiastis, C., Fragomeni, G., Fu, L. P., Guo, Y., Harris, D., Ioannou, A., Jeong, K-a (K.)., Konomi, S., Krömker, H., Kurosu, M., Lewis, J. R., Marcus, A., … Zhou, J. (2019). Seven HCI grand challenges. *International Journal of Human–Computer Interaction*, *35*(14), 1229–1269. https://doi.org/10.1080/10447318.2019.1619259

Straat, J. H., van der Ark, L. A., & Sijtsma, K. (2014). Minimum sample size requirements for mokken scale analysis. *Educational and Psychological Measurement*, *74*(5), 809–822. https://doi.org/10.1177/0013164414529793

Straat, J. H., van der Ark, L. A., & Sijtsma, K. (2016). Using conditional association to identify locally independent item sets. *Methodology*, *12*(4), 117–123. https://doi.org/10.1027/1614-2241/a000115

Straat, J. H., Hendrik Straat, J., van der Ark, L. A., & Sijtsma, K. (2013). Comparing optimization algorithms for item selection in Mokken scale analysis. *Journal of Classification*, *30*(1), 75–99. https://doi.org/10.1007/s00357-013-9122-y

Suárez-Álvarez, J., Pedrosa, I., Lozano, L. M., García-Cueto, E., Cuesta, M., & Muñiz, J. (2018). Using reversed items in likert scales: A questionable practice. *Psicothema*, *30*(2), 149–158.

Tajudeen, F. P., Bahar, N., Maw Pin, T., & Saedon, N. I. (2021). Mobile technologies and healthy ageing: A bibliometric analysis on publication trends and knowledge structure of mhealth research for older adults. *International Journal of Human–Computer Interaction*, *38*(2), 1–13.

Taufik, D., Kunz, M. C., & Onwezen, M. C. (2021). Changing consumer behaviour in virtual reality: A systematic literature review. *Computers in Human Behavior Reports*, *3*, 100093. https://doi.org/10.1016/j.chbr.2021.100093

Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley Publishing Company.

van der Ark, L. A. (2005). Stochastic ordering of the latent trait by the sum score under various polytomous IRT models. *Psychometrika*, *70*(2), 283–304. https://doi.org/10.1007/s11336-000-0862-3

van der Ark, L. A. (2007). Mokken scale analysis in R. *Journal of Statistical Software*, *20*(11), 1–19.

van der Ark, L. A. (2012). New developments in Mokken scale analysis in R. *Journal of Statistical Software*, *48*(5), 1–27.

van der Ark, L. A., & Bergsma, W. P. (2010). A note on stochastic ordering of the latent trait using the sum of polytomous item scores. *Psychometrika*, *75*(2), 272–279. https://doi.org/10.1007/s11336-010-9147-7

van der Ark, L. A., Rossi, G., & Sijtsma, K. (2019). Nonparametric item response theory and mokken scale analysis, with relations to latent class models and cognitive diagnostic models. In M. von Davier & Y. -S. Lee (Eds.), *Handbook of diagnostic classification models: Models and model extensions, applications, software packages* (pp. 21–45). Springer International Publishing.

van Laar, E., van Deursen, A. J. A. M., van Dijk, J. A. G. M., & de Haan, J. D. (2017). The relation between 21st-century skills and digital skills: A systematic literature review. *Computers in Human Behavior*, *72*, 577–588. https://doi.org/10.1016/j.chb.2017.03.010

Wensing, M., Paech, B., Roth, C., & Schwill, S. (2019). Learning, understanding and the use of information technology: A survey study among primary care physician trainees. *BMC Health Services Research*, *19*(1), 728. https://doi.org/10.1186/s12913-019-4615-y

Wessel, D., Attig, C., & Franke, T. (2019, September). ATI-S - an Ultra-Short scale for assessing affinity for technology interaction in user studies. In *Proceedings of Mensch und Computer 2019* (pp. 147–154). Association for Computing Machinery. https://doi.org/10.1145/3340764.3340766

Wetzel, E., & Carstensen, C. H. (2014). Reversed thresholds in partial credit models: a reason for collapsing categories? *Assessment*, *21*(6), 765–774. https://doi.org/10.1177/1073191114530775

Whitley, B. E. (1997). Gender differences in computer-related attitudes and behavior: A meta-analysis. *Computers in Human Behavior*, *13*(1), 1–22. https://doi.org/10.1016/S0747-5632(96)00026-X

Woods, C. M. (2006). Careless responding to Reverse-Worded items: Implications for confirmatory factor analysis. *Journal of Psychopathology and Behavioral Assessment*, *28*(3), 186–191. https://doi.org/10.1007/s10862-005-9004-7

Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the Three-Parameter logistic model. *Applied Psychological Measurement*, *8*(2), 125–145. https://doi.org/10.1177/014662168400800201

Yuan, K.-H., Liu, H., & Han, Y. (2021). Differential item functioning analysis without a priori information on anchor items: QQ plots and graphical test. *Psychometrika*, *86*(2), 345–377. https://doi.org/10.1007/s11336-021-09746-5

Zhang, X., Noor, R., & Savalei, V. (2016). Examining the effect of reverse worded items on the factor structure of the need for cognition scale. *PLoS One*, *11*(6), e0157795. https://doi.org/10.1371/journal.pone.0157795

Zijlmans, E. A. O., Tijmstra, J., van der Ark, L. A., & Sijtsma, K. (2018). Item-Score reliability as a selection tool in test construction. *Frontiers in Psychology*, *9*, 2298.

Zwitser, R. J., & Maris, G. (2016). Ordering individuals with sum scores: The introduction of the nonparametric rasch model. *Psychometrika*, *81*(1), 39–59. https://doi.org/10.1007/s11336-015-9481-x

About the Authors

Ville Heilala is a doctoral researcher at the Faculty of Information Technology at the University of Jyväskylä. He holds two Master's degrees, one in education and one in computer science. His research relates to understanding human learning using computational methods.

Riitta Kelly is an English teacher at the Centre for Multilingual Academic Communication, University of Jyväskylä, Finland. She is also a doctoral researcher in Applied Linguistics at the University of Jyväskylä.

Mirka Saarela is a research fellow at the Faculty of Information Technology at the University of Jyväskylä. Her research interests combine machine learning, explainable artificial intelligence, cognitive computing, and education.

Päivikki Jääskelä is an Adjunct Professor and a Senior Researcher at the Finnish Institute for Educational Research at the University of Jyväskylä, Finland. Her research focuses on university student agency, learner-centered pedagogy, and teacher development in higher education.

Tommi Kärkkäinen has been serving as a full professor of Mathematical Information Technology at the Faculty of Information Technology at the University of Jyväskylä since 2002. He has led almost 50 R&D projects, supervised over 50 Ph.D. students, and published over 180 peer-reviewed articles.

Appendix A. Translations of the introductory text and Likert options

The original ATI scale, the items, and the introductory text can be found from the original scale developers' article (Franke et al., 2019) and from their website.

Introductory text

"Kysymme sinulta teknisten järjestelmien käyttämisestä. Tekninen järjestelmä viittaa tässä yhteydessä kaikkiin appeihin, sovelluksiin ja myös varsinaisiin laitteisiin (esim. kännykkä, tietokone, televisio, auton navigaattori). Ole hyvä ja kerro missä määrin olet samaa tai eri mieltä seuraavien väitteiden kanssa."

Likert options

täysin eri mieltä (1), enimmäkseen eri mieltä (2), hieman eri mieltä (3), hieman samaa mieltä (4), enimmäkseen samaa mieltä (5), täysin samaa mieltä (6)

Appendix B. Properties of the scale items and the total scale score

Table B1. Coefficients of homogeneity for item pairs H_{jk} and individual items H_j without outliers and with outliers$^{(\dagger)}$.

Item	H_{jk}	$H_{jk}{}^{\dagger}$	H_j (SE)	H_j (SE)†
ATI1	0.43–0.80	0.39–0.76	0.66 (0.014)	0.62 (0.016)
ATI2	0.45–0.82	0.42–0.80	0.64 (0.016)	0.60 (0.017)
ATI3R	0.35–0.47	0.30–0.45	0.43 (0.027)	0.40 (0.027)
ATI4	0.47–0.81	0.42–0.78	0.64 (0.015)	0.60 (0.017)
ATI5	0.47–0.82	0.45–0.80	0.66 (0.014)	0.62 (0.016)
ATI6R	0.44–0.67	0.39–0.65	0.56 (0.020)	0.52 (0.022)
ATI7	0.35–0.67	0.30–0.65	0.58 (0.018)	0.54 (0.020)
ATI8R	0.46–0.57	0.42–0.57	0.52 (0.023)	0.48 (0.024)
ATI9	0.38–0.68	0.34–0.65	0.56 (0.021)	0.51 (0.022)

Homogeneity of the whole scale was $H = 0.58$ (0.54†); $SE = 0.015$ (0.016†).

Table B2. Total ATI scale score in different groups based on gender and faculty.

Group	*n*	*M/9*	*M*	*SD*	*min*	*max*	G_1	G_2	*SE*
All	757	3.7	33.3	9.2	9	54	−0.16	−0.55	0.34
Gender: Women	494	3.5	31.2	8.7	9	52	0.00	−0.50	0.39
Gender: Nonbinary	6	4.0	36.0	7.9	23	44	−0.75	0.32	3.24
Gender: Men	251	4.1	37.3	8.9	9	54	−0.62	0.18	0.56
Faculty: Hum & Soc.Sci	103	3.3	29.7	9.2	10	49	0.02	−0.88	0.90
Faculty: Sports	106	3.3	29.9	8.0	14	51	0.31	0.18	0.78
Faculty: Edu	85	3.4	30.6	8.8	9	51	0.00	−0.16	0.95
Faculty: Natural	162	3.6	32.6	9.5	9	53	−0.17	−0.56	0.75
Faculty: Business	76	3.8	33.8	8.6	11	50	−0.33	0.16	0.99
Faculty: Other	11	4.0	35.7	12.7	15	54	−0.20	−1.08	3.84
Faculty: IT	214	4.2	38.2	7.7	17	53	−0.40	−0.40	0.52

M is the mean of the composite sum score and M/9 is the mean Likert value.

Appendix C. Partial credit model

Parametric item response theory

The partial credit model (PCM) (Masters, 1982)—an extension of the dichotomous Rasch model for polytomous data—is the simplest of all polytomous IRT models widely used in various measurement and assessment scenarios (Masters, 2016, p. 110). Bond et al. (2020, p. 238) define a Rasch model as "a theoretical mathematical description of how fundamental measurement should operate with social/psychological variables". They continue that "no real, empirical data will ever fit Rasch's theoretical ideal," however, the question is more about how closely the model supports the measurement decisions one wants to make (Bond et al., 2020). One important decision a researcher usually wants to make is whether the respondents can be ordered based on their total score. While PCM is a strict model, it is the least strict of models that still have the property of stochastic ordering of latent variable by the total scale score (i.e., SOL by X_+) (Hemker et al., 1997; Ligtvoet, 2012; Sijtsma & Hemker, 2000; van der Ark, 2005). PCM has been used to analyze scales relating to, for example, attitudes (e.g., Aghekyan, 2020), technology interaction (e.g., Makransky et al., 2017), and behavioral intent (e.g., Chen & Jin, 2020).PCM estimates two parameters: the ability of the person in the latent variable and the item difficulty representing the locations of the item at a point in the latent variable continuum where it is equally likely to choose either of the extreme categories (Masters, 2016, p. 111). Another characteristic of PCM is that the category thresholds are allowed to vary between items, which in turn can give valuable information about the functioning of the response categories between items (Wetzel & Carstensen, 2014). The PCM model was fitted in *R* using *eRm* package (Mair & Hatzinger, 2007b) and conditional maximum-likelihood (CLM) procedure, which has mathematical and epistemological advantages (Mair & Hatzinger, 2007a).Unidimensionality was assessed using principal component analysis of residuals (PCAR). Under unidimensionality, it is expected that the partial credit model explains all the variance in the data, and the residuals represent random noise. Thus, unidimensionality could be supported if all eigenvalues obtained from the PCA of the residuals are less than two and there are no contrasting item loadings on the first component (Bond et al., 2020, p. 254). Local dependency (LD) was evaluated using Q_3 statistics, which is the Pearson correlation between the raw item residuals (Yen, 1984). The difference between the largest observed correlation with the average of the observed correlations denoted as $Q_{3,*}$ greater than 0.20 could indicate LD (Christensen et al., 2017).The goodness of fit at the item level was assessed using unweighted mean-square (i.e., outfit) and weighted mean-square (i.e., infit) statistics. Optimal infit and outfit values are close to 1.0, whereas statistically significant values greater than 1.0 indicate underfit (i.e., unpredictability and excess variation in the model), and values less than 1.0 indicate overfit (i.e., deterministic response patterns and less variation in the model) (Bond et al., 2020, p. 241). Overfitting can occur, for example, in the case of item redundancy, and it can cause smaller standard errors, inflated reliability, and local dependency (Bond et al., 2020, p. 241). On the other hand, underfit degrades the measurement quality and is in practice more important to diagnose than overfit (Bond et al., 2020, p. 241). Outfit and infit statistics were calculated based on the conditional residuals (Müller, 2020b) using *iarm* package (Müller, 2020a).We examined the category probability curves visually to find out possible disordered categories (i.e., reversed deltas or disordered thresholds). The ordering of category thresholds should follow the ordering of the actual response categories. Disordering of the categories can occur, for example, when a category is not the most likely category at any point along the latent variable (Wetzel & Carstensen, 2014). Disordered categories could indicate dependence among the underlying items; however, rash judgments should be avoided as disordering does not necessarily imply that the item is not functioning as expected (Adams et al., 2012). Another visual representation, person-item map (a.k.a, Wright map), depicts the respondents in terms of their ability and items in terms of their

Table C1. Item fit indices of the PCM ordered by the item location (δ_i).

Item	δ_i	Outfit	Infit
ATI9	0.25	1.06	1.03
ATI3R	0.44	1.81***	1.71***
ATI2	0.47	0.79**	0.79**
ATI4	0.50	0.75***	0.74***
ATI5	0.67	0.69***	0.66***
ATI7	0.99	1.05	1.02
ATI1	1.00	0.72***	0.70***
ATI6R	1.32	1.12	1.12
ATI8R	1.62	1.29***	1.24***

Infit and outfit $SE = 0.05$ for all items.
Note. Bonferroni adjusted.
$^{***}p < 0.001$, $^{**}p < 0.01$.

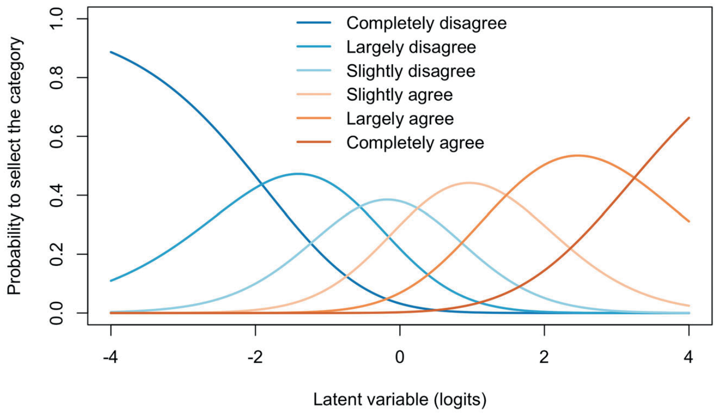


Figure C1. Item characteristic curve of the item ATI4 showed an ordered category structure. Similar ordering was found in items ATI1, ATI8R, and ATI9.

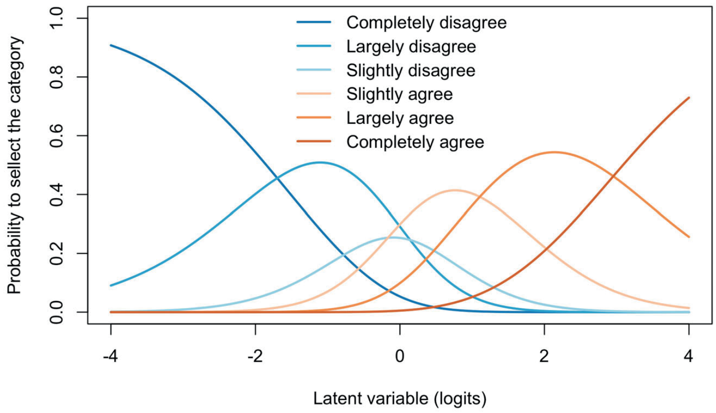


Figure C2. Item characteristic curve of the item ATI2 showed a slightly disordered category structure. Similar ordering was found in items ATI3R and ATI5.

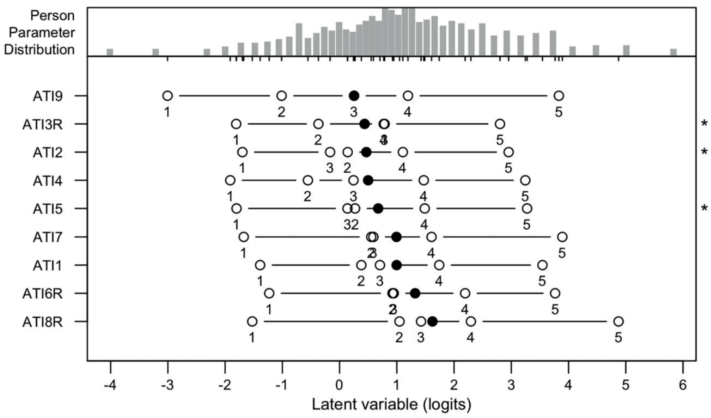


Figure C3. Person-item map of the PCM where items are ordered based on their location. Numbered circles represent the intersection points of adjacent item characteristic curves. Asterisk on the right marginal marks the items exhibiting category disordering.

location and thresholds along the latent variable continuum depicted as logit scale (Bond et al., 2020). Thus, a person-item map can be used to evaluate, for example, ordering the response categories and the coverage of the items along with the latent variable.

Results

All results for PCM were similar with and without outliers, and results without outliers are reported here. In the post-hoc analysis, the assumption of unidimensionality was assessed using PCA of the residual correlations. The largest eigenvalue was 2.19, which was slightly greater than the suggested threshold value of 2. Further examination showed contrasting item loadings. RW items and ATI7 showed negative loadings, and all other items showed positive loadings on the largest component. The results indicated that the residuals possibly contained unexplained variance, which would compromise the assumption of unidimensionality from the PCM point of view. Residual correlations ($\bar{Q}3 = -0.12$, $Q_{3,sd} = 0.16$, $Q_{3,min} = -0.36$, $Q_{3,max} = 0.23$, $Q_{3,*} = 0.35$) indicated the existence of local dependency. Seven item pairs out of 36 exhibited $Q_{3,*} > 0.20$, and two item pairs, ATI2–ATI4 and ATI2–ATI5, had $Q_{3,*} > 0.30$. The effect of including outliers was minimal for both assessing unidimensionality and the local dependency.Item fit indices (Table C1) indicated significant outfit and infit for all other items except ATI6R, ATI7, and ATI9. Significant overfit could be expected because of item content similarity. However, only items ATI3R and ATI8R showed significant underfit, which can degrade the measurement results (Bond et al., 2020, p. 241). For the item ATI3R, the misfit was extreme, indicating unpredictability and excess variation. The unpredictable pattern of the item ATI3R was also noticeable in the scalogram (Figure 3) and in other previous analyses.

Items ATI1, ATI4, ATI8R, and ATI9 showed a plain and ordered category structure (Appendix Figure C1). On the other hand, items ATI2, ATI3R, and ATI5 showed disordered thresholds between two adjacent categories (Appendix Figure C2), and items ATI6R and ATI7 showed one extremely narrow category. Notably, the middle response categories *Slightly disagree* and *Slightly agree* showed disordering or highly narrow range in the latent variable (Appendix Figure C3), indicating it was more probable to prefer the extreme categories.

The person-item map showed that the items cover a wide range of the latent variable continuum (Appendix Figure C3). Items ATI4 and ATI9 showed relatively evenly distributed category thresholds. Item ATI8R proved to be the most challenging item to agree with, providing information from the higher end of the continuum. Especially, item ATI9 showed several desirable properties: it covers a wide range of the latent variable with ordered and relatively equal thresholds, it has a central location in the middle of the continuum, and its item fit statistics are adequate.

Appendix D. Results for data including outliers

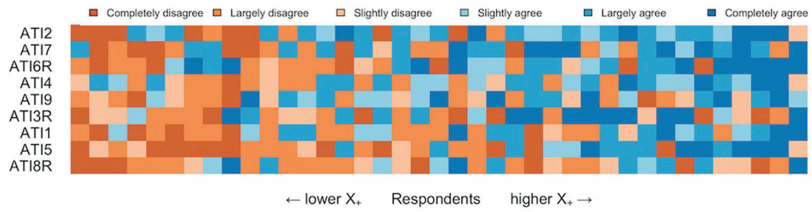


Figure D1. Scalogram containing only outliers ($n = 39$) suggested by MMCD75, respondents ordered by total score, and items ordered by total item sum.

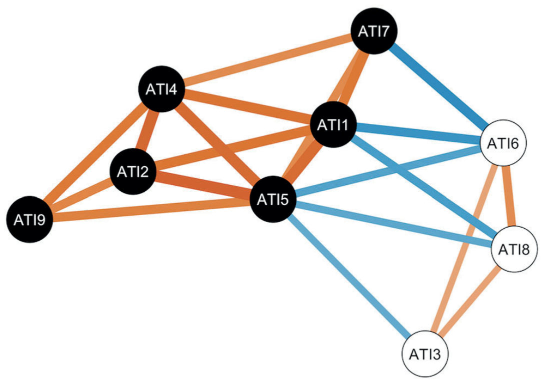


Figure D2. EGA for the complete data including outliers.