Ioana Jivet, Daniele Di Mitri, Jan Schneider, Zacharoula Papamitsiou, Mikhail Fominykh (Eds.)

Proceedings of the Doctoral Consortium of the Seventeenth European Conference on Technology Enhanced Learning

co-located with the Seventeenth European Conference on Technology Enhanced Learning (EC-TEL 2022)

Toulouse, France, September 12, 2022

Preface

The EC-TEL Doctoral Consortium is part of the EC-TEL since its beginning in 2005 and is part of the doctoral program of the European Association of Technology Enhanced Learning (EA-TEL). Besides the Doctoral Consortium, this program includes the JTEL Summer School. Together, these two events have been shaping and enriching the experiences of many young researchers on their PhD journey, leading to a community that addresses the transdisciplinary challenges of our field.

This volume contains papers presented at the Doctoral Consortium of the Seventeenth European Conference on Technology Enhanced Learning (EC-TEL 2022) held on September 12, 2022, in Toulouse, France. There were 12 proposals submitted and reviewed by at least two doctoral consortium program committee members and one doctoral candidate (peer). All 12 proposals were found eligible to be presented at the Doctoral Consortium event and were accepted to be published as full papers in the Doctoral Consortium proceedings. The papers in this volume completed the full reflection cycle that is characteristic of the EC-TEL Doctoral Consortium.

The EC-TEL Doctoral Consortium is designed as a training event for PhD candidates that seeks to improve the quality of their research. Junior research students are offered a comfortable networking-oriented space to present their theses advances, discuss their research plans with peers and more experienced researchers and further improve their writing and presentation skills. Besides an emphasis on gathering constructive feedback, the Doctoral Consortium provides a platform to explore and reflect on issues regarding methodology, research-related topics, supervision and career-related aspirations. This process starts with expressing the own research project and identifying the limitations and challenges of the present stage. The initial submissions were reviewed by the program committee. To strengthen the learning experience, every doctoral candidate had to review one other submission. The submission provides the foundation for the presentations at the Doctoral Consortium. The day was structured into four thematic sessions, each session consisting of presentations from three candidates followed by breakout groups in which each candidate received detailed feedback. Participants also had the opportunity to present their work as a poster at the main conference, which creates a unique opportunity to engage in discussions with the wider research community attending the conference. The PhD candidates received timely and relevant feedback, which needed to be addressed when preparing for the proceedings. All papers included in these proceedings have been reworked to address the comments of the reviewers and the participants of the event.

The submissions to this year's doctoral consortium show the continued relevance for PhD candidates to get qualitative feedback on their projects beyond the level of research papers. Receiving submissions from countries across and beyond Europe, the EC-TEL Doctoral Consortium shows its international relevance and potential impact. The variety of topics with both technological and educational focus represented at the doctoral consortium once again highlights the highly multidisciplinary nature of the TEL field. This is complemented by EATEL's activities for building the doctoral community including the series of webinars organized by the DETEL EU project.

We would like to express our gratitude to all senior researchers involved in the reviewing process. Especially, we would like to thank them for attending the Doctoral Consortium and for their valuable comments during the event. Finally, we are grateful to the conference organisers for supporting this event and for arranging the logistical and administrative side of this Doctoral Consortium.

We wish all PhD candidates a rewarding and productive continuation of their PhD journey.

November 2022

Ioana Jivet
Daniele Di Mitri
Jan Schneider
Zacharoula Papamitsiou
Mikhail Fominykh

Program Committee of the Doctoral Consortium

Torsten Zesch, FernUniversität in Hagen, Germany

Mar Perez-Sanagustin, Université Paul Sabatier Toulouse III, France

Andrea Horbach, University of Duisburg-Essen, Germany

Angela Fessl, Know-Center GmbH, Austria

Perry den Brok, Wageningen University and Research, The Netherlands

Renate Klaassen, Delft University of Technology, The Netherlands

Wilma Elston, Delft University of Technology, The Netherlands

Clara Schumacher, Humboldt Universität zu Berlin, Germany

Wilma Elston, Delft University of Technology, The Netherlands

Olga Viberg, KTH Royal Institute of Technology, Sweden

Laura von Rüden, Fraunhofer IAIS, Germany

Vitomir Kovanovic, The University of South Australia, Australia

Sebastian Dennerlein, University of Twente, The Netherlands

Ryan Baker, University of Pennsylvania, USA

Mutlu Cukurova, University College London, UK

Ioana Jivet, Goethe University Frankfurt & DIPF, Germany

Daniele Di Mitri, DIPF, Germany

Jan Schneider, DIPF, Germany

Zacharoula Papamitsiou, SINTEF, Norway

Contents

Learning analytics to support teachers in the challenge of overcoming the learning gaps in k-12 students

Erverson B. G. de Sousa[1], Rafael Ferreira Mello[1,2]

[1]Cesar School, Recife, Brazil

[2]Departamento de Computação, Universidade Federal Rural de Pernambuco, Recife, Brazil

Abstract

The emergency remote teaching caused by the covid-19 pandemic has potentiated the learning gaps of several students in Brazilian education, especially in the K-12 settings. Amidst the many challenges imposed by the pandemic, the adoption of digital tools in the school context has provided the generation of educational data, which can be collected and analyzed in order to provide evidence-based decision making, taking into account all the stakeholders in the teaching and learning process. Such decisions can provide for the personalization of learning, which aims to provide the student with educational resources that promote the building of weakened skills caused by learning gaps. The present thesis plan aims to present the work plan for the development of a Learning Analytics Dashboard tool for teachers in a basic education school in order to support data-driven pedagogical decision-making and to enable personalized monitoring of learning

Keywords

Learning analytics, learning gaps, k-12 students, personalized learning

1. Introduction

Schools have been facing new challenges due to the worldwide COVID-19 pandemic, which has been impacting, mainly, the teaching and learning process [1]. Many school institutions, even without the proper time and resources, had to migrate classes to the digital world, through educational apps and platforms [2]. The abrupt adoption of remote teaching evidenced the socio-educational precariousness of several countries, including Brazil. The strategy adopted to continue teaching did not reach some students and teachers, due to the context of social vulnerability [2] [3].

In the scenario of basic education, which comprises the levels of education from kindergarten to high school, such difficulties are potentiated, as it comprises one of the most important periods for students in this age group, the literacy process and the construction of mathematical skills, that serve as a base throughout their school career [4] [5]. These experiences and knowledge were completely affected due to adaptation to the new remote teaching scenario, increasing the learning gap that already existed in the context of Brazilian education [6] [7].

In this context, school management has the fundamental role of dealing not only with issues of improving educational indices, but also with concerns related to the physical and emotional health of its professionals, students and family members [2].This new reality, linked to the challenge of transposing face-to-face classes to the virtual environment, adds to the role of the school manager, who must take into account the current socio-educational reality, and improve his decision-making process to achieve the goals of the school. educational institution [3].

Amid so many challenges imposed by the pandemic, the adoption of digital tools in the school context has provided the generation of educational data, which can be collected and analyzed in order to provide evidence-based decision-making, taking into account all parties. interested in the teaching and learning process [8]. Decision making is a task that is part of the daily routine of a school manager, as well as the teacher who deals directly with the student and through his actions directly impacts student learning.

The data generated from student interactions with digital tools can be used to monitor, analyze, predict, intervene, recommend and, above all, improve the quality of the teaching and learning process [9].These are features recommended in teaching practice, which aims to maintain a personalized teaching and learning process, to meet the specific needs of each student.

Learning Analytics (LA) is an emerging field that addresses this context of educational data analysis, whose objective is the collection, analysis and reporting of data about students and the contexts in which they occur [10].Supporting the decision-making of managers, coordinators, teachers and other stakeholders in student learning. There have been applications of LA techniques in the context of basic education, among these applica-

Proceedings of the Doctoral Consortium of Seventeenth European Conference on Technology Enhanced Learning, September 12-16, 2022, Toulouse, France

ebgs@cesar.school (E. B. G. d. Sousa); rafael.mello@ufrpe.br (R. F. Mello)

0000-0002-6474-9606 (E. B. G. d. Sousa); 0000-0003-3548-9670 (R. F. Mello)

tions are those that enable data-based decision making for teachers and other stakeholders, especially with regard to personalization of the teaching and learning process [11]. On the other hand, most efforts in the use of LA are focused on higher education [12] [13] [14], lacking more research and tools that meet the specific demands of basic education schools [15].

1.1. Main goal

Develop a Learning Analytics tool for teachers of k-12 education school in order to support pedagogical decision-making based on data and enable personalized monitoring of learning.

1.2. Specific objectives

- Collect studies that used Learning Analytics in the context of basic education;
- Identify the demands of an elementary school to adopt Learning Analytics at an institutional level;
- Investigate the main problems and challenges associated with the adoption of Learning Analytics in the context of elementary schools;
- Design of a Learning Analytics tool to monitor students' learning progress;
- Conduct an evaluation of the adoption of the proposed tool in a primary school and verify if it supports teachers in how to deal with student learning gaps through personalization of teaching.

1.3. Research questions

1.3.1. Main question

How to support a k-12 education school to deal with the challenges arising from learning gaps, through Learning Analytics techniques?

1.3.2. The main question is divided into four sub-questions

- How can educational data analysis help address learning gaps?
- How to deal with ethical issues in the adoption of Learning Analytics in k-12 education?
- How can learning analytics techniques support the personalized monitoring of student learning in k-12 education?
- What is the context of a k-12 education school to adopt Learning Analytics?

2. State of the art

Data-based decision-making is an essential action in the school context, taking into account the growing generation of educational data provided by student interactions with digital tools, as well as traditionally existing data, such as grades and attendance [9].Within this context, educational performance indicators also play an important role in decision-making with the aim of improving the teaching and learning process. All data and information from the various sectors that make up the school can and should be used in order to provide insights and support decision-making in a timely manner with the aim of promoting personalization and learning recovery.

2.1. Background and context

2.1.1. Personalized monitoring of learning in blended learning

Due to the global pandemic of covid-19, schools had to adapt the way they interact with students, starting to use more disruptive teaching-learning tools and strategies to guarantee remote classes, such as Google Meet and Zoom, which were used. to enable synchronous classes [16]. With regard to educational strategies, there was a massive adoption of hybrid teaching models, since, soon after the beginning of the vaccination period for health professionals, many schools began to partially return to the face-to-face model, adopting this approach [17].

Blended learning is "a model of formal education that is characterized by merging two modes of teaching: traditional and online, also valuing interaction and collective and collaborative learning" [18]. In the systematic review conducted by [19], where they investigated what types of hybrid learning models exist, having found six types. The following models were found: supplementary, inverted classroom, rotational laboratory, study rotation, synchronous collaborative hybrid and dual-collaborative group.

With a different approach to the traditional teaching model, blended learning has as its specific characteristic a more personalized learning, respecting the students' own pace and understanding that people learn in different ways [20]. Based on this understanding, it is possible to offer students learning that addresses their learning gaps and can enable students to learn more individually and effectively.

2.1.2. Application of Learning Analytics in k-12 Education

Learning Analytics (LA) is an emerging research field that aims to measure, collect, analyze and report data about students and their contexts, as well as understand

and optimize learning and the environment in which it takes place [21].

Despite substantial growth in the application of learning analytics to improve teaching and learning in the last decade, most of these works have focused on higher education issues and contexts [22] [23]. With a wider adoption of digital educational technologies in primary education, recently accelerated by remote emergency classes due to Covid-19, there has been a greater awareness of using LA to accompany and personalize the learning process [24].

In the systematic review of the literature conducted by [11] 42 studies were identified that applied LA techniques in the context of basic education, among the approaches are, data distillation for human judgment, prediction, educational data mining, discovery with models and clustering. Most of these approaches developed isolated works in some sector of the school, but did not address the context of adopting learning analytics in an institutional way. Some aspects must be considered with regard to the use of AL in basic education, an institutional diagnosis must be taken into account to understand the needs of the context of a particular school, the ethical issues that are generated must be taken into account. from the use of the data, it is necessary to use more diversified techniques that take into account, mainly, the personalized accompaniment of the learning and, finally, to use explainability techniques (Explainable artificial intelligence) in the algorithms used to support students, teachers and educational managers [11].

2.1.3. Use of Learning Analytics in formative assessment to support data-driven pedagogical decision making

Assessment has three general functions: diagnose, control and classify. These three functions are represented, respectively, by the types of existing assessments: diagnostic, formative and summative.The diagnostic evaluation, according to [25] , “the fundamental objective is to analyze the situation of each student before starting a certain teaching-learning process, to become aware of the starting points, and to adapt the process to the detected needs”. The summative assessment, on the other hand, takes into account all the content taught, usually divided by two months, and at the end of this process, a test is carried out to verify the acquisition of knowledge [26].

Formative assessment aims to monitor students’ learning during classes, in daily activities and is concerned with ”determining the degree of mastery of a given learning task and indicating the part of the task not mastered” [27]. Unlike summative assessment, the focus of formative assessment is to collect data to reorient the teaching and learning process, pointing out its weaknesses, allowing necessary changes during the school period, in daily practice [28].

In view of the objective of formative assessment, which aims to accompany students, collecting evidence of their learning process, learning analytics techniques can be used to deal with the measurement, collection, analysis and reporting of these collected data, enabling teachers and other stakeholders valuable information about students during the construction of their knowledge, taking into account the individual learning pace. Through these data, it is impossible to make a pedagogical decision based on data, since with the increasing use of educational technologies, in the context of basic education, more data is generated during student interaction made possible through formative assessments made available by teachers daily.

Support for teachers’ decision-making has gained a lot of notoriety in recent studies in the area of learning analytics for basic education [29]. And with regard to personalized monitoring of learning, through the increasing use of digital technologies in basic education, it is possible to empower teachers to deal with problems arising from lag and learning gaps, and help students to recover their learning [30].

2.2. Related works

In order to verify the importance of the challenges presented in this thesis plan, a systematic review of the literature was carried out, in order to obtain the state of the art of publications that addressed the application of learning analytics in the context of high school, and later it was A survey of studies was carried out, directly from the databases, to update these studies, as well as to identify studies that addressed the use of learning analytics in the context of basic education as a whole to support teachers and/or managers in pedagogical decision-making based on in data. Taking this context into account, some studies were identified that aimed to address issues similar to this thesis plan.

The work done by [9] uses the various data generated by educational information systems, such as: learning management systems or virtual learning environment, student diary, library system, digital repository, etc. The authors address that due to the use of these digital tools in the school context, there has been a significant increase in the volume and variety of data that can be captured, stored and analyzed in order to improve student learning and school effectiveness. In this study, they took a comprehensive approach to the use of learning analytics in Bulgarian education, and developed six machine learning models to support decision-making based on data from stakeholders in that context. The models were developed to support students, teacher monitors, classroom teachers, administrators, parents and educational inspectors. Four models were evaluated, for students, monitor teachers, classroom teachers and parents, and showed promising results.

In the project developed by [31] learning analytics dashboards were developed to help teachers make quick and effective decisions regarding student learning activities in the classroom. The proposed dashboard was enhanced taking into account the needs of teachers, with a user experience and usability suitable for teaching practice, taking into account the dynamics presented in basic education. An important feature of this study is the provision of information through real-time dashboards to speed up teachers' decision making. The dashboard presented for the educational context was originally developed for the business context, however it was adapted to be used by teachers. The final prototype was evaluated by 9 teachers, and it was found to have a high potential to support pedagogical decision-making. As a point of improvement, the teachers participating in the dashboard evaluation pointed out the need to use data from external tools, which are already part of the school context.

The work led by [32] addressed an experiment carried out with five high school teachers, who were monitored during a school year. Teachers used information provided by learning analytics in their classrooms through data provided from computer-based assessments. Such information served as a basis for the planning of classes, which enabled a more individualized and personalized teaching-learning approach. Teachers reported that the insights extracted from the data collaborated in their teaching practice, highlighting the detailed information about each student, task and responses. In the classroom, teachers used such detailed insights to provide feedback to low-performing students and it was found that those students who had these learning gaps performed better after performing the data intervention.

Finally,[33] investigated the role of learning analytics to assess formative assessments, with the aim of using a data-driven approach to inform teachers about changes in their teaching practices and how they impact the development of student learning. The authors highlighted that one of the most challenging tasks for teachers is designing, managing and evaluating formative assessments, and this is one of the main reasons for not using formative assessments as a form of feedback for students and for teachers themselves to adjust their teaching strategies throughout the school year. One of the ways to overcome these challenges, according to the authors, was the use of learning analytics techniques that were employed in the study with the purpose of facing such difficulties and providing personalized feedback on a large scale. Briefly, the data collected from formative assessments were analyzed using learning analytics and provided recommendations that supported students in a self-regulated learning approach, and enabled teachers to reorient their planning and teaching practices.

Taking into account these approaches, how to use learning analytics techniques in formative assessments, in the context of blended learning, to collect, analyze and report educational data for teachers in a basic education school to monitor the learning process and support the decision-making process. data-driven pedagogical decision-making to help students address their learning gaps?

3. Work plan

The purpose of this work is to develop a descriptive and predictive Learning Analytics Dashboard - LAD, using data collected from Google Classroom and Khan Academy, to support teachers in pedagogical decision making in the classroom, in order to identify, monitor and propose interventions to deal with students' learning gaps.

3.1. Submission of the proposal

The proposal is to use data from computer-based assessments, with the support of two educational platforms, Google Classroom and Khan Academy, which are used to manage classes and activities in the context of online and face-to-face classes.

The APIs (Application Programming Interface) provided by both platforms will be used to access and form the data repository, which will serve as input for the construction of the LAD. The availability of descriptive and predictive data analysis through LADs is the most common way to fulfill the Learning Analytics cycle, which has as a crucial objective, in addition to measuring, collecting and analyzing educational data, to provide reports on this data. from student interactions on digital educational platforms, enabling teachers to make evidence-based pedagogical decision-making [34].

Through the use of LAD, teachers will be able to track student performance in real time on the Khan Academy and Google Classroom platform, as well as have access to predictive results based on student interactions. In the context of blended learning, using the rotational laboratory approach, students participate, in addition to the traditional classroom lesson, they also interact with digital devices, where they will have the purpose of continuing the class started in the classroom. Among the most common activities carried out by students are: research on the internet, answering online activities, developing individual or collaborative textual productions, etc.

Students will use Chromebook devices to carry out classes through the rotational laboratory approach, which are notebooks that use the Chrome OS operating system and are generally used in the school context. With

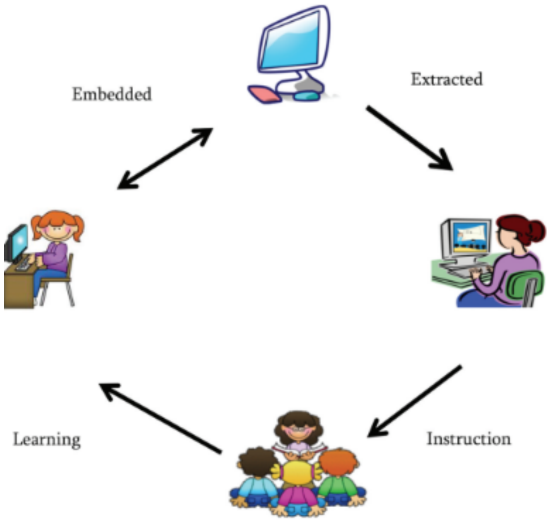

Figure 1: Embedded and extracted analytics in computer-based assessments.

the chromebooks, data regarding online activities will be collected through the Khan Academy API, and data regarding student interactions in the classroom will be obtained through the classroom API and Google Chrome's Sync function, which tracks logs from the browser.

LAD will support the teacher in decision making while students are working on an assignment, in the classroom, and will provide in real time which students will need support and what kind of support will be needed.

In order to support computer-based formative assessment, through Classroom and Khan Academy, for the construction of the LAD, the Learning Analytics - extracted analytics strategy will be used. There are two types of Learning Analytics strategies, the embedded analytics which refers to the data that is used to inform the student and/or adapt tasks to the students' skill levels without teacher intervention. And extracted analytics refers to the data that is presented for interpretation and provides teachers with information about the learning process and its results, where it is possible to personalize teaching and learning in the classroom [35]. Figure 1 illustrates the difference between the two approaches.

3.2. Method

The present work will use applied research as a type of study, which according to [36] are "research aimed at acquiring knowledge with a view to applying it in a specific situation", where the need to produce knowledge for the application of its results is the motivation to "contribute to practical ends, aiming at the immediate solution of the problem encountered in reality" [37].

For research purposes, it is characterized as descriptive, as it aims to describe the characteristics of certain populations or phenomena and "can also be elaborated with the purpose of identifying relationships between variables" [36]. Its approach will be qualitative for the analysis of research data, according to Gil (2002, p. 133) "qualitative analysis depends on many factors, such as the nature of the data collected, the size of the sample, the research instruments and the theoretical assumptions that guided the investigation" [38].

Regarding the technical procedure, the study is classified as action research, which is defined as a type of empirically based research and has a "close association with an action or with the resolution of a collective problem and in which researchers and representative participants of the situation or problem are involved in a cooperative or participatory way" [39].

In order to answer the research questions and achieve the objective of this study, the data collection tools will be the use of forms, observations and interviews with teachers of the Portuguese Language and Mathematics subjects, of elementary school 2, of the Escola Professor Olindina Roriz Dantas, as well as monitoring student performance through summative assessments made available every two months and through formative assessments made available by educational platforms, discussed in the previous topic.

As a methodology for the data mining process, the CRISP-EDM will be adopted, which is a version adapted for the educational context of the consolidated standard of data mining and knowledge discovery aimed at the CRISP-DM industry. CRISP-EDM fully follows the six steps of the original model, but with educational data mining particularities (RAMOS et al., 2020).

References

[1] E. M. Onyema, N. C. Eucheria, F. A. Obafemi, S. Sen, F. G. Atonye, A. Sharma, A. O. Alsayed, Impact of coronavirus pandemic on education, Journal of Education and Practice 11 (2020) 108–121.

[2] M. R. Peres, Novos desafios da gestão escolar e de sala de aula em tempos de pandemia, Revista de Administração Educacional 11 (2020) 20–31.

[3] D. X. F. Giordano, A pandemia e as consequências no setor educacional: Desafios para os gestores escolares, Colóquios-Geplage-PPGED-CNPq (2021) 28–35.

[4] M. de Queiroz, F. G. A. de Sousa, G. Q. de Paula, Educação e pandemia: impactos na aprendizagem de alunos em alfabetização, Ensino em Perspectivas 2 (2021) 1–9.

[5] L. G. Ferreira, L. G. Ferreira, G. C. Zen, Alfabetiza-

ção em tempos de pandemia: perspectivas para o ensino da língua materna, fólio-Revista de Letras 12 (2020).
[6] K. L. d. Oliveira, E. Boruchovitch, A. A. A. d. Santos, Reading and school performance in portuguese and mathematics in elementary school, Paidéia (Ribeirão Preto) 18 (2008) 531–540.
[7] M. Buselli, K. P. Estevão, M. F. Sambugari, A importância da alfabetização matemática no ciclo i do ensino fundamental, Revista Eletrônica de Ciências Humanas 3 (2020).
[8] K. Schildkamp, W. Kuiper, Data-informed curriculum reform: Which data, what purposes, and promoting and hindering factors, Teaching and teacher education 26 (2010) 482–496.
[9] S. Gaftandzhieva, M. Docheva, R. Doneva, A comprehensive approach to learning analytics in bulgarian school education, Education and Information Technologies 26 (2021) 145–163.
[10] G. Siemens, D. Gasevic, Guest editorial-learning and knowledge analytics, Journal of Educational Technology & Society 15 (2012) 1–2.
[11] E. B. de Sousa, B. Alexandre, R. F. Mello, T. P. Falcão, B. Vesin, D. Gašević, Applications of learning analytics in high schools: a systematic literature review, Frontiers in Artificial Intelligence 4 (2021) 737891.
[12] P. Leitner, M. Khalil, M. Ebner, Learning analytics in higher education—a literature review, Learning analytics: Fundaments, applications, and trends (2017) 1–23.
[13] H. Waheed, S.-U. Hassan, N. R. Aljohani, M. Wasif, A bibliometric perspective of learning analytics research landscape, Behaviour & Information Technology 37 (2018) 941–957.
[14] A. Charitopoulos, M. Rangoussi, D. Koulouriotis, 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 (2020) 371–430.
[15] C. Cechinel, X. Ochoa, H. Lemos dos Santos, J. B. Carvalho Nunes, V. Rodés, E. Marques Queiroga, Mapping learning analytics initiatives in latin america, British Journal of Educational Technology 51 (2020) 892–914.
[16] A. Schleicher, The impact of covid-19 on education: Insights from education at a glance 2020, Retrieved from oecd. org website: https://www. oecd. org/education/the-impact-of-covid-19-on-education-insights-education-at-a-glance-2020. pdf (2020).
[17] J. R. R. Lima, A implementação do ensino híbrido no período pós-pandemia, Revista Ibero-Americana de Humanidades, Ciências e Educação 7 (2021) 10–10.
[18] H. Staker, M. B. Horn, Classifying k-12 blended learning., Innosight institute (2012).
[19] E. P. Schiehl, I. Gasparini, Modelos de ensino híbrido: Um mapeamento sistemático da literatura, in: Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE), volume 28, 2017, p. 1.
[20] L. Bacich, A. T. Neto, F. de Mello Trevisani, Ensino híbrido: personalização e tecnologia na educação, Penso Editora, 2015.
[21] G. Siemens, R. S. d. Baker, Learning analytics and educational data mining: towards communication and collaboration, in: Proceedings of the 2nd international conference on learning analytics and knowledge, 2012, pp. 252–254.
[22] K. C. Li, H. K. Lam, S. S. Lam, A review of learning analytics in educational research, in: International Conference on Technology in Education, Springer, 2015, pp. 173–184.
[23] M.-R. Sancho, A. Cañabate, F. Sabate, Contextualizing learning analytics for secondary schools at micro level, in: 2015 international conference on interactive collaborative and blended learning (icbl), IEEE, 2015, pp. 70–75.
[24] M. B. Horn, H. Staker, The rise of k-12 blended learning, Innosight institute 5 (2011) 1–17.
[25] B. Oliveras, N. Sanmartí, La lectura como medio para desarrollar el pensamiento crítico, Educación química 20 (2009) 233–245.
[26] M. Taras, Assessment–summative and formative–some theoretical reflections, British journal of educational studies 53 (2005) 466–478.
[27] B. S. Bloom, et al., Handbook on formative and summative evaluation of student learning. (1971).
[28] L. Cortesão, Formas de ensinar, formas de avaliar: breve análise de práticas correntes de avaliação, Reorganização curricular do ensino básico: avaliação das aprendizagens: das concepções às novas práticas (2002).
[29] V. Kovanovic, C. Mazziotti, J. Lodge, Learning analytics for primary and secondary schools, Journal of Learning Analytics 8 (2021) 1–5.
[30] B. M. Batubara, The problems of the world of education in the middle of the covid-19 pandemic, Budapest International Research and Critics Institute (BIRCI-Journal): Humanities and Social Sciences 4 (2021) 450–457.
[31] X. Luo, Supporting k-12 teachers' decision making through interactive visualizations: A case study to improve the usability of a real-time analytic dashboard, 2020.
[32] W. Admiraal, J. Vermeulen, J. Bulterman-Bos, Teaching with learning analytics: how to connect computer-based assessment data with classroom instruction?, Technology, Pedagogy and Education

29 (2020) 577–591.
[33] R. H. Sagarika, R. Kandakatla, A. Gulhane, Role of learning analytics to evaluate formative assessments: Using a data driven approach to inform changes in teaching practices, Journal of Engineering Education Transformations 34 (2021) 550–556.
[34] B. A. Schwendimann, M. J. Rodriguez-Triana, A. Vozniuk, L. P. Prieto, M. S. Boroujeni, A. Holzer, D. Gillet, P. Dillenbourg, Perceiving learning at a glance: A systematic literature review of learning dashboard research, IEEE Transactions on Learning Technologies 10 (2016) 30–41.
[35] W. Greller, H. Drachsler, Translating learning into numbers: A generic framework for learning analytics, Journal of Educational Technology & Society 15 (2012) 42–57.
[36] A. C. Gil, Como elaborar projetos de pesquisa. são paulo: Atlas, 2006. gil, antônio carlos, Como elaborar projetos de pesquisa 5 (2010).
[37] B. Básica, B. Complementar, Metodologia científica, São Paulo: Pearson Prentice Hall (2007).
[38] A. C. Gil, et al., Como elaborar projetos de pesquisa, volume 4, Atlas São Paulo, 2002.
[39] M. Thiollent, Metodologia da pesquisa-ação, in: Metodologia da pesquisa-ação, 1988, pp. 108–108.

Towards a computer-assisted Computational Thinking (CT) assessment system in higher education

Xiaoling Zhang, Marcus Specht[a]

[a] Delft University of Technology, Van Moori Broekmanweg 6, Delft, 2628XE, The Netherlands

Abstract

With the vision to promote CT to a wider group of audiences, this PhD project explores the formative assessment of CT skills in Programming Education to support students to learn CT skills in Higher Education. In this project, we plan to investigate the importance of CT in the context of Higher Education, explore the relationship between CT skills and programming skills, build a model to assess learners' CT skills and develop a computer-assisted assessment system with automated components to enhance students' CT competences in Higher Education. Mixed-method research methodologies will be employed in distinct phases of the project accordingly. A system which allows formative assessment of CT skills will be iteratively designed and constructed throughout the project. The outcome of the project should support the CT learning process, make CT more visible for people from diverse backgrounds and empower them with a CT mindset to embrace the digitalization of society.

Keywords 1

Computational Thinking, Computer-Assisted Assessment, Higher Education, Educational Technology

1. Introduction

1.1. Digitalisation and Computational Thinking

Living in an era of digitalisation, digital elements is everywhere. For instance, education, healthcare and governance, fundamentals to a modern society, are developing towards a digital direction [1-3]. This has a huge influence on employment and skills, such as the increasing unemployment rate, and the increasing demand for digital skills in the labour market [4]. To empower people the capability of living and working in such a digitalized society, governments, and education institutions from distinct levels world-wide have been striving to promote education of computer-based technologies and skills varying from academy to industry. Among skills being mentioned, digital skills, problem-solving skills, and computational thinking (CT) are the top few most mentioned skills and are regarded as fundamental skills in workplaces [5-7, 28].

Computational Thinking is closely related to the development of digitalisation in different domains and changes the professional competencies need for these professions. First proposed by Papert as procedural thinking [8] and then being promoted by Wing [9], a considerable amount of research has been conducted to define CT in the past few decades. Though there is no agreed-upon theoretical or operational definition so far, existing works share main components of CT, which are problem decomposition, abstraction, pattern recognition and algorithm [9-15]. Besides studying the operational and theoretical definition of CT, massive amounts of studies have been conducted globally to investigate topics around CT education, such as

Proceedings of the Doctoral Consortium of Seventeenth European Conference on Technology Enhanced Learning, September 12-16, 2022, Toulouse, France
EMAIL: x.zhang-14@tudelft.nl (A. 1); m.m.specht@tudelft.nl (A. 2);
ORCID: 0000-0003-0951-0771 (A. 1); 0000-0002-6086-8480 (A. 2);

pedagogical contents, didactic strategies, integration of CT into other disciplines [16-26].

People of almost all ages can be participants in these studies, however, most of the existing research focuses specifically on K-12 settings, with an increasing number of studies conducted in Higher Education over the last decade. Existing work in K-12 settings has explored a considerable range of topics regarding learning and teaching CT in both science, technology, engineering, and mathematics (STEM) and non-STEM disciplines, results in a flourishment of development in tools and activities for teaching and learning CT, both CS-unplugged such as bebras challenge and Lego construction and CS oriented such as programmable robotics, micro-bits, code.org, Scratch, Alice [20]. While being regarded as crucial competence for learners in higher education, the development of CT, compared to CT in K-12 setting, is still in its infancy. Increased attention has been paid to CT in Higher Education in recent years, most of which are related to Computer Science (CS) major, and few are in non-CS major disciplines [26]. In their literature review, Lyon and Magana identified several issues existing in current CT education which makes it difficult for students to understand CT, including unclear definition, lack of assessment methods, unclear use of CT in classrooms [26]. They also stressed the necessity of a clearer definition of CT and called for more implementation of CT in Higher Education and studies.

With current insights into existing literature, it is obvious to conclude that CT is closely related to developments of digitalisation in different domains and changes the professional competencies needed for these professions. However, it is still unclear how to embed CT in different curricula and how to develop transdisciplinary CT skills. Therefore, researchers need to conduct studies to establish a comprehensive and more complete system for the purpose of enhancing people's CT competencies.

1.2. Computational Thinking and Programming Education in Higher Education

Learners of diverse backgrounds learn CT with various purposes and learners' target objectives considering the proficiency level also differ accordingly on learner's level of proficiency. Therefore, it is important to know what the necessary skills are to be developed in higher education, what proficiency level of CT is expected for people from distinct domains and in what way should CT be incorporated in different domains in Higher Education. Programming education is frequently used for fostering CT in higher education; visual programming in Scratch and Alice as well as text programming in Python, C, C++, Java have been used for teaching CT in K-12 settings as well as in Higher Education settings [39-40]. However, it remains a controversial topic whether everyone should learn to code. For example, Shein acclaimed that "Not everyone needs coding skills but learning how to think like a programmer can be useful in many disciplines" [35]. Therefore, it would be important to study the role of Programming Education.

CT and programming skills are closely interlinked and are both challenging for novice learners [29, 30]. However, a significant drop-out rate can be found in programming education on novice learners due to distinct difficulties students meet during their learning process [31]. Pane et al. [32] found that the ability to solve problems using programming skills so that the solution can be transformed and executed by computing agents does not come naturally for learners in CS studies. Additionally, studies also suggest that the absence of strategic tools can lead to deficient performance in learning to program [33-34].

To overcome these challenges, it is necessary to conduct research in both programming skills and CT skills and the relationship between them, which has been seldom researched.

Through qualitative and quantitative analyses, Selby [38] built a preliminary model to reveal connections between CT skills and programming activities using Bloom's taxonomy. However, it does not demonstrate in detail how CT can be measured in programming. Thus, it is necessary to carry out studies on how to empower students to use CT as a strategic tool for programming and gain CT knowledge through learning to program.

In brief, the following questions should be studied regarding CT and Programming Education in Higher Education:

- What skills are necessary for students in different domains in Higher Education?
- What is the role of Programming Education for students from different domains in Higher Education?
- How are programming skills and CT skills related and how to foster CT skills via programming?

1.3. Formative Assessment and Feedback Generation

Novice programmers who are new to programming are faced with challenges such as misunderstanding the programming concepts, misusing the language syntax, and understanding poorly the feedback generated from the interpreter or compiler [31]. Alternative approaches to overcome these issues can be, for instance, enhancing teachers' pedagogical content knowledge, developing more effective didactic strategies, using formative assessment to provide feedback.
Assessment and feedback are essential elements in different learning theories which are used to assist students in the learning process [41]. Assessment is presented in two categories in general, formative assessment and summative assessment. Formative assessment is defined as assessment for learning while summative assessment as assessment of learning [42]. Formative assessment generally consists of teacher observation, conventional assessment, oral presentation and so on. According to Paul Black & Dylan Wiliam [43], formative assessment remains incomplete until it has resulted in feedback and action on the part of the instructor and/or learner. Therefore, a formative assessment is all about feedback. According to Hattie and Timperley [45], feedback is one of the most crucial factors for efficient learning.

The development of formative assessment in Programming Education is still at an early age though there has been lots of research on intelligent tutoring systems which assess students' solutions in recent years. Computer-assisted learning environments provide the opportunity to automate the assessment and considerable work has been conducted to assess works in STEM disciplines automatically [44]. In terms of Programming Education, Grover [42], in the Raspberry Pi Foundation Computing Education Research Seminar, strived to promote the concept of formative assessment in CS for K-12. In contrast, no existing study explicitly facilitates formative assessment either in computing education or in Programming Education specifically in Higher Education.

While most of the assessments being conducted on CT and Programming Education are summative, there is some work that applies formative assessment measures in their implementations. These implementations focused on merely part of programming education and none of these works incorporated CT into programming education, making them infeasible for assessing CT in Programming Education. Meanwhile, some studies aimed at supporting students in learning to program, mostly in the form of automated assessment systems and intelligent tutoring systems for programming exercises. In their literature review, Keuning et al. [47] reported that most of the elaborate feedback provided by the systems reviewed focus on the identification of mistakes and no further suggestions on how to proceed and fix the problem. This, however, can impede students from enhancing their performance according to the feedback model defined by Hattie and Timperley [45]. Therefore, it is necessary to conduct research to explore formative assessment of CT in Programming Education in order to assist students in the learning process to enhance their CT in Programming Education.

With the vision to make CT skills more accessible and tangible in the context of Programming Education for learners from different domains, this project aims to develop formative assessment components to improve students' performance in learning to program and gaining CT skills.

2. Theoretical Background

To address the questions mentioned in the last section, theories on formative assessment and theoretical models of CT and Programming Education are crucial. Therefore, they are being investigated to ensure the reliability of the conduction of the project. CT and Programming Education will be first introduced with a focus on Brennan and Resnick's operational framework [16] and Bloom's taxonomy on Programming Education. Then follows theory for formative assessment and feedback models with a focus on Hattie's feedback model and the theory of formative assessment from Paul Black & Dylan Wiliam [43]. The theories are identified as the backbone in the implementation of this project.

2.1. Computational thinking and programming education (Bloom's Taxonomy)

Although there are no agreed-upon operational and theoretical definitions, definitions given by researchers and educators share the same elements in their definition. Wing defined CT operationally with the concepts of abstraction and automation [9]. Having components used in Wing's definition, Barr and Stephenson [46] included also problem decomposition, algorithmic thinking, data collection, analysis and representation and simulation to define CT. Similarly, Selby's definition of CT consists of abstraction, decomposition, generalization, evaluation and algorithmic design [38]. Four main components of CT can be identified from existing definitions: problem decomposition, pattern recognition, abstraction and algorithmic design.

Deriving from the main CT components, Brennan and Resnick [38] proposed an operational framework of CT which is frequently used in CT studies and the framework relates quite close to programming concepts and skills. Three dimensions constitute the framework: computational concepts, computational practices and computational perspectives. These components are recognizable in other disciplines and practices as well, which is consistent with Denning's description CT: it is nothing new, it is the way of thinking about the world shaped by the current technologies [50]. This framework considers elements comprehensively from both a knowledge perspective and a psychology perspective and it is a framework that can be practically used for setting learning objectives, designing pedagogical contents, and assessing students' performance [48].

CT concepts and CT practices involved in this framework [48] are some of the indicators that measure CT competences through programming concepts and practices. Studies have been conducted to map programming skills and CT skills as well as using Bloom's taxonomy and SOLO taxonomy to differentiate various levels of cognition for both CT and programming skills [36, 37]. Assessment of CT through assessing Scratch codes in Dr. Scratch with the framework presented by Brennan [38] is an example of how CT can be matched in Programming Education [49]. Selby [39] developed a model which discovers the relationship between CT skills and programming activities by using Bloom's taxonomy. This model can serve as the backbone in fostering CT via programming and vice versa.

2.2. Formative assessment and feedback generation

Having a CT framework and a model which maps CT to programming using cognitive levels in Bloom's taxonomy is insufficient for this project as the aim of this project is to enhance students' CT skills via formative assessment. Therefore, this subsection will introduce theories on formative assessment and models for generating feedback as formative assessment is said to be all about feedback [42].

Assessment is identified as one of the fundamental elements in all learning theories in education [41]. Formative assessment is defined as assessment for learning, and it is expected to result in feedback and action on the part of the instructor and/or learner if formative assessment is implemented. Thus, feedback is crucial in formative assessment, which is consistent with "Feedback plays a crucial role in learning" [27].

The efficiency of the feedback is influenced by the kind of formative feedback provided and the learner characteristics. Under the definition given by Boud and Molloy [51], feedback is formative, and it can be used to improve learners' performance. Another type of feedback is summative feedback, typically consists of grades or percentage of evaluation, which informs the learner about the performance. However, this type of feedback is usually too superficial to be useful for learners. Therefore, formative feedback is of more importance for the purpose of improving learning.

Different definitions and models have been investigated regarding feedback generation both in general and for studies in specific domains. Boud and Molloy define feedback as a process in which the learners improve their work with the given information which presents the discrepancy and similarities between learners' work and the expected standards [51]. Hattie and Timperley [45] described a model for feedback which is also in a formative way. The model aims to answer learners' questions about where they are, how they should proceed and where they should arrive. In this model, feedback is categorized into "task level", "process level", "self-regulation level" and "self-level", with findings indicating self-level the most ineffective one.

Having a model of feedback is insufficient for generating the most effective feedback for learners, extra facets should be considered when generating feedback. In Le and Pinkwart's work [52], programming exercises supported in learning environments were categorized into three classes according to the level of ill-definedness of the programming problem. As Hattie and Timperley [45] pointed out that feedback should target students at appropriate levels, it would be necessary to also consider Narciss's [53] categorization of feedback in computer-assisted learning environments according to the aspects of the instructional context. Narciss [53] has identified eight types of feedback components, five of them are elaborated feedback component and are intended to "improve learner's performance": knowledge about task constraints (KTC), knowledge about concepts (KC), knowledge about mistakes (KM), knowledge about how to proceed (KH) and knowledge about Meta-cognition (KMC). Combining the context to be assessed, the type of exercises to be assessed and the feedback level to provide, a strategy for generating feedback can be devised.

In sum, this project will first focus on identification of the need for CT and the role of Programming Education in different disciplines. Then, the focus will be shifted to the measurement of CT skills and programming skills and the relationship between these two sets of skills. Based on studies conducted, this project will then explore feedback generation and develop feedback generation strategies to promote CT for students from different domains and enhance their performance in CT skills and programming skills. The following definitions will be used for the remainder of the proposal:

- CT competencies: according to Brennan's framework, CT competencies refer to CT concepts, CT practices and CT perspectives.
- Programming skills: including conceptual knowledge, syntactic knowledge and strategic knowledge and programming style.
- Indicators for CT skills and programming skills: Any features, instruments that provide a sign or a signal of CT competence and programming skills.
- Formative assessment: A kind of assessment which provides feedback to the learner and it is an assessment for learning.

3. Research Questions

The research will be guided by the following research questions:

RQ1. How are CT skills and programming skills being conceptualised and measured?

1. What are indicators and assessment methods for CT competence and programming skills?
2. What systems and domains are using the indicators and assessments for CT competence and programming skills?
3. How to evaluate the validity of the indicators/assessment?

After collecting the indicators for CT competencies and assessment methods, techniques used for formative assessment and feedback generation and the effect of feedback should be investigated to provide the basis for design feedback generation strategies. Therefore, the second research question is:

RQ2. How should feedback be provided to support developing CT skills and programming skills, and how should formative assessment be implemented in this process?

1. What formative assessment and feedback generation strategies are used for the development of programming skills and CT competence?
2. What are the effects of different types of feedback on motivation, learning gain, and CT performance?
3. What empirical knowledge has been established regarding the effect of providing feedback on the development of CT competence and programming skills??

4. How to use formative assessment and generate feedback to support the development of CT and programming skills?

Based on the results obtained by answering the questions above, the next step is to contextualize the feedback and thus employ formative assessments for learners from different educational backgrounds. To achieve the goal, the following questions should be studied:

RQ3. How can Programming Education and learning of CT be contextualised and embedded in different educational domains?

1. How important are links between curricular tasks and CT skills?
2. What role can transfer learning play in the contextualisation of CT?
3. What are the means to contextualise and embed CT learning in different domains?
4. What is the impact of contextualised teaching of CT skills on student motivation and understanding?

4. Design and Methods

The research is organized in four phases. In the first phase a desktop research/systematic literature review will be used to identify relevant works to get an overview of state-of-the-art regarding the topic being studied in this project - formative assessment for supporting students from different disciplines in the process of learning CT in the context of Programming Education in Higher Education. The following factors will be identified in this phase: indicators used for assessment and assessment methods for CT in Programming Education; formative assessment and feedback generation; empirical experiences of CT in different domains. The indicators identified in the first phase can then be used to develop an assessment model for CT in the context of Programming Education and a CT dashboard to present learners' progress and CT level. Exploratory research in the form of formative studies will be employed in this phase. Phase three will focus on the development of strategies for feedback generation and formative assessment based on the assessment model and the CT dashboard built in phase two. In the last phase, an integrated study will be conducted to evaluate the tool developed and refine the system according to different needs from people of different backgrounds. In parallel, design and development of the formative assessment tool for CT in the context of Programming Education will be carried out throughout the lifecycle of the project. In addition to that, the design, development and testing of the prototype will be iteratively proceeded. The plan for the workflow is provided in the diagram shown in Figure 1 (in the Appendix.

Phase 1 Desktop research - Literature review

In this phase, a systematic literature review will be conducted to get a holistic overview of formative assessments for supporting learners in different disciplines to learn CT in the context of Programming Education. This process will follow the PRISMA statements and the PRISMA diagram, including defining research questions, collecting literature, screening, checking eligibility of the literature, data extraction and analysis of extracted results. RQ1.1, RQ1.2, RQ2.1 and RQ3.1 will be addressed in this phase. The outcome of this phase will be indicators used for assessment and assessment methods for CT in Programming Education; a comprehensive overview of formative assessment and feedback generation; empirical experiences of CT in different domains.

Phase 2 Exploratory research/ Formative studies - Build up the assessment model and a CT Dashboard

This phase begins with interviews with different target groups. The aim of the interview is to identify the necessity of CT skills and the role of Programming Education for learners with diverse backgrounds. In combination with the indicators and assessment methods identified in Phase 1, assessment models can then be prototyped according to the result from a qualitative analysis of the interviews. The interviews should also clarify the embedding of the CT skills in the different study contexts and the relevance for student and educators' goals in the different curricula. According to the goals and models a CT dashboard will be developed. To ensure the usability of the models and the CT dashboard, a usability study will be conducted in a programming course for students and the models and CT dashboard will be refined accordingly. Once the usability of the model is verified, quasi experimental studies will then come into play to examine the effect of using the assessment model and CT dashboard.

In this phase, RQ1.3, RQ2.2 and RQ2.3 will be studied, and an assessment model based on the indicators and assessment methods found in Phase 2 will be developed. This will include a participatory design and prototype of a CT dashboard. The design and the development of the models and the CT dashboard will proceed iteratively.

Phase 3 Develop feedback and formative assessment based on assessment model and CT Dashboard

This phase will focus on addressing RQ2.4, which is about developing proper feedback generation strategy to present to students their CT competencies and programming skills based on the strategies for feedback generation and formative assessment identified in Phase 1 and the CT assessment prototype and CT dashboard developed in Phase 2. Formative studies will be conducted to iteratively develop the feedback generation model. Student models will be identified in this phase by using data such as analysis of students' code, student's competence profile and analysis of students' performance. At the end of this phase, strategies for providing feedback and formative assessment should be identified.

Phase 4 Evaluation - Integrated study on the developed formative assessment tool

The result from Phase 3 will provide a basis to address RQ3.2 to RQ3.4 in this phase. Considering the factors which are important in adapting feedback for learners from different domains identified in phase 1, RQ3.2 to RQ 3.4 will be addressed by conducting an integrated study which includes both case studies and an evaluation study to contextualise the model developed and embed it into different educational domains and verify the validity and the effectiveness of the designed system. This integrated study aims to evaluate the tool developed and refine the system according to diverse needs from people of different backgrounds such that CT can be promoted further to a wider audience.

5. Acknowledgements

Xiaoling Zhang: Conceptualization, Methodology, Data Collection, Analysis, Writing - Original Draft, Writing – Review & Edit, Visualization, Resources

Marcus Specht: Conceptualization, Methodology, Writing – Review & Edit

This work is a part of a PhD project funded by Center for Education and Learning at Leiden-Erasmus-Delft Universities (LDE-CEL).

6. References

[1] Dillenbourg, P. (2016). The Evolution of Research on Digital Education. Int J Artif Intell Educ 26, 544–560 (2016). https://doi.org/10.1007/s40593-016-0106-z.

[2] Duggal, R., Brindle, I., Bagenal, J. (2018, January 15). Digital healthcare: regulating the revolution. doi: https://doi.org/10.1136/bmj.k6.

[3] Holzer, M., Kim, Seang-Tae. (2006) Digital Governance in Municipalities Worldwide (2005) : A Longitudinal Assessment of Municipal Websites Throughout the World. United Nations Public Administration Network. http://unpan1.un.org/intradoc/groups/public/documents/aspa/unpan022839.pdf.

[4] Schwab,K., Sala-i-Martín, X. (2013).The Global Competitiveness Report 2013–2014: Full Data Edition. URI: http://hdl.handle.net/11146/223.

[5] D Barr, J Harrison, L Conery. (2011). Computational thinking: A digital age skill for everyone. Learning & Leading with Technology, 2011 - ERIC.

[6] Francisco José García-Peñalvo, Antònio José Mendes,Exploring the computational thinking effects in pre-university education,Computers in Human Behavior,Volume 80,2018, Pages 407-411, ISSN 0747-5632, https://doi.org/10.1016/j.chb.2017.12.005.

[7] Anita JUŠKEVIČIENĖ, Valentina DAGIENĖ. (2018). Computational Thinking Relationship with Digital Competence.

[8] Papert, S. Mindstorms: Children, Computers, and Powerful Ideas. Basic Books, 1980.

[9] Wing, J. Computational thinking, Commun. ACM 49, 3 (Mar. 2006), 33–35.

[10] Wing, J. 2008. Computational thinking and thinking about computing. Philosophical Transactions of The Royal Society A, 366, 3717-3725.

[11] Wing, J. 2011. Research Notebook: Computational Thinking - What and Why?

The Link. Pittsburgh, PA: Carneige Mellon.
[12] Computer Science Teachers Association Task Force. 2011. K–12 Computer Science Standards, New York, ACM.
[13] Hu, C. 2011. Computational thinking: what it might mean and what we might do about it. Proceedings of the 16th annual joint conference on Innovation and technology in computer science education. Darmstadt, Germany: ACM.
[14] Guzdial, M. 2011. A Definition of Computational Thinking from Jeannette Wing. Computing Education Blog [Online]. Available from: http://computinged.wordpress.com/2011/03/22/a-definition-of-computational-thinking-from-jeanette-wing/ [Accessed 30-11-2020].
[15] Guzdial, M. 2012. A nice definition of computational thinking, including risks and cyber-security. Computing Education Blog [Online]. Available from: http://computinged.wordpress.com/2012/04/06/a-nice-definition-of-computational-thinking-including-risks-and- cyber-security/ [Accessed 30-11-2020].
[16] Brennan, K., & Resnick, M. 2012, New frameworks for studying and assessing the development of computational thinking. Paper presented at the Annual Meeting of the American Educational Research Association, Vancouver, BC.
[17] Z. Berkaliev et al., Initiating a programmatic assessment report, Primus 24 (2014), 403–420. https://doi.org/10.1080/10511970.2014.893939.
[18] P. Curzon et al., Developing computational thinking in the classroom: a framework, Comput. Sch. (2014), http://eprints.soton.ac.uk/369594/10/DevelopingComputationalThinkingInTheClassroomaFramework.pdf.
[19] C. Evia, M. R. Sharp, and M. A. Perez-Quinones, Teaching structured authoring and DITA through rhetorical and computational thinking, IEEE Trans. Prof. Commun. 58 (2015), 328–343. https://doi.org/10.1109/TPC.2016.2516639.
[20] S. Grover and R. Pea, Computational thinking in K-12: A review of the state of the field, Educ. Res. 42 (2013), 38–43. https://doi.org/10.3102/0013189X12463051.
[21] K. Jaipal-jamani and C. Angeli, Effect of robotics on elementary preservice teachers' self-efficacy, Sci. Learn. Comput. Think. (2017), 175–192. https://doi.org/10.1007/s10956-016-9663-z.
[22] Y. Jeon and T. Kim.The effects of the computational thinking-based programming class on the computer learning attitude of non-major students in the teacher training college, J. Theor.Appl. Inf. Technol. 95 (2017), 4330–4339.
[23] B. Kim, T. Kim, and J. Kim, Paper-and-pencil programming strategy toward computational thinking for non-majors: Design your solution, J. Educ. Comput. Res. 49 (2013), 437–459. https://doi.org/10.2190/EC.49.4.b
[24] Y. Lan, Exploration on database teaching based on computational thinking, Bol. Tec. Bull. 55 (2017), 363–370. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85038935540&partnerID=40&md5=010e1ae2c24dbf9eaa18b000844ece22.
[25] C. Mouza et al., Resetting educational technology coursework for pre-service teachers: A computational thinking approach to the development of technological pedagogical content knowledge (TPACK), Australas. J. Educ. Technol. 33 (2017). https://doi.org/10.14742/ajet.3521.
[26] Lyon, J. & Magana, A. (2020). Computational thinking in higher education: A review of literature. Computer Applications in Engineering Education. 28. 10.1002/cae.22295.
[27] A. L. S. O. de Araujo, W. L. Andrade, and D. D. S. Guerrero, “A systematic mapping study on assessing computational thinking abilities,” in Frontiers in Education Conference (FIE), 2016 IEEE. IEEE, 2016, pp. 1–9.
[28] Francisco José García-Peñalvo, Antònio José Mendes. Exploring computational thinking effects in pre-university education. Computers in Human Behavior. Volume 80. 2018. Pages 407-411. ISSN 0747-5632. https://doi.org/10.1016/j.chb.2017.12.005.
[29] M. Tedre. Many paths to computational thinking. Paper presented at the TACCLE

3 final conference, Brussels, Belgium (2017).

[30] P. Denning, M Tedre, P Yongpradit. Misconceptions about computer science. Communications of the ACM. Volume 60, Number 3 (2017), Pages 31-33.

[31] Yizhou Qian and James Lehman. 2017. Students' Misconceptions and Other Difficulties in Introductory Programming: A Literature Review. ACM Trans. Comput. Educ. 18, 1, Article 1 (December 2017), 24 pages. DOI :https://doi.org/10.1145/3077618.

[32] Pane, J. F., Ratanamahatana, C. A. & MYERS, B. A. 2001. Studying the language and structure in non-programmers' solutions to programming problems. International Journal of Human-Computer Studies, 54, 237-264.

[33] Robins, A., Rountree, J. & Rountree, N. 2003. Learning and Teaching Programming: A Review and Discussion. Computer Science Education, 13, 137 - 172.

[34] Saknini, V. & Hazzan, O. 2008. Reducing Abstraction in High School Computer Science Education: The Case of Definition, Implementation, and Use of Abstract Data Types. J. Educ. Resour. Comput., 8, 1-13.

[35] Esther Shein. 2014. Should everybody learn to code? Commun. ACM 57, 2 (February 2014), 16–18. DOI:https://doi.org/10.1145/2557447.

[36] Susana Masapanta-Carrión and J. Ángel Velázquez-Iturbide. 2018. A Systematic Review of the Use of Bloom's Taxonomy in Computer Science Education. In Proceedings of the 49th ACM Technical Symposium on Computer Science Education (SIGCSE '18). Association for Computing Machinery, New York, NY, USA, 441–446. DOI:https://doi.org/10.1145/3159450.3159491.

[37] David Ginat and Eti Menashe. 2015. SOLO Taxonomy for Assessing Novices' Algorithmic Design. In Proceedings of the 46th ACM Technical Symposium on Computer Science Education (SIGCSE '15). Association for Computing Machinery, New York, NY, USA, 452–457. DOI:https://doi.org/10.1145/2676723.2677311.

[38] Cynthia C. Selby. 2012. Promoting computational thinking with programming. In Proceedings of the 7th Workshop in Primary and Secondary Computing Education (WiPSCE '12). Association for Computing Machinery, New York, NY, USA, 74–77. DOI:https://doi.org/10.1145/2481449.2481466.

[39] Cynthia C. Selby. 2014. How Can Teaching of Programming Be Used to Enhance Computational Thinking Skills? University of Southampton, Faculty of Social and Human Sciences, PhD Thesis, pagination.

[40] Valerie J. Shute, Chen Sun, Jodi Asbell-Clarke. Demystifying computational thinking. Educational Research Review. Volume 22. 2017. Pages 142-158. ISSN 1747-938X. https://doi.org/10.1016/j.edurev.2017.09.003.

[41] Dale H. Schunk. (2012). Learning theories an educational perspective sixth edition.

[42] Shuchi Grover. (2020). Formative Assessment for Students in CS Classrooms, https://www.youtube.com/watch?v=0ZuSqsJQRFg&feature=emb_title. [last access: 2020-11-16]

[43] Black, P., Wiliam, D. Developing the theory of formative assessment. Educ Asse Eval Acc 21, 5 (2009). https://doi.org/10.1007/s11092-008-9068-5.

[44] Barana, A., Conte, A., Fioravera, M., Marchisio, M., Rabellino, S.; A model of formative automatic assessment and interactive feedback for STEM. In: Proceedings of 2018 IEEE 42nd Annual Computer Software and Applications Conference, pp. 1016–1025. IEEE Computer Society Conference Publishing Services (CPS), Tokyo, Japan (2018).

[45] Hattie, J., & Timperley, H. (2007). The power of feedback. Review of educational research, 77(1), 81-112.

[46] V. Barr and C. Stephenson, Bringing computational thinking to K-12: What is involved and what is the role of the computer science education community? ACM Inroads 2 (2011), 48–54.

[47] Keuning, H., Jeuring, J. T., & Heeren, B. J. (2019). A Systematic Literature Review of Automated Feedback Generation for

Programming Exercises. ACM Transactions on Computing Education, 19(1), [3]. https://doi.org/10.1145/3231711

[48] Yeni, S., & Hermans, F. (2019). Design of CoTAS: Automated computational thinking assessment system. In TACKLE 2019: 2nd Systems of Assessments for Computational Thinking Learning workshop: Proceedings of the 2nd Systems of Assessments for Computational Thinking Learning workshop (TACKLE 2019) (Vol. 2434). (CEUR Workshop Proceedings).

[49] Jesús Moreno-León and Gregorio Robles. 2015. Dr. Scratch: a Web Tool to Automatically Evaluate Scratch Projects. In Proceedings of the Workshop in Primary and Secondary Computing Education (WiPSCE '15). Association for Computing Machinery, New York, NY, USA, 132–133. DOI:https://doi.org/10.1145/2818314.2818338.

[50] Peter J. Denning and Matti Tedre. (2019). Computational Thinking.

[51] Boud, D., & Molloy, E. K. (2013). Feedback in higher and professional education: Understanding it and doing it well. Routledge. https://doi.org/10.4324/9780203074336.

[52] Sebastian Gross, Bassam Mokbel, Barbara Hammer, and Niels Pinkwart. 2015. Learning Feedback in Intelligent Tutoring Systems. Künstliche Intelligenz 29, 4 (2015), 413–418.

[53] Susanne Narciss. 2008. Feedback strategies for interactive learning tasks. Handbook ofresearch on educational communications and technology (2008), 125–144.

7. Appendix

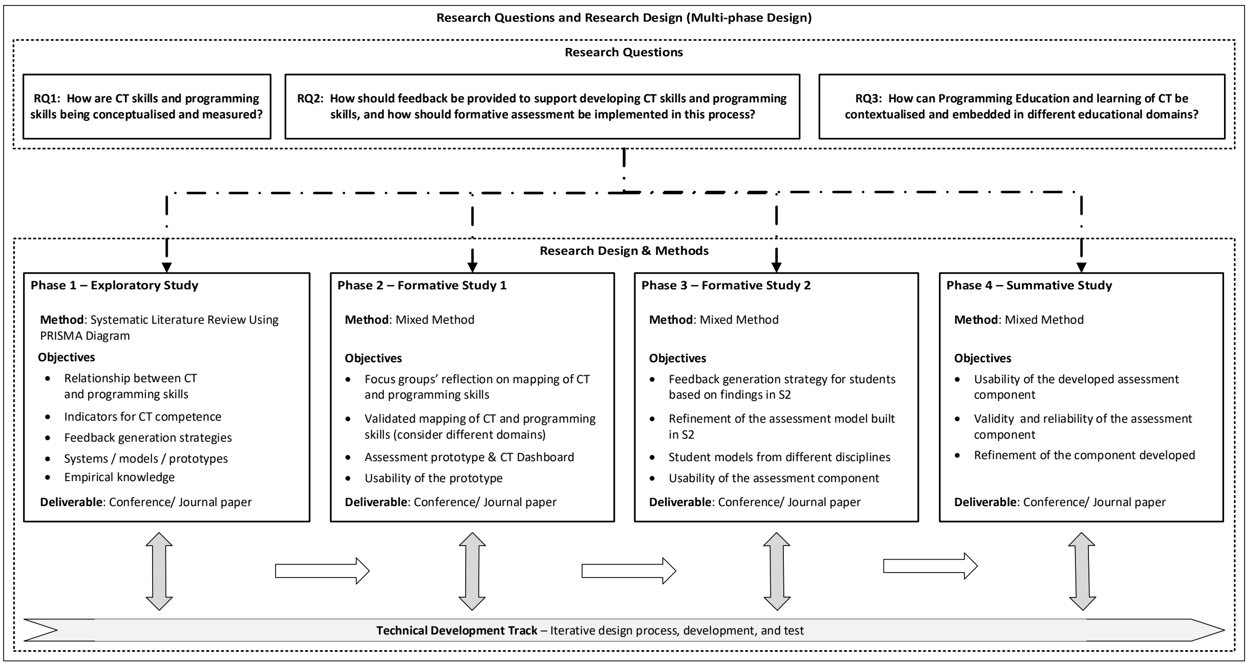

Figure 1. The whole PhD research plan with the main goals presented for each year. The system for providing feedback will be iteratively designed and developed throughout the project lifecycle.

Supporting student motivation through Social Comparison

Aditya Joshi

Utrecht University, 3584 CC Utrecht, The Netherlands

Abstract

Learning Analytics provides a methodology for the collection and analysis of learning data. Pedagogical research has always been inspired by ideas from applied psychology to discover and evaluate methods to boost motivation and engagement of students. Past research has shown that people often compare themselves with their peers in various contexts, including education. Social comparison has proven to be an effective motivation factor. Most of the recent research is based on using leaderboards to motivate individual comparison or open social student models to enable comparison with the course average. However, students' preferences towards social comparison can vary. For example, some people tend to compare upwards, while others mostly compare downwards, and some do not rely on social comparison at all.

This research aims to study how Social Comparison can be used to motivate students. In particular, it focuses on its effects on students' behavior, engagement, and performance on students from different demographics, with different psychological and motivation profiles. We will explore more adaptive approaches towards social comparison which adjust the direction and the magnitude of social comparison to suit students' needs and preferences.

Keywords

Learning Analytics, Motivation, Self-Regulated Learning, Social Comparison

1. Introduction

Social Comparison (SC) is the ability and tendency to gain self-evaluations by comparing oneself with their peers. It is an innate human trait and has been observed in kids as young as two years. We evaluate our opinions, skills, abilities, and achievements by comparing ourselves to others to define the self. Due to this, Social Comparison is a strong motivator and has always been leveraged in avenues ranging from commercial advertising to political discourse. Technology Enhanced Learning (TEL) environments provide feedback and supportive interfaces to help the students understand their progress towards the learning goals. Knowingly, or unknowingly, educational tools introduce social comparison as a tool through gamification elements such as leaderboards and halls-of-fame. In my research, I aim to design and evaluate mechanisms for adaptive Social Comparison.

1.1. Social Comparison

Festinger [1] proposed the theory of Social Comparison in 1954 which stands on the premise that humans have an innate desire to evaluate their abilities and opinions. A person's understanding about the situation and their abilities together have a bearing on their behavior. However, this requires accessing abilities even when objective information is not available, and then, they do so by comparing themselves with each other. It is found [2] that people may compare downwards to increase their subjective well-being which may enhance their self-esteem.

The concept of SC has been observed [3, 4] in children as young as preschoolers. Veroff [4] proposed that the concept of achievement begins in elementary school students, while the social comparison orientation increases as they grow older, the autonomous achievement orientation drops. As they grow up, they emphasize on demonstrating superior performance in comparison to others. In a usual classroom, the reward system provokes students to compare themselves socially. Similar effects were observed by Seidner et. al. [5] who noticed that the sense of pride of older students is affected more by comparing their performances with their peer rather than mastery.

The INCOM Scale [6] was developed for measuring individual differences in Social Comparison Orientation. It was found that two factors were responsible for explaining 38% and 10% of the variance. These two fectors reflect the perception of abilities and orientation based on Social comparison. They explained that such SC information may help ascertain the SC behavior of individuals and provide them interventions accordingly. It is also studied that the demographics like age, sex, race, or socio-economic status can be a factor of who students compare themselves with [7, 8, 9]. Studies [8] also show that students generally prefer compare themselves with friends and aquitances. Besides the target of comparison, the direction of comparison is also different [10]. Students may compare at the same level (laterally) [11], or upwards [12, 13, 7] or downwards [6, 14]. Thus it can be summarized that the target and preference of social com-

Proceedings of the Doctoral Consortium of the Seventeenth European Conference on Technology Enhanced Learning, September 12–16, 2022, Toulouse, France.
a.joshi@uu.nl (A. Joshi)
0000-0002-7115-4635 (A. Joshi)

parison may change depending on target, time, mood, motivation etc.

1.2. Leveraging Social Comparison to Improve Learning

Though SC is a psychological trait, SC can be further leveraged or manipulated by a researcher, a teacher, or a learning aid including learning support systems. SC can also serve as a feedback mechanism for self regulated learning [15]. Activity and progress visualizations [16, 17], and student model based tools [18, 19, 20] for conveying SC information as a feedback to the students have been created. However due to difference of how we perceive SC information, there is a need of the theories and implementations that provide adaptive SC based on temporal, demographic and situational differences. This paper proposes my ideas and plans to deeply understand the mechanisms to use Social Comparison in Technology Enhanced Learning (TEL) and study the need of adaptive SC that is capable of motivating and engaging a student based on their behavior and motivation profile.

Rest of this paper is organized as follows. Section 2 discusses how TEL systems leverage Social Comparisons, their different forms and the ideas which have been proposed in the past two decades. Section 3 explains the research plan by highlighting the research problem, research questions, explaining the concrete tasks that are planned for the next stage, and details of the TEL system that will be behind this research. Section 4 gives a brief outline of the experiments which are planned in the next few years. The Section 5 concludes this paper with discussion about the expected outcomes of this research, and the role of those outcomes in designing TEL systems that use SC effectively.

2. Social Comparison in TEL

SC has been an active area for research in the past decades. The idea of SC was first studied in depth by Festinger [1], who wrote that “There exists, in the human organism, a drive to evaluate his opinions and abilities.” He mentioned that people have a constant need to evaluate their abilities and test the validity of their opinions. Social Comparison [21] ”consists of comparing oneself with others in order to evaluate or to enhance some aspects of the self.” The effect of social comparison has been widely studied in education and pedagogical research.

This relates with the idea of Self Regulated Learning (SRL) [22], which is described as a cyclical process with three stages, namely, Forethought, Performance, and Self-reflection. Social comparison is active and affects students’ decisions and actions at all the three stages [23], and thus, TEL systems should provide support to the students during all the three stages.

Social comparison as a tool in Technology Enhanced Learning has been implemented in form of comparative charts or leaderboards [24, 25, 26] for a long time. They have been found effective in improving engagement and participation. However systems focussed on improving the SC feedback and studying the effects of SC are relatively recent topics.

One of the popular works in using comparative visualizations in education is Comtella [27] which was originally designed to motivate cooperative user behavior in peer to peer networks. It was proposed for exchanging resources and services in research/study groups by persuading the user to particupate in the sharing community through attractive and informative visualizations. This shows user’s contribution in form of a star whose color, brightness, shape depends on user’s interest, contribution and cooperation. It created a visualization that compares the whole class on multiple parameters in a single view and was validated [28, 29] to improve participation and contribution in the classroom. Their work highlighted the importance of developing the right visualization with respect to the goal.

Progressor [30, 31, 32] introduced Social Visualizations in an interface that helps students find relevant resources. It was observed that due to social comparison, class leaders provide a guidance to the rest of the students, and eventually lead to more engagement, and thus higher success rates.

Several ideas related to Open Student Models [33, 34] have been explored. Reading Circle [35] combines the idea of Open Student Models and Social Comparison to encourage students to read. A textbook reading support interface called Reading Mirror [36] shows SC information uses a grid-like interface that shows a student’s own progress and the class average with respect to sections/chapters of a textbook. It was found that most of the students felt that SC information altered their behavior positively. More recent works have used these interfaces [37] for encouraging motivation and engagement. An interface called Mastery Grids [38, 19] is a chart that shows students’ performance and compares with the class average. This form of visualization was shown to improve the motivation and engagement of the students.

A recent implementation [39] that gives younger students an understanding of their mastery of concepts in the achieving multiplication table fluency and can be used to give additional information, including SC cues. A study on the effect of a dashboard widget [40] for Massively Open Online Courses (MOOCs) that provides students more crisp infromation about their progress as well as SC cues improves the course completion rates. A dartboard like interface for multidimensional comparison

From the past works, it is evident that accessing SC information helps students achieve more motivation and engagement, and leads to a higher success rates. Meanwhile it has also been observed [41] that peer comparison doesn not necessarily improve, but in some cases, hamper the motivation of students. In some cases that though students prefer personalized recommendations, they may not find peer comparisons as useful or motivating. It was also found [42] that students' SC own preferences do not necessarily align with their best interests.

This leads to a challenge of analyzing the design as well as the effect of social comparison at a finer granularity. Social Comparison for better learning experiences needs further exploration. I plan to explore methods and create adaptive SC interfaces that can be effective tool to promote meaningful learning.

3. Research Plan

The main objective of this research is to devise effective mechanisms for using adaptive Social Comparison to improve students' motivation, engagement and learning outcome.

3.1. Research Problem

Typically, in most TEL and e-Learning softwares, SC information is usually provided to all the users the same way. However it has been found [43] that demographic and cultural backgrounds have a significant influence on self-construals based on social comparison which may affect the motivating factors. Apart from demographics, the SC orientation and direction also determines whether a person is motivated, challenged, or demotivated by SC information [6, 44]. That is, someone might get inspiration from someone who's performing better than them, while someone else may feel dissatisfied, or envy. [45] mentioned that though we all engage in social comparisons all the time, some people are more concerned and influenced by social comparison than others.

The differences in perception and effects of SC don't end with demographics and personality - but even at individual level, they expand over temporal and contextual dimensions. We engage in comparisons with others over time[46, 47] or our own past selves[48].

The popular Social Comparison approaches don't capture all these dimensions of social comparison and implement a one-size-fits-all solution regardless of individual and contextual differences. The issue with standard one-size-fits-all approaches is that though they work in some cases, they might affect some users rather negatively.[49]

This leads to the idea of a system that adapts the SC interface to the user based on their demographics, social comparison orientation, motivation profile and psychological profile.

This can be divided into following research questions:
RQ1. What are the current state of the art interactions to show social comparison and what are their effects on students' learning experience?
RQ2. How is the effect of social comparison on motivation and engagement related with personality traits?
RQ3. Are there distinct effects of using different types and direction of SC interfaces with different students? What are these effects?
RQ4. How to match a student with a social comparison method fine-tuned to promote their learning?

These will be studied and validated through a Learning Support System that will be used to supplement students' learning experience. Some of the initial experiments have been thoroughly discussed and planned to occur in the academic year 2022-23. The students will be asked to use Studylens as a Learning Support System (LSS) that will allow them to attempt ungraded assessment tests related to the concepts in the course, and I will analyze their performance and engagement with respect to the interface provided.

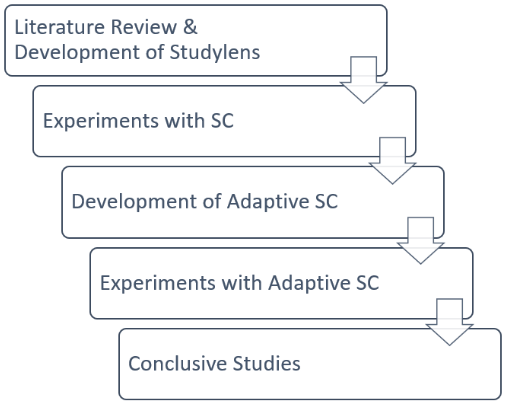

Figure 1: Overview of the Research Plan

3.2. Preparation and Setup

The first activities towards this goal are to study the existing research in Social Comparison in TEL and study the methods and tools used to convey the SC information to the students. This will be executed in parallel with development of a TEL system that can act as a Learning Support System.

3.2.1. Systematic Literature Review

A systematic literature review helps aggregate the existing research and ideas related to how SC has been used in TEL and what are the state of the art methods to use

SC as a tool to motivate students. This literature review is driven by the following research questions:

- What are the common ways of conveying SC information in TEL tools?
- What are the effects of context and direction of Social Comparison?
- What are common systems that allow students to actively engage with Social Comparison?
- What are the unexplored directions of utilizing SC in Education?

This is being performed with a hybrid methodology based on SPIDER [50] and PRISMA [51]. SPIDER helps summarize the study on the basis of (S) Sample size, (PI) Phenomenon of Interest, (D) Study Design, (E) Evaluation and (R) Research type. PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analysis) provides a structure for conducting the literature search and summarizing the analysis in a detailed manner. Meanwhile for each research that is included in the study, we also explore what was the SC method used and how was its effect studied.

3.2.2. Development of Studylens

Studylens is a Learning Support System built at the Utrecht University. It is a relatively lean implementation of the system which has been used with several university courses over the past few years [42]. The new implementation has been designed to have only the most necessary features that enable us to closely study the impact of social comparison.

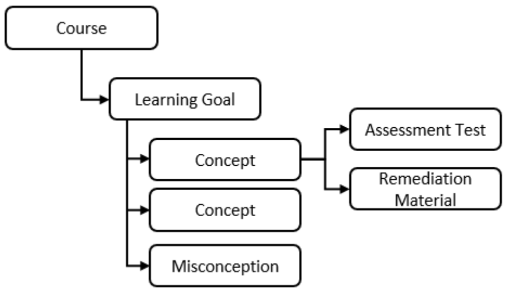

Figure 2: Organization of course content in Studylens

Studylens provides students a dashboard for exploring their expertise of the concepts covered in a course. It allows students to attempt self-assessment quizzes, which are associated with concepts and misconceptions that are part of a course inventory. When a student chooses to view the results of the test, they are shown their performance over each concept, and accordingly, remidiation material is recommended. Figure 2 shows the organization of a course into constituent Learning Goals each of which have one or more Concepts related to it.

When the next Learning Goal is activated, the student can take the self assessment test. The Knowledge Map is updated that helps student get feedback about their expertise of the topic. Figure 3 shows the current version which is expected to be further updated. In the social comparison setting, the student is shown the average performance of their peers as well. For research purposes, visibility of the social comparison widgets is configurable to provide a different view to each student based on their experiment group. The student can explore their knowledge and take the right remedial action through a list of learning resources.

At the time of writing, Studylens is planned to be used as a part of courses related to Evolutionary Biology at Utrecht University in The Netherlands. The courses are conducted over three-month terms and expected to be taken by 120-480 students. Studylens is recommended to the students as a self evaluation tool that can help them find their strong and weak points, and recommend remediation material to improve their understanding of the topics.

3.2.3. Technical Details

Studylens is built with Flask, a Python web framework at the backend. The database is MySQL, and the front end is based on a popular Javascript framework that provides a highly extensible component based design. The system is designed to provide user interfaces based on the experiment groups a user is allotted to.

A minimalistic Learning Record Store (LRS) is implemented in the database that stores users' activities in terms of actor (the student), verb (loading an activity, answering a question etc), and object (question or learning material). At later stage, this may be replaced by a full fledged LRS based on research requirements.

4. Planned Experiments

In the second year of this project (July 2022-June 2023), we have planned a project to explore different methods for personalized support with focus on exploring SC as a vehicle to motivate students with non-mandatory educational content. Studylens will be used as a learning support system for three courses at the department of Biology. These are all related to Evolutionary Biology at Year 1 and Year 2 of their undergraduate degree program.

Students will be able to take formative tests and a dashboard (Knowledge Map) will display their progress of mastery with respect to the Concepts covered in these courses. The interactive dashboard will help them explore the topics, learning goals, concepts better and de-

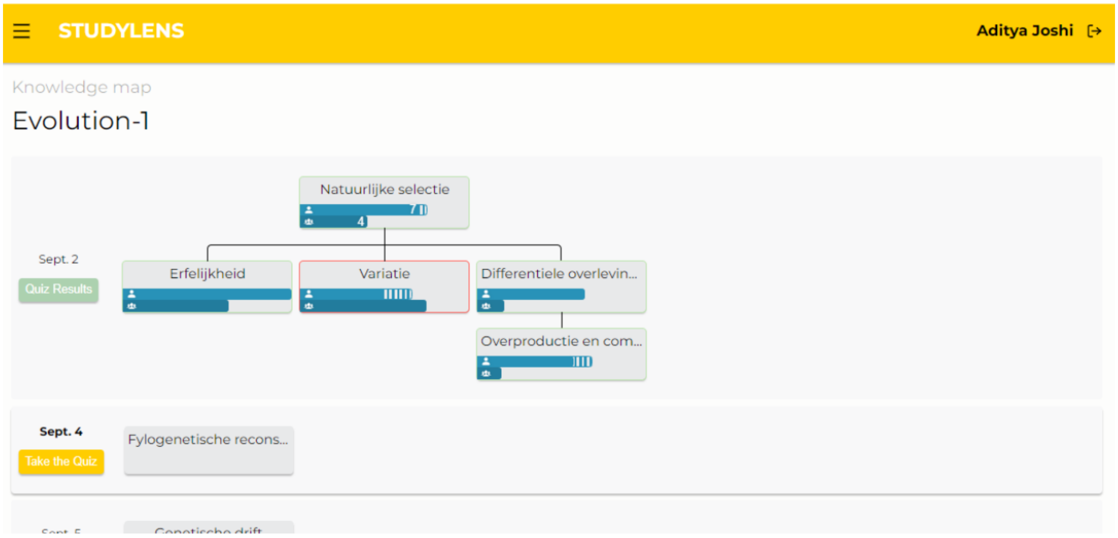

Figure 3: Screenshot of Studylens Knowledge Map

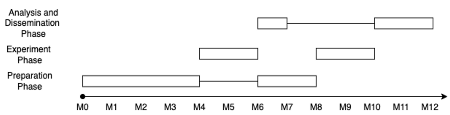

Figure 4: Experiments in Academic Year 2022-23

cide the learning activities that can help them fill the gaps in their learning.

At certain points in these courses, we will also use motivation inventories to understand different motivation profiles. We will also monitor app logs and User Experience logs to monitor user's engagement with the application. This can be used for comparative studies based on controlled experiments. We are specifically interested in analyzing the effect of the proposed SC interface in the TEL system on students' motivation, engagement and learning. We will examine user activity in the system, their grades outside the system, and possible changes in their motivation profile.

In these experiments, we will attempt to collect and analyze the data to be able to answer second and third research questions mentioned in the previous section. This will allow us to validate our hypothesis about the impact of social comparison information. The fourth research question may need further refinement based on the results of these experiments. The timeline of these experiments is shown in Figure 4. This offers us enough data collection, analysis and further improvements to the system.

4.1. Social Comparison and Motivation Inventories

A sub-task during preparations for the first experiment will be to study and develop questionnaires that can help us understand students' perceptions and inclinations based on different types and directions of social comparison and motivation profiles.

The need of this research may lead to creation of an inventory that can provide us insight into how SC affects a students' motivation. We are currently exploring using the items from Goal Achievement Framework[52] for studying performance and mastery orientation. Identification Contrast Scale[53] has been used to study effect of social comparison on cancer patients. We will test if this can be modified to use in studying SC in educational setup. Another highly popular inventory based on Self Determination Theory [54, 55] is a 22-item Motivation inventory that focuses on studying intrinsic motivation. 6-item Social comparison Concern Questionnaire [56] examines SC concern that can help us support the claim for adaptive SC in TEL.

4.2. Privacy and Ethical Concerns

We have thoroughly analyzed the privacy and ethical concerns related to any experiment of this kind. To mitigate the privacy risks, we have devised that the user's details in the system will be synthetic, and the teachers of the courses would map the user ids with actual students in the class. Meanwhile the teachers will not have access to the database or any internals of our system. This creates a safe barrier, thus allow anonymity during the data collection and analysis. Meanwhile use of the software, data collection, and participation in the motivation profiling surveys will be voluntary.

The aim of any TEL system is to devise the ideas that lead to discovery of more effective learning methodologies. This research can potentially impact how the Social Comparison information is visualized and used in learning softwares.

4.3. Plans for Future Work

The first set of experiments will be concluded by the middle of 2023 which would give adequate insights on the factors that determine the effect of SC on students' motivation and learning outcome. This will help in the development and refinement of effective interfaces for conveying SC information. This would be followed by providing adaptive SC interfaces to the students and comparing their effects with respect to student controlled and static social comparison.

5. Expected Contributions

This research will contribute to the empirical knowledge in Technology Enhanced Learning and Pedagogy domains. The outcomes of this research will allow us to gain a thorough and coherent understanding about how Social Comparison affect different behavior profiles, and create a system that adapts to a learner's behavior and provides them the Social Comparison cues that will motivate them.

The learning support system being built as a part of this research, Studylens, will be used to help students towards Self Regulated Learning. Though the experiments that are planned in the next year are related to Biology students, the tool and the ideas are domain independent and can be easily applied to other subjects and areas like computer science and soft skill training. We believe this research would lead to adaptation of SC methods that help the students achieve their learning goals.

References

[1] L. Festinger, A theory of social comparison processes, Human Relations 7 (1954) 117–140. URL: https://doi.org/10.1177/001872675400700202. doi:10.1177/001872675400700202.

[2] T. A. Wills, Downward comparison principles in social psychology., Psychological Bulletin 90(1) (1981) 245–271. doi:https://doi.org/10.1037/0033-2909.90.2.245.

[3] judith A. Chafel, Social comparisons by young children: A structural analysis., Early Childhood Research Quarterly 1(2) (1986) 155–165. doi:https://doi.org/10.1016/0885-2006(86)90026-8.

[4] J. Veroff, Social comparison and the development of achievement motivation, Achievement-related motives in children (1969) 46–101.

[5] L. B. Seidner, D. J. Stipek, N. D. Feshbach, A developmental analysis of elementary school-aged children's concepts of pride and embarrassment., Child Development 59 (1988) 367–377. doi:10.2307/1130316.

[6] F. X. Gibbons, B. P. Buunk, Individual differences in social comparison: Development of a scale of social comparison orientation, Journal of Personality and Social Psychology 76(1) (1999) 129–142. doi:https://doi.org/10.1037/0022-3514.76.1.129.

[7] H. Blanton, B. P. Buunk, F. X. Gibbons, H. Kuyper, When better-than-others compare upward: Choice of comparison and comparative evaluation as independent predictors of academic performance., Journal of personality and social psychology 76 (1999) 420.

[8] C. J. Meisel, C. J. Blumberg, The social comparison choices of elementary and secondary school students: The influence of gender, race, and friendship, Contemporary Educational Psychology 15 (1990) 170–182.

[9] I. Régner, J.-M. Monteil, Low-and high-socioeconomic status students preference for in-group comparisons and their underpinning ability expectations, Revue Internationale De Psychologie Sociale-International Review of Social Psychology 20 (2007) 87–104.

[10] P. Dijkstra, H. Kuyper, G. Van der Werf, A. P. Buunk, Y. G. van der Zee, Social comparison in the classroom: A review, Review of educational research 78 (2008) 828–879.

[11] F. Dumas, P. Huguet, J.-M. Monteil, C. Rastoul, J. B. Nezlek, Social comparison in the classroom: Is there a tendency to compare upward in elementary school, Current Research in Social Psychology 10 (2005) 166–187.

[12] I. Régner, F. Loose, A. Smeding, F. Dumas, A social side of academic achievement goals: The influence of ethnic and gender identities., in: 16th General Meeting of the European Association of Social Psychology, 2011.

[13] B. P. Buunk, F. X. Gibbons, Toward an enlightenment in social comparison theory, in: Handbook of Social Comparison, Springer, 2000, pp. 487–499.

[14] J. Suls, L. Wheeler, A selective history of classic and neo-social comparison theory, in: Handbook of social comparison, Springer, 2000, pp. 3–19.

[15] D. Persico, M. Passarelli, F. Manganello, F. Pozzi, F. M. Dagnino, A. Ceregini, G. Caruso, Automatic feedback, self-regulated learning and social com-

parison: A case study, Qwerty-Open and Interdisciplinary Journal of Technology, Culture and Education 15 (2020) 27–44.
[16] B. Yousuf, O. Conlan, Visen: Motivating learner engagement through explorable visual narratives, in: G. Conole, T. Klobučar, C. Rensing, J. Konert, E. Lavoué (Eds.), Design for Teaching and Learning in a Networked World, Springer International Publishing, Cham, 2015, pp. 367–380.
[17] S. Govaerts, K. Verbert, J. Klerkx, E. Duval, Visualizing activities for self-reflection and awareness, in: X. Luo, M. Spaniol, L. Wang, Q. Li, W. Nejdl, W. Zhang (Eds.), Advances in Web-Based Learning – ICWL 2010, Springer Berlin Heidelberg, Berlin, Heidelberg, 2010, pp. 91–100.
[18] J. Fritz, Video demo of umbc's "check my activity" tool for students, Educause Quarterly 33 (2010). URL: https://www.learntechlib.org/p/109175.
[19] J. Guerra, R. Hosseini, S. Somyurek, P. Brusilovsky, An intelligent interface for learning content: Combining an open learner model and social comparison to support self-regulated learning and engagement, in: Proceedings of the 21st international conference on intelligent user interfaces, 2016, pp. 152–163.
[20] P. Brusilovsky, S. Somyürek, J. Guerra, R. Hosseini, V. Zadorozhny, P. J. Durlach, Open social student modeling for personalized learning, IEEE Transactions on Emerging Topics in Computing 4 (2015) 450–461.
[21] J. Suls, R. Martin, L. Wheeler, Social comparison: Why, with whom, and with what effect?, Current Directions in Psychological Science 11 (2002) 159–163. URL: https://doi.org/10.1111/1467-8721.00191. doi:10.1111/1467-8721.00191. arXiv:https://doi.org/10.1111/1467-8721.00191.
[22] B. J. Zimmerman, Becoming a self-regulated learner: Which are the key subprocesses?, Contemporary Educational Psychology 11 (1986) 307–313. URL: https://www.sciencedirect.com/science/article/pii/0361476X86900275. doi:https://doi.org/10.1016/0361-476X(86)90027-5.
[23] R. Pérez-Álvarez, J. Maldonado-Mahauad, M. Pérez-Sanagustín, Tools to support self-regulated learning in online environments: Literature review, in: V. Pammer-Schindler, M. Pérez-Sanagustín, H. Drachsler, R. Elferink, M. Scheffel (Eds.), Lifelong Technology-Enhanced Learning, Springer International Publishing, Cham, 2018, pp. 16–30.
[24] B. Huang, K. F. Hew, Do points, badges and leaderboard increase learning and activity: A quasi-experiment on the effects of gamification, in: Proceedings of the 23rd International Conference on Computers in Education, Society for Computer in Education Hangzhou, China, 2015, pp. 275–280.
[25] P. Fotaris, T. Mastoras, R. Leinfellner, Y. Rosunally, Climbing up the leaderboard: An empirical study of applying gamification techniques to a computer programming class., Electronic Journal of e-learning 14 (2016) 94–110.
[26] S. Bai, K. F. Hew, M. Sailer, C. Jia, From top to bottom: How positions on different types of leaderboard may affect fully online student learning performance, intrinsic motivation, and course engagement, Computers & Education 173 (2021) 104297.
[27] H. Bretzke, J. Vassileva, Motivating cooperation on peer to peer networks, in: P. Brusilovsky, A. Corbett, F. de Rosis (Eds.), User Modeling 2003, Springer Berlin Heidelberg, Berlin, Heidelberg, 2003, pp. 218–227.
[28] J. Vassileva, Harnessing p2p power in the classroom, in: J. C. Lester, R. M. Vicari, F. Paraguaçu (Eds.), Intelligent Tutoring Systems, Springer Berlin Heidelberg, Berlin, Heidelberg, 2004, pp. 305–314.
[29] J. Vassileva, L. Sun, Evolving a social visualization design aimed at increasing participation in a class-based online community, International Journal of Cooperative Information Systems 17 (2008) 443–466.
[30] I. H. Hsiao, P. Brusilovsky, Motivational social visualizations for personalized e-learning, in: 21st Century Learning for 21st Century Skills, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 153–165.
[31] I.-H. Hsiao, J. Guerra, D. Parra, F. Bakalov, B. König-Ries, P. Brusilovsky, Comparative social visualization for personalized e-learning, in: Proceedings of the International Working Conference on Advanced Visual Interfaces, 2012, pp. 303–307.
[32] I.-H. Hsiao, F. Bakalov, P. Brusilovsky, B. König-Ries, Progressor: Social navigation support through open social student modeling, New Review of Hypermedia and Multimedia 19 (2013) 112–131. doi:10.1080/13614568.2013.806960.
[33] S. Bull, See yourself write: A simple student model to make students think, in: A. Jameson, C. Paris, C. Tasso (Eds.), User Modeling, Springer Vienna, Vienna, 1997, pp. 315–326.
[34] D. Hartley, A. Mitrovic, Supporting learning by opening the student model, in: S. A. Cerri, G. Gouardères, F. Paraguaçu (Eds.), Intelligent Tutoring Systems, Springer Berlin Heidelberg, Berlin, Heidelberg, 2002, pp. 453–462.
[35] J. Guerra, D. Parra, P. Brusilovsky, Encouraging online student reading with social visualization, volume 1009, 2013.
[36] J. Barria-Pineda, P. Brusilovsky, D. He, Reading mirror: Social navigation and social comparison for electronic textbooks, in: iTextbooks@AIED, 2019.
[37] B. Rahdari, P. Brusilovsky, K. Thaker, J. Barria-

Pineda, Using knowledge graph for explainable recommendation of external content in electronic textbooks, 2020.

[38] T. Loboda, J. Guerra, R. Hosseini, P. Brusilovsky, Mastery grids: An open source social educational progress visualization, 2014. doi:10.1007/978-3-319-11200-8_18.

[39] A. Leonardou, M. Rigou, J. Garofalakis, Adding social comparison to open learner modeling, in: 2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA), 2019, pp. 1–7. doi:10.1109/IISA.2019.8900763.

[40] D. Davis, I. Jivet, R. F. Kizilcec, G. Chen, C. Hauff, G.-J. Houben, Follow the successful crowd: Raising mooc completion rates through social comparison at scale, in: Proceedings of the Seventh International Learning Analytics & Knowledge Conference, LAK '17, Association for Computing Machinery, New York, NY, USA, 2017, p. 454–463. URL: https://doi.org/10.1145/3027385.3027411. doi:10.1145/3027385.3027411.

[41] C. Herodotou, A. Boroowa, M. Hlosta, B. Rienties, What do distance learning students seek from student analytics? (2020).

[42] S. Sosnovsky, Q. Fang, B. d. Vries, S. Luehof, F. Wiegant, Towards adaptive social comparison for education, in: European Conference on Technology Enhanced Learning, Springer, 2020, pp. 421–426.

[43] S. Guimond, N. R. Branscombe, S. Brunot, A. P. Buunk, A. Chatard, M. Désert, D. M. Garcia, S. Haque, D. Martinot, V. Yzerbyt, Culture, gender, and the self: variations and impact of social comparison processes., Journal of personality and social psychology 92 (2007) 1118.

[44] S. Y. Park, Y. M. Baek, Two faces of social comparison on facebook: The interplay between social comparison orientation, emotions, and psychological well-being, Computers in Human Behavior 79 (2018) 83–93.

[45] A. P. Buunk, F. X. Gibbons, Social comparison orientation: A new perspective on those who do and those who don't compare with others. (2006).

[46] N. E. Frye, B. R. Karney, Being better or getting better? social and temporal comparisons as coping mechanisms in close relationships, Personality and Social Psychology Bulletin 28 (2002) 1287–1299.

[47] F. Wolff, F. Helm, F. Zimmermann, G. Nagy, J. Möller, On the effects of social, temporal, and dimensional comparisons on academic self-concept., Journal of Educational Psychology 110 (2018) 1005.

[48] S. Redersdorff, S. Guimond, Comparing oneself over time: The temporal dimension in social comparison. (2006).

[49] M. D. Hanus, J. Fox, Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance, Computers & education 80 (2015) 152–161.

[50] A. Cooke, D. Smith, A. Booth, Beyond pico: The spider tool for qualitative evidence synthesis, Qualitative Health Research 22 (2012) 1435–1443. doi:10.1177/1049732312452938, pMID: 22829486.

[51] D. Moher, L. Shamseer, M. Clarke, D. Ghersi, A. Liberati, M. Petticrew, P. Shekelle, L. Stewart, D. Altman, A. Booth, A. Chan, S. Chang, T. Clifford, K. Dickersin, P. Gøtzsche, J. Grimshaw, T. Groves, M. Helfand, J. Higgins, E. Whitlock, Preferred reporting items for systematic review and meta-analysis protocols (prisma-p) 2015 statement, Systematic reviews (2015).

[52] A. J. Elliot, H. A. McGregor, A 2x2 achievement goal framework, in: American Psychological Association, 2001, pp. 501–519.

[53] K. I. Van Der Zee, B. P. Buunk, R. Sanderman, G. Botke, F. Van Den Bergh, The big five and identification–contrast processes in social comparison in adjustment to cancer treatment, European Journal of Personality 13 (1999) 307–326.

[54] E. L. Deci, R. M. Ryan, Intrinsic Motivation and Self-Determination in Human Behavior, Springer, New York, NY, 1985.

[55] R. M. Ryan, E. L. Deci, Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being., American psychologist 55 (2000) 68.

[56] M. Micari, P. Pazos, Worrying about what others think: A social-comparison concern intervention in small learning groups, Active learning in higher education 15 (2014) 249–262.

Towards a multi-faceted framework for planning and evaluating innovation in Engineering Education

Erna Engelbrecht , Remon M. Rooij and Marcus M. Specht

Delft University of Technology, Mekelweg 5, Delft, 2628 CD, The Netherlands

Abstract

For universities, educational change at institutional level is a slow process [1], [2]. To keep up with societal and technological advancement, education innovation project leaders at universities need practical guidelines and procedures in place that will enable sustainable and scalable innovation that can meet the needs of industry as we transition from Industry 4.0 to Industry 5.0 [3]. To develop such guidelines and procedures, we need to conduct socially responsible, evidence-based educational research [4]. This paper is part of a larger study during which we will conceptualize the planning and evaluation of innovation in engineering education at the Delft University of Technology (TU Delft). From this conceptualization, a framework for planning and evaluation of education innovation will emerge. The data collection process will take place in six phases: (1) Exploration of the problem (2) feasibility studies; (3) conceptualization and development of the framework; (4) piloting of the framework and its associated processes; (5) field study; and lastly, (6) evaluation of the framework. This paper provides an initial overview of the literature, as well as an explanation of the proposed research methodology.

Keywords 1

Innovation, Higher Education, engineering education, research methodology, concept mapping

1. Introduction

The COVID pandemic, conflict with world powers, the consequent fast tracking of energy transition, and the exponential advancement of technology brings about novel problems that need novel solutions. As a consequence, education is in need of transformation [3], [5], [6]. Universities of technology are responsible for the education of engineers who need to be equipped with holistic skill sets for dealing with an increasingly unpredictable future.

Unfortunately, universities are slow to change [1], [2] and innovations are often short-lived [7]. Consequently, time and money is spent with little to no impact, while graduates may find themselves insufficiently prepared to work in an unpredictable and unstable world [8].

There is a need for socially responsible, evidence-based educational research [4] to produce practical guidelines and appropriate measurement instruments that can support sustainable innovation in engineering education that meet the needs of future graduates and an ever-changing society [9]–[11].

In this paper we describe the initial plan for a research initiative during which we will develop a multifaceted innovation framework that can guide the planning and evaluation of innovation initiatives in Higher Engineering Education (HEE). This framework will serve project teams and individuals at all levels, including educators, educational support staff and management. It is envisioned that this

Proceedings of the Doctoral Consortium of Seventeenth European Conference on Technology Enhanced Learning, September 12–16, 2022, Toulouse, France
EMAIL: e.engelbrecht@tudelft.nl (A. 1); r.m.rooij@tudelft.nl (A. 2); m.m.specht@tudelft.nl (A. 3)
ORCID: 0000-0001-8731-347X (A. 1); 0000-0002-2784-1892 (A. 2); 0000-0002-6086-8480 (A. 3)

framework would help to align, for example, its users' goals, expectations, resource allocation and communication flows.

The purpose of this endeavor is to facilitate the feasibility, impact and sustainability of innovations in engineering education. To this end, the following research questions will be addressed:

1. *How can we define the contextual characteristics that influence innovation in HEE?*
2. *How can we conceptualize the planning and evaluation of innovation in engineering education?*
3. *To what extent can this conceptualization be applied to ensure feasibility, sustainability and impact of education innovation that aligns HEE with the needs of society and industry?*

Each research question will be addressed during the different phases of a larger research project. The research questions will be refined after a more in-depth literature review has been conducted.

2. Theoretical background

This study is initiated at a time when a global pandemic, conflict with world leaders, energy transition and data privacy is dominating Western media. The question is whether or not continuation of our current education system will suffice in preparing our engineering students for such an unpredictable and insecure future. For example, the COVID pandemic led to a shift in how many companies do business, and pushed industry and education towards online and hybrid methods. At the same time emergency energy transition plans are being developed as a consequence of the conflict in Eastern Europe.

What kind of engineering professionals do we need in such a rapidly changing world? What kind of curriculum agility do we need in these kinds of circumstances? Does the engineering education community need to wait for the next crisis for large scale innovation and fundamental changes to take place?

This review of the literature first provides a brief introduction to why innovation in engineering education is needed. Next, the facilitation of innovation and the consequences of unguided, unsupported innovation is discussed. We then look at a number of existing frameworks for innovation and the evaluation thereof, before positioning the current study.

2.1 Why innovate?

There are various definitions of innovation discussed in detail in the literature [12]–[14]. For the purpose of this study, however, education innovation will be defined as: *Any change that significantly increases the impact on education processes.*

This initial definition will be further informed and refined as the research project develops. Currently, the definition is purposefully open to interpretation to allow for flexibility and freedom for exploration until a more comprehensive definition emerges.

Why is innovation in engineering education needed? The world is changing fast due to societal and technological developments, and HEE needs to keep up the pace. Some authors [15]–[17] argue that a new type of engineering graduate is needed for taking on global problems in an unpredictable and probably unstable future [8] as we transition to Industry 5.0 [3], [18], [19]. There are more works providing a lengthier discussion on this matter [8], [13], [20], however, we will briefly touch on it here as well. This is not to say that we can predict the future to determine with accuracy what skills our (future) graduates will need – we can only make educated guesses.

The literature speculates, for example, on the significance of automation, the Internet-of-Things, Artificial Intelligence, and big data [21]–[23].

In addition to technological developments, there are also growing concerns of global problems such as data privacy, climate change, pollution, food insecurity and a need for energy transition. Our 'educated guessing' could therefore focus on tasks that cannot (yet) be performed by machines, or tasks performed in collaboration with machines that require human intervention, for example, critical thinking and ethical decision-making.

Furthermore, our graduates will also need durable skills such as digital literacy, analytical thinking, resilience and problem-solving [3], [6], [18].

Education innovation not only happens top-down (instruction from institutional and faculty managers, program leaders, lawmakers and policy makers), but also takes place bottom-up. These innovations are often driven by educators or course teams, student feedback, changes in the field (and consequent updating of course content), funding (or lack thereof) and/or increase in student numbers. Such innovations tend to be introduced incrementally, which might lead to loss of coherence within the program [1].

To keep programs up to date, course content, curricula and teaching methods need coordinated renewal strategies. In fact, not only do we need renewal, but more fundamental transformation is needed to ensure coherence in curricula that equips our graduates with the skills needed to face our (rapidly changing) real world problems.

2.2 Facilitation of innovation

At the start of the pandemic we found ourselves in an emergency situation where we were forced to find alternative methods for conducting everyday business. Many educators hastened to get their courses online, while others were more reluctant to adapt, hoping that life would get back to normal soon. During this time, institutions were forced to adjust and innovate quickly. At TU Delft, pockets of innovation initiatives became more visible as practitioners were trying to find alternatives and reaching out for help. However, most of these initiatives were somewhat painful, uncoordinated, and sporadic at best, since there was no emergency plan in place.

Educators who have been teaching using the blended course format seemed to have adapted more quickly to the situation than those who were newer to online education [24]. The authors go on to explain that centralized support initiatives were emerging, and as the pandemic progressed, an increasing amount of cooperation and exchange of information was observed. Unfortunately, communication thereof did not always seem to reach those who needed it [24].

One example of this is the large number of educators opting to use Zoom for presenting their lectures online, despite it neither having been an approved, nor centrally supported at TU Delft. In fact, the sheer number of Zoom users was so overwhelming that the university was forced to negotiate licensing agreements with the service provider, and produce guidelines for best practices.

At the time of writing, there were plans for eventually phasing out many of these ‘emergency online education’ tools and replacing them with policy compliant alternatives. In hindsight, what was needed was a framework for educators and support personnel to evaluate the feasibility and suitability of the tool; guidance for good practices during usage; and eventually making informed decisions by evaluating how it was used, its impact, and to determine how to go forward. Addressing this need will be the main objective of this study.

The intention here would not be to create an additional hurdle, but rather to equip practitioners with a framework for making better decisions that are more sustainable in the long run in all aspects of the education process. The framework should open communication lines between various levels of stakeholders to ensure feasibility, impact, sustainability, and dissemination of education innovations in the engineering domain.

2.3 Scoping existing education innovation evaluation frameworks

To position this research initiative in the research field, an initial literature search was done using Google Scholar. This was chosen to get a general idea of what is already available on this topic. Once the research project has been approved, a more rigorous search will be conducted, as described further on in *Research methodology* in section 3.

In this section we will provide a brief introduction to five evaluation frameworks. The overview will identify similarities and differences in the elements which the frameworks consist of, as well as any patterns that might emerge.

By investigating formative, summative and illuminative evaluation goals, a 10-step process model was proposed [25] which defines the stages in the process of evaluating education innovations. According to this model, both the academic context and the governing policies need to be taken consideration in the first stage, as these can have a ‘significant impact on innovative practices’.

When defining the academic context, the author included the curriculum, the teaching processes, and learning. In terms of policy, both policies at institutional level, as well as policies that govern the tertiary education sector were taken into account. This initial step of defining the context and policy framework is then followed by defining the goals of the evaluation; identification of stakeholders; aspects of the innovation and criteria for evaluation; data collection and analysis; as well and dissemination of the findings.

Another process-based framework [2] maps out the process of innovation in Higher Education, and includes the following:

- Identifying the current stage of the innovation implementation process and associated challenges. The stages are (1) recognition of need, (2) planning, (3) initiating, and (4) institutionalization.
- Determining the aim, type, nature and measures to institutionalize the innovation.
- Identifying the innovation itself, the problem it addresses, and the people involved in the innovation activity.
- Evaluating the learning curve and adjusting aims and methods for institutionalization.
- Analyzing potential factors that might affect institutionalization of an innovation (opportunity, compatibility and agency).

This framework provides a very useful insight on the complexity and instructiveness of the innovation process itself. By taking these elements into account, the framework can provide a starting point for identifying elements for consideration to minimize potential pitfalls that could hinder dissemination of innovations.

[26] attempted to develop a more contextualized evaluation methodology. Although the framework was developed with the purpose of evaluating courses, instead of innovations in education, it is worth looking at the framework to inform the evaluation (application) process of the framework under development in the current study. The framework includes the following aspects: purpose (of the evaluation), content (what to evaluate), usage (by whom the analysis will be done and how the results will be shared), and method (when and how evaluations should be done).

[27] developed a framework that serves to ensure responsible innovation. It informs the framework under development in that it addresses the following four dimensions: anticipation (being in touch with social and technological change), reflexivity (adjusting behavior based on past experiences), inclusion (involving a wider circle of contributors), and responsiveness (adapting in response to changing circumstances). These dimensions align with the underpinning reasons for the need for innovation, discussed earlier in this review, and according to the authors, have emerged from public debate on new developments in science and technology.

[15] developed the Course Innovation Framework (CIF) with which to analyze multiple aspects of course innovation. Aimed at policy makers and educators, this framework provides input for analyzing, mapping out and making decisions on course innovations. Using Curriculum Development Theory [28] as part of the conceptual foundation, the intended, implemented and attained forms of innovation were taken into consideration. Within the CIF framework, different stages of the course innovation life cycle, as well as different processes of innovation are considered. Furthermore, the framework is both informed by the literature and policy (top-down), as well as practice and interviews (bottom-up).

From this brief discussion, the following preliminary conclusions can be drawn:

- The impact on student learning should be one of the main aspects of a framework, as it gives an important indication of the impact of the innovation;
- Stakeholders should be another key element – not only the students, but the educators themselves, and management.
- The institutional context and the policies that apply to it can have implications for the dissemination process of innovations; and
- Education innovations should serve a specific purpose. More strongly put, it should solve a specific problem. The framework should help to conceptualize the problem and how it can be solved.

Based on these points we can already identify important elements that will define 'innovation' in this study. Besides, of course, it being novel, it should have a (positive) impact on its stakeholders, be compliant with policy requirements and be fit for purpose by solving some or other problem.

2.4 Positioning the framework to be developed during this study

Although many authors have investigated innovation evaluation and evaluation frameworks in the past, each of them was conducted within their unique institutional and educational contexts. It could be assumed that the discussion on evaluation frameworks for HEE will continue to evolve organically as the world changes and education follows suit. The present study aims to contribute to this evolution, specifically in the light of global challenges that urgently need to be considered in the renewal and development processes in engineering education.

Both top-down and bottom-up innovation can flourish when managerial support is in place and open communication lines are maintained. If not, innovation initiatives are stifled, making it more difficult (and costly) to bring about change. The intended evaluation framework aims to contribute in that regard: increasing the autonomy and impact of all levels of innovative project leaders, ensuring that their innovations contribute to the shared goals of the degree program and/or institution.

Therefore, the framework to be developed should be comprehensive enough to serve as a multi-stakeholder instrument that can be applied firstly as a forecasting tool to determine education innovations' potential, feasibility and fit within the institutional context and assist in the planning and design phases; secondly to inform the implementation process; and thirdly for the assessment of those innovations in terms of impact, sustainability, and dissemination.

In addition to this, this study aims to contribute to the discussion on fundamental changes needed in engineering education . In an attempt to accomplish this, the framework will be developed in collaboration with various engineering education innovation project leaders. This will be done by building on existing innovation initiatives of educators, and in turn, support with dissemination of their work. Ultimately, a consolidated, multi-stakeholder framework will emerge that can be applied widely across the institution, aligning innovation practice bilaterally.

3. Research methodology

To address the research questions, the data collection for this sequential mixed methods study will be done in six phases. The following table summarizes the phases that will be undertaken in the current study:

Table 1
Research phases, based on **[29]**

Phases	Description
Phase 1: Exploration of the problem through secondary data collection	Systematized literature review, PRISMA Analysis of innovation project documentation
Phase 2: Feasibility study	Testing initial framework design Interview project leaders for feedback Reflection, and implementation of improvements
Phase 3: Primary data collection and analysis; and development of intervention	Group concept mapping in collaboration with project leaders Development of initial framework
Phase 4: Prototyping	Piloting framework Interview/focus group discussions with project leaders/project groups Reflection and implement improvements after each iteration
Phase 5: Field study	Apply framework to innovation initiatives – at least 1 x before, 1 x during and 1 x after implementation of innovation Reflection and implementation of improvements after each iteration
Phase 6: Feedback and reflection	Evaluation of framework

During the first phase, the problem itself and its context will be explored.

This phase aims to address the first research question:

1. How can we define the contextual characteristics that influence innovation in HEE?

A systematized literature review will be conducted for an in-depth theoretical understanding of the context within which innovation in engineering education should take place. Considering the advancement of technology and developments in society at large, education needs to be updated to be able to meet the demand of skills and knowledge needed in the future, as discussed earlier.

The systematized method for literature review will be followed to ensure academic rigor similar to a systematic review, while allowing for some flexibility to complete the review in good time. In fact, a systematized review is recommended for post-graduate research [30].

During Phase 2 we will conduct two feasibility studies. First, we will test the primary data collection process that will take place in Phase 3. After Phase 3 (development of the framework) has been completed, another feasibility study will be conducted to test the implementation process and usability of the framework itself (in effect extending Phase 2 beyond Phase 3). Improvements will be made by reflecting on how the process went, and based on interviews with participants of the feasibility studies.

During Phase 3 the primary, mixed method data will be collected by means of Concept Mapping [31]. Here, project leaders will be guided through a brainstorming session to generate ideas on how the planning and evaluation of innovations should be conducted. These ideas will then be analyzed by means of a cluster analysis and multidimensional scaling to sort, rank and map the ideas. Use of this technique enables the researcher to fill gaps where knowledge is incomplete or uncertain by collecting information which a group of experts have reached consensus on [31].

Based on this conceptualization, a framework for education innovation will be developed. Phase 3, therefore, will aim to address the second research question:

2. How can we conceptualize the planning and evaluation of innovation in engineering education?

During Phases 3 – 5, the research participants will consist of the project leaders from innovation initiatives at TU Delft. Project leaders can include Educators, Educational Advisors and Managers from the eight TU Delft faculties and the department of Teaching and Learning Services (TLS) at TU Delft. The selection of education innovations which the participants are involved in will be made to include, but are not limited to, for example, education technology, teaching methodology, learning environments, and course content. During Phase 1 of the study, a list will be drafted of participants to include, from which they will be selected. During the selection process, the optimal number of participants will be decided on to get a fair demographic representation of participants, their innovation initiatives and the phases they are in.

Phases 4 – 6 will focus on the third research question:

3. To what extent can this conceptualization be applied to ensure feasibility, sustainability and impact of education innovation that aligns HEE with the needs of society and industry?

This leads us to Phase 4, where application of the evaluation framework will be piloted on a small scale on education innovation cases to test for feasibility, applicability and impact of the framework. This will be followed by focus groups/interviews involving project leaders and peers for the purpose of feedback and reflection for improvement, before continuing onto the next phase. The data will be analyzed, based on which preliminary conclusions can be drawn.

Then, during the fifth phase, the field study will be carried out by applying the framework to education innovation initiatives. Innovations for this study will be chosen based on the phases that they are in – before, during, and after implementation.

For Phases 4 – 5, at least three iterations will be done, starting with simpler innovations with a small scope, and then scaling up to larger innovation initiatives. The size and scope of the initiatives will be determined relative to each other and can be as simple as, for example (hypothetically speaking), using a new tool for a single activity vs. migration to a new learning management system.

Lastly, Phase 6 will follow, where the framework will be evaluated by means of questionnaires. The questionnaires will be sent to project leaders and other stakeholders to evaluate the usefulness, impact (internal and external), and validity of the framework. Project leaders as well as Comenius and Education Fellows from the 4TU (four Universities of Technology in the Netherlands) will be included during Phase 6. The evaluation process will be done for all three stages of innovation projects – before, during and after implementation.

This process will be repeated until the framework is sufficiently validated.

Any problems experienced, or points for improvement during iterations, will be dealt with before moving on to the next iteration. Additional iterations will be added if it is found that three iterations are insufficient to draw strong conclusions, or if an iteration has failed for some reason or another.

By combining qualitative and quantitative data, a holistic view of the feasibility, impact, sustainability, and dissemination of innovations that are guided by the evaluation framework can be captured. As explained, this will be conducted in iterations, with moments for reflection for improvement in-between phases.

4. Ethical considerations and data management

The research will not impact on human subjects and there is no foreseen conflict of interest or risk involved. A detailed data management plan will be drawn up in consultation with a TU Delft Data Steward. The data management plan will detail how the data will be indexed and made accessible, and reusable. All data collected during this research initiative will be stored on a password protected database on the TU Delft server, as well as the 4TU.ResearchData[2] repository for scientific research data in the Netherlands.

5. Dissemination of research

The research progress and results will be shared at conferences, journal publications, poster presentations and workshops. The main topics intended are as follows:

- Literature review – innovation trends and contexts, and the way forward
- Research methodology
- Data collection, analysis and discussion of results;
- Literature review on innovation frameworks and comparison with own intervention;
- Application of the intervention developed, and discussion of feedback received on its application; and
- Evaluation of intervention and discussion of final results of the study.

Furthermore, cross-departmental sessions will be held to share progress and new insights with Teaching and Learning Services (TLS) at TU Delft. Lastly, workshops will be provided to other PhD candidates on lessons learned during the research process.

6. Conclusion

This study will attempt to conceptualize the process and evaluation of innovation needed to meet the demand of industry and society. This conceptualization will serve project leaders of innovation initiatives both bilaterally and during the planning and evaluation phases of their innovation initiatives.

By providing the right support, tools and processes in place for planning and evaluating innovation, educators and teaching teams will be more equipped to implement feasible, sustainable and meaningful educational change that will enable us to train holistically educated engineers.

7. Acknowledgements

The research initiative is funded by the 4TU Centre for Engineering Education (CEE) and will be conducted in collaboration with Marcus Specht as Supervisor and Remon Rooij as Promotor.

8. References

[1] C. Bajada, P. Kandlbinder, and R. Trayler, 'A general framework for cultivating

[2] http://researchdata.4tu.nl

innovations in higher education curriculum', *High. Educ. Res. Dev.*, vol. 38, no. 3, pp. 465–478, Apr. 2019, doi: 10.1080/07294360.2019.1572715.

[2] Y. Cai, 'From an Analytical Framework for Understanding the Innovation Process in Higher Education to an Emerging Research Field of Innovations in Higher Education', *Rev. High. Educ.*, vol. 40, no. 4, pp. 585–616, 2017, doi: 10.1353/rhe.2017.0023.

[3] S. GüğerçiN, 'How Employees Survive In The Industry 5.0 Era: In-Demand Skills Of The Near Future', *Int. J. Discip. Econ. Adm. Sci. Stud. IDEAstudies*, vol. 7, no. 31, pp. 524–533, Jan. 2021, doi: 10.26728/ideas.452.

[4] T. C. Reeves, J. Herrington, and R. Oliver, 'Design research: A socially responsible approach to instructional technology research in higher education', *J. Comput. High. Educ.*, vol. 16, no. 2, pp. 96–115, Mar. 2005, doi: 10.1007/BF02961476.

[5] L. Guàrdia, D. Clougher, T. Anderson, and M. Maina, 'IDEAS for Transforming Higher Education: An Overview of Ongoing Trends and Challenges', *Int. Rev. Res. Open Distrib. Learn.*, vol. 22, no. 2, pp. 166–184, Jan. 2021, doi: 10.19173/irrodl.v22i2.5206.

[6] S. Grabowska, S. Saniuk, and B. Gajdzik, 'Industry 5.0: improving humanization and sustainability of Industry 4.0', *Scientometrics*, vol. 127, no. 6, pp. 3117–3144, Jun. 2022, doi: 10.1007/s11192-022-04370-1.

[7] C. J. Bonk and D. A. Wiley, 'Preface: Reflections on the waves of emerging learning technologies', *Educ. Technol. Res. Dev.*, vol. 68, no. 4, pp. 1595–1612, Aug. 2020, doi: 10.1007/s11423-020-09809-x.

[8] A. Kamp, *Engineering Education in the Rapidly Changing World: Rethinking the Vision for Higher engineering Education*. TU Delft, Faculty of Aerospace Engineering, 2016.

[9] S. Brooks, S. Brown, D. Davis, and J. Lebeau, 'Adoption of Technological Innovations: A Case Study of the ASSESS Website', *Adv. Eng. Educ.*, vol. 4, no. 1, 2014.

[10] T. C. Reeves and L. Lin, 'The research we have is not the research we need', *Educ. Technol. Res. Dev.*, vol. 68, no. 4, pp. 1991–2001, Aug. 2020, doi: 10.1007/s11423-020-09811-3.

[11] T. C. Reeves and P. M. Reeves, 'Reorienting educational technology research from things to problems', *Learn. Res. Pract.*, vol. 1, no. 1, pp. 91–93, Jan. 2015, doi: 10.1080/23735082.2015.1008120.

[12] C. J. Fraser, G. Duignan, D. Stewart, and A. Rodrigues, 'Overt and Covert: Strategies for Building Employability Skills of Vocational Education Graduates', *J. Teach. Learn. Grad. Employab.*, vol. 10, no. 1, pp. 157–172, 2019.

[13] F. E. H. M. Smulders, A. Kamp, and C. Fortin, 'The CDIO framework and new perspectives on technological innovation', presented at the The 14th International CDIO Conference, 2018. [Online]. Available: https://repository.tudelft.nl/islandora/object/uuid%3A27702436-f7d8-4cc9-ad09-ebb3f9020d01

[14] S. G. Vincent-Lancrin, G. Jacotin, J. Urgel, S. Kar, and C. González-Sancho, *Measuring Innovation in Education: A Journey to the future*. Paris: OECD Publishing, 2017.

[15] V. C. Tassone, H. J. A. Biemans, P. den Brok, and P. Runhaar, 'Mapping course innovation in higher education: a multi-faceted analytical framework', *High. Educ. Res. Dev.*, pp. 1–15, Oct. 2021, doi: 10.1080/07294360.2021.1985089.

[16] P. den Brok, *Cultivating the growth of life-science graduates: on the role of educational ecosystems*. 2018. Accessed: Feb. 18, 2022. [Online]. Available: https://doi.org/10.18174/458920

[17] A. Kamp, *Navigating the landscape of higher engineering education: coping with decades of accelerating change ahead.* 2020.

[18] J. Mitchell and D. Guile, 'Fusion Skills and Industry 5.0: Conceptions and Challenges', in *Insights Into Global Engineering Education After the Birth of Industry 5.0*, M. Bouezzeddine, Ed. IntechOpen, 2022. doi: 10.5772/intechopen.100096.

[19] *SOCIETY 5.0: a people-centric super-smart society.* Place of publication not identified: SPRINGER Verlag, SINGAPOR, 2020.

[20] R. Klaassen, M. Dijk, R. Hoope, D. Ceulemans, A. Kamp, and M. Van der Sanden, 'A DESIGN-BASED VISION ON FUTURE ROLES IN ENGINEERING', 2018.

[21] J. Grodotzki, T. R. Ortelt, and A. E. Tekkaya, 'Remote and Virtual Labs for Engineering Education 4.0', *Procedia Manuf.*, vol. 26, pp. 1349–1360, 2018, doi: 10.1016/j.promfg.2018.07.126.

[22] M. Hernandez-de-Menendez, C. A. Escobar Díaz, and R. Morales-Menendez,

‘Engineering education for smart 4.0 technology: a review’, *Int. J. Interact. Des. Manuf. IJIDeM*, vol. 14, no. 3, pp. 789–803, Sep. 2020, doi: 10.1007/s12008-020-00672-x.

[23] R. A. Ramirez-Mendoza, R. Morales-Menendez, H. Iqbal, and R. Parra-Saldivar, ‘Engineering Education 4.0: — proposal for a new Curricula’, in *2018 IEEE Global Engineering Education Conference (EDUCON)*, Tenerife, Apr. 2018, pp. 1273–1282. doi: 10.1109/EDUCON.2018.8363376.

[24] N. L. Wahls, GM. Dijkstra, and G. M. Ouwehand, ‘Blending Your Education: Lessons Learned During COVID’, 2022. [Online]. Available: https://digitelpro.eadtu.eu/images/Blending_Your_Education__Lessons_Learned_During_COVID.pdf

[25] C. Jacobs, ‘The Evaluation of Educational Innovation’, *Evaluation*, vol. 6, no. 3, pp. 261–280, Jul. 2000, doi: 10.1177/13563890022209280.

[26] C. Nygaard and D. Z. Belluigi, ‘A proposed methodology for contextualised evaluation in higher education’, *Assess. Eval. High. Educ.*, vol. 36, no. 6, pp. 657–671, Oct. 2011, doi: 10.1080/02602931003650037.

[27] J. Stilgoe, R. Owen, and P. Macnaghten, ‘Developing a framework for responsible innovation’, *Res. Policy*, vol. 42, no. 9, pp. 1568–1580, Nov. 2013, doi: 10.1016/j.respol.2013.05.008.

[28] J. van den Akker, ‘Curriculum Perspectives: An Introduction’, in *Curriculum Landscapes and Trends*, Dordrecht: Springer Netherlands, 2004, pp. 1–10. doi: 10.1007/978-94-017-1205-7_1.

[29] A. E. Kelly, R. A. Lesh, and J. Y. Baek, Eds., *Handbook of design research methods in education: innovations in science, technology, engineering, and mathematics learning and teaching*. New York : London: Routledge, 2008.

[30] M. J. Grant and A. Booth, ‘A typology of reviews: an analysis of 14 review types and associated methodologies: A typology of reviews, *Maria J. Grant & Andrew Booth*’, *Health Inf. Libr. J.*, vol. 26, no. 2, pp. 91–108, Jun. 2009, doi: 10.1111/j.1471-1842.2009.00848.x.

[31] W. M. K. Trochim, ‘An introduction to concept mapping for planning and evaluation’, *Eval. Program Plann.*, vol. 12, no. 1, pp. 1–16, Jan. 1989, doi: 10.1016/0149-7189(89)90016-5.

Measuring and supporting self-regulated learning in blended learning contexts

Esteban Villalobos[1,*], Mar Pérez-Sanagustin[1], André Tricot[2] and Julien Broisin[1]

[1]IRIT, Université de Toulouse, CNRS, Toulouse INP, UT3, Toulouse, France
[2]Université Paul-Valéry Montpellier 3, EPSYLON

Abstract
Despite the positive effects of Blended Learning (BL), several studies have shown that students require high levels of self-regulation to succeed in these types of practices. Still, there is little understanding of how students organize their learning in BL authentic contexts. This paper presents the objectives and current status of a project that seeks to understand how students' Self-regulated Learning (SRL) strategies manifest themselves in BL contexts holistically and how to foster it through technological solutions. The contributions of this project will be three-fold. First, we aim to develop novel analytical and technological solutions to understand better the dynamics of how self-regulated learning unveils in BL contexts. Second is the development of a dashboard-based support tool for students and teachers. And third, we will provide evaluations of the analytical framework and support tool in authentic BL contexts. We expect that these contributions will provide the community with a better understanding of the dynamics of SRL in BL.

Keywords
Self-regulated Learning, Blended Learning, Learning Analytics

1. Introduction

In the last few years, we have seen Blended Learning (BL) approaches becoming more varied and commonly applied [1]. This methodology consists in combining online and traditional in-person activities [2]. Nonetheless, while BL has been shown to have positive effects on learning, many students often have problems regulating their study [3, 4, 2]. This has prompted a growing interest in finding out how to understand and support students' self-regulation abilities in BL.

Self-regulated Learning (SRL) is defined as a complex process that combines meta-cognitive, motivational, and emotional processes [5]. Recent literature shows that students' SRL ability is a good predictor of their behavior and success in a course [6]. However, most studies on SRL have been conducted in online contexts and little is known about how these processes manifest in BL [3]. Recent works show that students' SRL manifests differently depending on pedagogical decisions, such as the learning context and course modality [3, 7, 8, 9]. For example, Matcha et al. [9] compared students' strategies in a BL course, in a Flipped Classroom (FC), and in Massive Open Online Courses (MOOCs), showing that students used similar strategies in BL and FC modalities, but these differed from the tactics used in MOOCs. Moreover, [3] showed that BL students used SRL strategies less often than online students. Overall, there seems to be a strong connection between the course design, the learners' SRL ability profile, and the learning strategies in the course [9, 7].

To support students' SRL, researchers propose different mechanisms. One of these mechanisms is using dashboard-based tools. These tools provide learners with information about their progress. Although most of these tools have been designed and evaluated in online environments with encouraging results [10], only a few works show how students incorporate them into their learning strategies and have an impact on their behavior in BL courses [11, 12].

In order to give meaningful SRL support in BL it is important to understand how different external factors (e.g., the influence of the teacher or face-to-face classes) and internal factors (e.g., students' self-regulation abilities) affect learners in these contexts. These factors influence how students will interact with the learning material along the course. This represents a particular challenge in TEL, as it implies that strategies observed will be heavily influenced by the dynamics of the system in which the students operate [13, 14]. This points out the need to develop new holistic approaches to understand the SRL behavior of the students better.

This work is part of a 3-year thesis starting in October 2021, in which we expect to contribute to the TEL domain by addressing these gaps. Specifically, we propose: (1)

Proceedings of the Doctoral Consortium of the Seventeenth European Conference on Technology Enhanced Learning, September 12–16, 2022, Toulouse, France.
*Corresponding author.
esteban.villalobos@irit.fr (E. Villalobos); mar.perez-sanagustin@irit.fr (M. Pérez-Sanagustin); andre.tricot@univ-montp3.fr (A. Tricot); julien.broisin@irit.fr (J. Broisin)
0000-0002-6026-3756 (E. Villalobos); 0000-0001-9854-9963 (M. Pérez-Sanagustin); 0000-0003-4005-7338 (A. Tricot); 0000-0001-8713-6282 (J. Broisin)

studying new analytics techniques to understand the development of SRL strategies in BL holistically and (2) developing technological solutions to support SRL in BL.

2. Objectives and research questions

The general objective of this project is to investigate the SRL strategies used by learners in BL scenarios and to propose and evaluate a Learning Analytics (LA) technological solution based on user-centered dashboards (for teachers and students) to support those strategies that maximize learners' performance. Three main objectives are derived from this general objective:

- **Objective 1**: To propose an analytical framework to study in a holistic manner how students' SRL strategies manifest in BL contexts.
- **Objective 2**: To design a LA dashboard-based solution for teachers and students to support SRL in BL.
- **Objective 3**: To evaluate the impact of LA solution on students' learning strategies and teachers' decision-making in BL scenarios.

2.1. Measuring SRL in BL

Different methods have been proposed for studying how SRL manifests in different learning contexts, especially in online learning environments. These range from using self reported data [15] to detecting tactics and strategies by using the trace data collected from the course's LMS [16, 7, 17, 18, 9]. The latter has seen many contributions from the field of Learning Analytics (LA). Some examples of these analytical approaches have used techniques derived from temporal analysis and sequence mining [17, 16]. Some studies have also made the connection between these techniques and the SRL theory [16]. Fan et al. [16] suggests this theoretical backbone may allow us to overcome the limitations of the context-specific nature of LA to perform pedagogical interventions that go beyond course setting.

Most of these methods have been applied in online settings, and very few have been applied in Blended Learning settings. The currently applied methods are limited in capturing the impact of factors such as teacher interventions and face-to-face classes. In fact, current research applying existing methods in Blended Learning encounters difficulties in providing indicators on run-time, as well as in giving a temporal meaning to the collected data. From this, we derive the following research question:

- **RQ1:** How can pre-existing LA methods and techniques be adapted and combined with qualitative methods to create an analytical framework for characterizing the dynamics of students' strategies in BL?

2.2. Supporting SRL in BL

Researchers have proposed different approaches to support students' SRL processes [19]. The most common approaches explored are educational prompts and integrated support systems [20]. These solutions transform raw data into 'actionable insights' to produce behavioral changes in the students [21]. So far, most of this prior work has been conducted in online settings, such in Massive Open Online Courses (MOOCs), in which students have low interaction with the teacher [20]. These studies suggest that dashboards could be an appropriate approach for supporting SRL strategies. In particular, the strategies of goal setting, strategic planning, time management, and monitoring have been shown to be more effective for promoting students' motivation and impact on course performance.

There are still very few studies looking at these solutions BL contexts (e.g., [22, 23]). These works in BL have two main limitations. First, the tools focus on supporting the students directly, usually overlooking the role of the teacher. Second, while some tools are based on theoretical models for SRL, there is still much to understand about their impact on students' SRL strategies. This posses the following research questions for the project:

- **RQ2:** How useful (interpretable, actionable, and comprehensive) are the existing indicators provided in the SRL-support dashboard for students and teachers?
- **RQ3:** How do SRL support tools influence students' strategies and teachers' decision-making in BL scenarios?

3. Project Methodology

Design Based Research (DBR) will be used as a methodological approach, which combines experiments in real-world settings with theoretical models [24]. The interventions will be based on the NoteMyProgress (NMP) tool [25], a Moodle plug-in that delivers dashboards with self-regulation indicators in the course to both students and teachers (see Figure 1). Three experimental cycles will be carried out to improve the tool and the analytical frameworks in an iterative way. After each cycle, the results will be published as part of the LASER project following an Open Science Framework.

Figure 1: Examples of visualizations in the NoteMyProgress plug-in

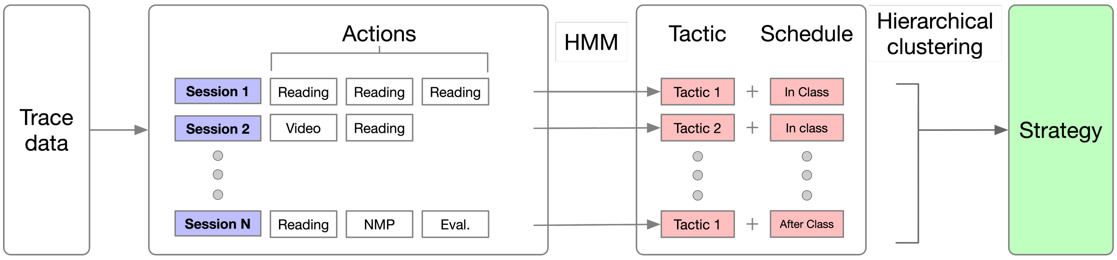

Figure 2: Analytical approach used to evaluate the first design cycle [26].

4. Current Results: First Design Cycle

The first cycle focused on studying students' behavior in BL. This cycle had three research questions:

1. How do students' learning tactics and strategies manifest along the BL course?
2. Does the NMP tool, designed to support students' SRL, have an effect on their learning tactics and strategies?
3. Is there a relationship between students' learning strategies, course performance, and SRL ability profile?

This intervention took place between September 2021 and January 2022. The study consisted on 241 students from two university courses. At the beginning of the course, students completed the informed consent for participation and a questionnaire to assess their level of SRL. Midway through the course (week 6), they were introduced to NMP and invited to refer to it to assess their study strategies [27]. At the end of the course, they were asked to complete a questionnaire on their sense-making of the tool [25].

The evaluation of one of these courses is detailed in [26]. Here, we extended an analytical approach proposed in Fincham et al. [17] and analyzed the results with respect to students' SRL ability profile, final performance, and previous achievements. The approach consists of the following steps:

1. **Separating the activity of the students into sessions.** These correspond to a sequence of actions not separated by more than 30 minutes of inactivity.
2. **Detecting the underlying tactic of each session**. A tactic is defined as the underlying process that a student is applying in a given period of time [17]. We used a Hidden Markov Model (HMM) in order to detect students' tactics.
3. **Detecting students' strategies**. Under the analytical approach proposed by Fincham et al. [17], strategies are defined as sequences of tactics applied by the students. In order to include the context of the BL course, we included in this model the timing with respect to the face-to-face sessions.
4. **Analyzing relationships between strategies and students' profile**. We analyzed how different tactics and strategies applied by the students related to their SRL ability profile, course performance, and previous achievements.

We found that students' strategies were correlated with their previous achievements (GPA) and their self-reported Self-Regulation ability. We also found that the tactics used by the students varied across modalities and

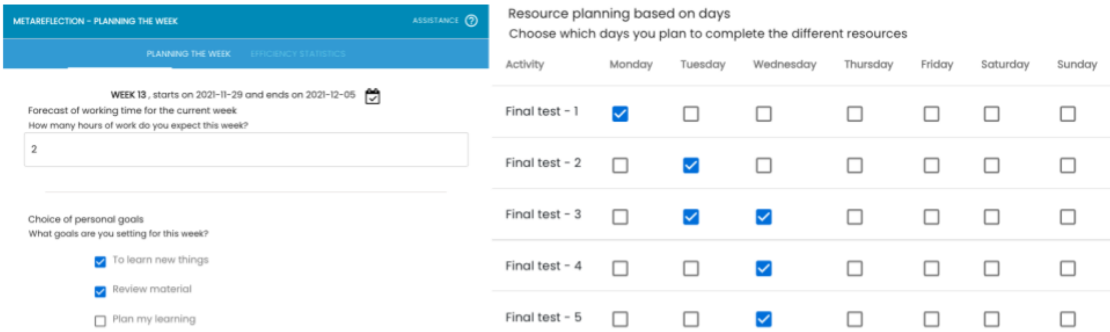

Figure 3: Examples of the 'Student planning and goal setting' functionalities added to NMP

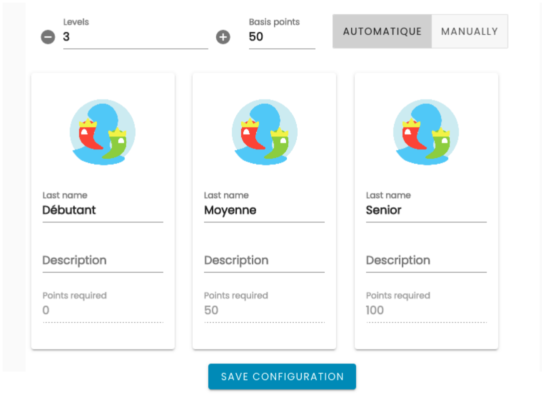

Figure 4: Example of the 'Gamification' functionalities added to NMP

were based on the pedagogical decisions of the course. In terms of the usage of NMP we found that even though some students incorporated the SRL support tool into their learning tactics, the use of the tool was relatively sparse. We also found that, even if the use of the tool was not mandatory, most of the students interacted with the indicators relating to Strategic Planning.

While this gives us some insight into the performance of the students in the course, this methodology still has some limitations. Mainly, since the methods applied are "memory-less", we are losing information on the temporal dynamics of the events. Also, this methodology only allows us to do a retrospective analysis of the course. This limits our capability to perform meaningful interventions on run-time.

5. Future work: Second Design Cycle

The second design cycle focuses on the role of the teacher in the BL course, as well as on students' behavior when they use support for planning their course. This cycle will take place between September 2022 and January 2023. Based on the insights from the first cycle, new developments were made to NMP. We developed new functionalities of student planning and goal setting (see Figure 3), and gamification (see Figure 4).

We aim to evaluate this intervention based on the temporal dynamics of the students. Our goal is to understand how external factors (such as feedback and gamification) and internal factors (such as student planning) affect the students' SRL behavior. Following the recent works by [14, 28, 29], we will study how context-dependent and context-independent indicators behaviors throughout the course and their potential to give meaningful information to students and teachers. In the short term, we will be following behavior-based indicators already studied in the literature to provide students feedback week to week. In the long term, we are looking to develop indicators based on point processes to capture more complex temporal behavior from the students. This study will be done in collaboration with the Millennium Nucleus Student Experience in Higher Education in Chile (NMEDSUP) to see how this work can be extended to different institutions and contexts.

6. Contribution to TEL domain

This work aims at advancing research in TEL, and in particular in the study of SRL in BL scenarios, with three contributions. Firstly, we expect to provide the community with an analytical framework for understanding the dynamics of SRL in BL in a holistic manner and taking into consideration temporal aspects. These tools will help in analyzing data but also in proposing indicators that could serve researchers doing interventions on run-time. Second, we contribute with the NMP tool, a functional tool that both teachers and students could use to support SRL, and its evaluation in authentic contexts. The current version of the tool is already openly available[1]. And third, we expect to contribute with exemplary scenarios on how to apply our analytical framework in BL.

These contributions will have implications at the theoretical level, the analytical level, and the teaching practices level. We expect that our analytical framework and proposed tool can give the community greater insights into how to understand the different factors that affect the dynamics of SRL in BL. We hope that this allows the community to have a better understanding of how to support SRL in a holistic manner.

Acknowledgments

This paper has been partially funded by the ANR LASER (156322). The authors acknowledge PROF-XXI, which is an Erasmus+ Capacity Building in the Field of Higher Education project funded by the European Commission (609767-EPP-1-2019-1- ES-EPPKA2-CBHE-JP). This publication reflects the views only of the authors and funders cannot be held responsible for any use which may be made of the information contained therein.

References

[1] K. Pelletier, M. McCormack, J. Reeves, J. Robert, N. Arbino, 2022 EDUCAUSE Horizon Report, Teaching and Learning Edition (2022) 58.

[2] C. R. Graham, Blended learning systems: Definition, current trends, future directions, in: Handbook of blended learning: Global Perspectives, local designs, San Francisco, CA: Pfeiffer Publishing, Brigham Young University, USA, 2004.

[3] J. Broadbent, Comparing online and blended learner's self-regulated learning strategies and academic performance, The Internet and Higher Education 33 (2017) 24–32. URL: https://www.sciencedirect.com/science/article/pii/S1096751617300398. doi:10.1016/j.iheduc.2017.01.004.

[4] J. Broadbent, M. Fuller-Tyszkiewicz, Profiles in self-regulated learning and their correlates for online and blended learning students, Educational Technology Research and Development 66 (2018) 1435–1455. URL: http://link.springer.com/10.1007/s11423-018-9595-9. doi:10.1007/s11423-018-9595-9.

[5] E. Panadero, A Review of Self-regulated Learning: Six Models and Four Directions for Research, Frontiers in Psychology 8 (2017) 422. URL: http://journal.frontiersin.org/article/10.3389/fpsyg.2017.00422/full. doi:10.3389/fpsyg.2017.00422.

[6] J. Maldonado-Mahauad, M. Pérez-Sanagustín, P. M. Moreno-Marcos, C. Alario-Hoyos, P. J. Muñoz-Merino, C. Delgado-Kloos, Predicting Learners' Success in a Self-paced MOOC Through Sequence Patterns of Self-regulated Learning, in: V. Pammer-Schindler, M. Pérez-Sanagustín, H. Drachsler, R. Elferink, M. Scheffel (Eds.), Lifelong Technology-Enhanced Learning, volume 11082, Springer International Publishing, Cham, 2018, pp. 355–369. URL: http://link.springer.com/10.1007/978-3-319-98572-5_27. doi:10.1007/978-3-319-98572-5_27, series Title: Lecture Notes in Computer Science.

[7] Y. Fan, W. Matcha, N. A. Uzir, Q. Wang, D. Gašević, Learning Analytics to Reveal Links Between Learning Design and Self-Regulated Learning, International Journal of Artificial Intelligence in Education 31 (2021) 980–1021. URL: https://link.springer.com/10.1007/s40593-021-00249-z. doi:10.1007/s40593-021-00249-z.

[8] D. Gašević, N. Mirriahi, S. Dawson, S. Joksimović, Effects of instructional conditions and experience on the adoption of a learning tool, Computers in Human Behavior 67 (2017) 207–220. URL: https://linkinghub.elsevier.com/retrieve/pii/S0747563216307270. doi:10.1016/j.chb.2016.10.026.

[9] W. Matcha, D. Gašević, N. Ahmad Uzir, J. Jovanović, A. Pardo, L. Lim, J. Maldonado-Mahauad, S. Gentili, M. Pérez-Sanagustín, Y.-S. Tsai, Analytics of Learning Strategies: Role of Course Design and Delivery Modality, Journal of Learning Analytics 7 (2020) 45–71. URL: https://learning-analytics.info/index.php/JLA/article/view/7008. doi:10.18608/jla.2020.72.3.

[10] R. Pérez-Álvarez, J. Maldonado, M. Pérez-Sanagustín, Tools to Support Self-Regulated Learning in Online Environments: Literature Review: 13th European Conference on Technology Enhanced Learning, EC-TEL 2018, Leeds, UK, September 3-5, 2018, Proceedings, 2018, pp. 16–30. doi:10.1007/978-3-319-98572-5_2.

[11] M. Pérez-Sanagustín, D. Sapunar-Opazo, R. Pérez-Álvarez, I. Hilliger, A. Bey, J. Maldonado-Mahauad, J. Baier, A MOOC-based flipped experience: Scaffolding SRL strategies improves learners' time management and engagement, Computer Applications in Engineering Education 29 (2021) 750–768. URL: https://onlinelibrary.wiley.com/doi/10.1002/

[1] https://gitlab.com/laser-anr/notemyprogress-plug-in

cae.22337. doi:10.1002/cae.22337.

[12] M. Yoon, J. Hill, D. Kim, Designing supports for promoting self-regulated learning in the flipped classroom, Journal of Computing in Higher Education 33 (2021) 398–418. URL: https://link.springer.com/10.1007/s12528-021-09269-z. doi:10.1007/s12528-021-09269-z.

[13] S. Dawson, S. Joksimovic, O. Poquet, G. Siemens, Increasing the Impact of Learning Analytics, in: Proceedings of the 9th International Conference on Learning Analytics & Knowledge, ACM, Tempe AZ USA, 2019, pp. 446–455. URL: https://dl.acm.org/doi/10.1145/3303772.3303784. doi:10.1145/3303772.3303784.

[14] J. Jovanović, M. Saqr, S. Joksimović, D. Gašević, Students matter the most in learning analytics: The effects of internal and instructional conditions in predicting academic success, Computers & Education 172 (2021) 104251. URL: https://linkinghub.elsevier.com/retrieve/pii/S0360131521001287. doi:10.1016/j.compedu.2021.104251.

[15] M. Zhou, P. H. Winne, Modeling academic achievement by self-reported versus traced goal orientation, Learning and Instruction 22 (2012) 413–419. URL: https://linkinghub.elsevier.com/retrieve/pii/S0959475212000217. doi:10.1016/j.learninstruc.2012.03.004.

[16] Y. Fan, J. Saint, S. Singh, J. Jovanovic, D. Gašević, A learning analytic approach to unveiling self-regulatory processes in learning tactics, in: LAK21: 11th International Learning Analytics and Knowledge Conference, ACM, Irvine CA USA, 2021, pp. 184–195. URL: https://dl.acm.org/doi/10.1145/3448139.3448211. doi:10.1145/3448139.3448211.

[17] E. Fincham, D. Gašević, J. Jovanović, A. Pardo, From Study Tactics to Learning Strategies: An Analytical Method for Extracting Interpretable Representations, IEEE Transactions on Learning Technologies 12 (2019) 59–72. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85045204080&doi=10.1109%2fTLT.2018.2823317&partnerID=40&md5=84a758598f985bd200d40eab1fb0e45c. doi:10.1109/TLT.2018.2823317.

[18] D. Gasevic, J. Jovanovic, A. Pardo, S. Dawson, Detecting Learning Strategies with Analytics: Links with Self-reported Measures and Academic Performance, Journal of Learning Analytics 4 (2017). URL: https://learning-analytics.info/index.php/JLA/article/view/5085. doi:10.18608/jla.2017.42.10.

[19] A. Devolder, J. van Braak, J. Tondeur, Supporting self-regulated learning in computer-based learning environments: systematic review of effects of scaffolding in the domain of science education: Scaffolding self-regulated learning with CBLES, Journal of Computer Assisted Learning 28 (2012) 557–573. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.1365-2729.2011.00476.x. doi:10.1111/j.1365-2729.2011.00476.x.

[20] J. Wong, M. Baars, D. Davis, T. Van Der Zee, G.-J. Houben, F. Paas, Supporting Self-Regulated Learning in Online Learning Environments and MOOCs: A Systematic Review, International Journal of Human–Computer Interaction 35 (2019) 356–373. URL: https://www.tandfonline.com/doi/full/10.1080/10447318.2018.1543084. doi:10.1080/10447318.2018.1543084.

[21] R. L. Jørnø, K. Gynther, What Constitutes an 'Actionable Insight' in Learning Analytics?, Journal of Learning Analytics 5 (2018). URL: https://learning-analytics.info/index.php/JLA/article/view/5897. doi:10.18608/jla.2018.53.13.

[22] W.-J. Shyr, C.-H. Chen, Designing a technology-enhanced flipped learning system to facilitate students' self-regulation and performance, Journal of Computer Assisted Learning 34 (2018) 53–62. URL: https://onlinelibrary.wiley.com/doi/10.1111/jcal.12213. doi:10.1111/jcal.12213.

[23] C. Michel, E. Lavoué, S. George, M. Ji, Supporting Awareness and Self-Regulation In Project-Based Learning through Personalized Dashboards, International Journal of Technology Enhanced Learning 9 (2017) 204–226. URL: https://hal.archives-ouvertes.fr/hal-01384211. doi:10.1504/IJâĎą.2017.084500.

[24] P. Reimann, Design-Based Research, in: L. Markauskaite, P. Freebody, J. Irwin (Eds.), Methodological Choice and Design: Scholarship, Policy and Practice in Social and Educational Research, Methodos Series, Springer Netherlands, Dordrecht, 2011, pp. 37–50. URL: https://doi.org/10.1007/978-90-481-8933-5_3. doi:10.1007/978-90-481-8933-5_3.

[25] M. Pérez-Sanagustín, R. Pérez-Álvarez, J. Maldonado-Mahauad, E. Villalobos, C. Sanza, Designing a moodle plugin for promoting learners' self-regulated learning in blended learning, in: Proceedins of the Seventeenth European Conference on Technology-Enhanced Learning - EC-TEL '22, Toulouse, France, In press.

[26] E. Villalobos, M. Pérez-Sanagustín, C. Sanza, A. Tricot, J. Broisin, Supporting self-regulated learning in bl: Exploring learners' tactics and strategies, in: Proceedins of the Seventeenth European Conference on Technology-Enhanced Learning - EC-TEL '22, Toulouse, France, In press.

[27] P. R. Pintrich, E. V. D. Groot, Motivational and Self-Regulated Learning Components of Classroom Academic Performance (1990) 8.

[28] J. Jovanovic, N. Mirriahi, D. Gašević, S. Dawson, A. Pardo, Predictive power of regularity of pre-class activities in a flipped classroom, Computers & Education 134 (2019) 156–168. URL: https://www.sciencedirect.com/science/article/pii/S0360131519300405. doi:10.1016/j.compedu.2019.02.011.
[29] M. Saqr, J. Jovanovic, O. Viberg, D. Gašević, Is there order in the mess? A single paper meta-analysis approach to identification of predictors of success in learning analytics, Studies in Higher Education (2022) 1–22. URL: https://www.tandfonline.com/doi/full/10.1080/03075079.2022.2061450. doi:10.1080/03075079.2022.2061450.

Ethical FRAPPE – an adapted draft framework for ethical AIED

Bhoomika Agarwal, Corrie Urlings, Giel van Lankveld and Roland Klemke

Open Universiteit, Valkenburgerweg 177, 6419 AT Heerlen, The Netherlands

Abstract

Artificial Intelligence (AI) is pervading our lives in numerous ways today. It is important to apply ethical principles to guide the development and usage of AI systems to prevent harms or discrimination through AI algorithms. This has led to various ethical regulations and guidelines being formed at the corporate, national and supra-national level. The EU AI Act classifies the usage of AI in education as 'high-risk' as "such systems may violate the right to education and training as well as the right not to be discriminated against and perpetuate historical patterns of discrimination" [1, p. 26]. However, there has been little attention paid to ethics in AI in Education (AIED) in literature and there is only one existing framework to ethically guide AIED. AIED ethics is complex as it has to combine both general AI ethics and the ethics of educational technology. We aim to create a theoretical framework for AIED, comprising implementation guidelines for developers and organizational users of AI in education. In this paper, an existing draft framework by Holmes et al. is adapted by using insights from literature in the ethics of AI, ethics of educational technology and ethics of AIED.

Keywords

Artificial Intelligence, Education, Ethics

1. Introduction

Artificial Intelligence (AI), once a buzzword, is now a reality. It is being used in many aspects of our lives including healthcare, transport, communication, agriculture, finance and education. The usage of AI in classrooms and in education is promising and provides opportunities to improve the education process with technological innovations. AI has been applied in educational contexts in automation of administrative processes and tasks, curriculum and content development, instruction, and students' learning processes [3]. AI systems have enabled early detection and redress of learning shortcoming by analyzing student data - thereby providing a more customized learning experience for students [3]. Over the past decade, the use of AI tools to support or enhance learning has grown exponentially [4]. In a recent literature review, Chen et al. looked at 20 years of AI in Education (AIED) from 2000 to 2019 and shared several relevant findings: (a) the domain of AIED has received increased interest in the last few years, owing to the positive effect of AI on learning performance; (b) there is an increase in AIED literature over the years, showing an increased scientific output; (c) AIED research is especially found in interdisciplinary journals with a dual focus on education and technology [5].

Ethics plays an important role in guiding the usage of AI in our lives. As defined by Potter Stewart, "Ethics is knowing the difference between what you have a right to do and what is right to do" [6]. It is important to ethically guide the development and usage of AI for several reasons. The primary reason is that AI is being increasingly integrated into our lives and therefore has the potential for widespread influence and direct control over people's lives. This means that it could negatively or unfairly impact numerous lives with far-reaching consequences. AI technologies are being developed at a high speed to automate tasks that are traditionally done by humans. The parties implementing the automation of tasks are at risk of not fully considering the ethical consequences in an effort to improve efficiency and save costs [7]. When such automated tasks involve any sort of decision-making by AI, the decisions can impact the personal well-being of individuals and have a potential for dangerous consequences.

The EU AI Act [1] classifies the usage of AI in education as 'high-risk' as "such systems may violate the right to education and training as well as the right not to be discriminated against and perpetuate historical patterns of discrimination" [1, p. 26]. In addition, ethics for AIED have not been discussed at the forefront of national AI policy strategies [8]. Schiff examined 24 national AI policy strategies from G-7 and OECD countries and other important global actors such as India, China, Russia, Singapore and Malta [8]. The author found that remarkable attention has been paid to AI ethics in general, but this did not imply that attention has been paid to ethics in AIED in particular. Schiff also noted that the missing role

Proceedings of the Doctoral Consortium of the Seventeenth European Conference on Technology Enhanced Learning, September 12–16, 2022, Toulouse, France

bhoomika.agarwal@ou.nl (B. Agarwal); corrie.urlings@ou.nl (C. Urlings); giel.vanlankveld@ou.nl (G. v. Lankveld); roland.klemke@ou.nl (R. Klemke)

0000-0002-7347-8465 (B. Agarwal)

of education as a sector is an anomaly because many of these national AI policy documents "discuss the use of AI not only for healthcare, but also for transportation, agriculture, finance, and many other sectors" [8]. In addition, of the 4-5 countries that discussed AIED as a tool for teaching, learning and educational administration, none of them commented on or discussed AIED ethics. This is a cause of concern as there is no consideration about the ethical approach to AIED among policymakers [8]. Until now, there exists only one framework for ethical AIED developed by 'The Institute for Ethical AI in Education' aimed at those making procurement and application decisions regarding AIED [9].

AIED ethics is complicated as it has to consider both general AI ethics and the ethics of educational technology. On the one hand, there is an overlap between the ethics of AI, ethics of educational technology and ethics of AIED - suggesting that they should draw inspiration from each other [2]. On the other hand, the usage of AIED systems raises concerns such as the autonomy of teachers, responsibility and accountability for decisions made by AIED systems, impact of potential discrimination by AIED systems through historical biases, explainability of AIED systems, etc [2]. Owing to these concerns raised by AIED systems, AIED ethics deserves attention and there is a need to develop an ethical framework for guiding AIED ethics that is targeted at developers and organizational users of AIED.

Keeping the limited attention to AIED ethics in mind, we aim to create an ethical framework for AIED using the Ethical FRAPPE - a set of high-level ethical principles for AIED that are derived in this paper. This paper aims to answer the following research question: "Can Ethical FRAPPE be used to construct an exhaustive ethical framework for AIED?" Multiple steps are necessary to answer this research question: (1) Define the properties and aspects of an exhaustive ethical framework from literature; (2) Identify the ethical principles that can be used to form an exhaustive ethical framework for AIED; (3) Identify current and possible future use-case scenarios that an ethical framework for AIED can be applied to, such that the framework can be future-proof and evolve as AI evolves. However, several of these steps are out of scope for this paper. In this paper, we focus on the second step. As part of the second step, we build upon an existing draft framework for ethical AIED by Holmes et al. using insights from literature. The other two steps are planned as part of the future work, as described in section 5.

2. Background

A framework for AIED should aim to combine both the ethics of AI and the ethics of educational technology into a single framework, considering the overlap between these two domains. Thereby, 2.1 looks into the ethics of AI and 2.2 looks into the ethics of educational technology individually. 2.3 examines the overlap between the above two domains and looks at existing AIED ethics frameworks.

2.1. Ethics of AI

The ethics of AI in general have been studied extensively and numerous frameworks and policies have been developed for AI ethics. The inventory of AI Ethics guidelines by the Algorithm Watch [10] comprises 167 different guidelines on a corporate, national and supra-national level. Among these, some frameworks are notable. The Asilomar AI principles developed by the Future of Life Institute [11] has been adopted by 1797 AI and Robotics researchers and 3923 others. Furthermore, the 'Ethics Guidelines for trustworthy AI' have been proposed by the European Union [12]. The guidelines are encompassed in the 'AI Act', which is a proposed European law to regulate the usage of AI [1].

Floridi et al. encouraged an ethical approach to AI to incorporate the benefits of AI and mitigate the potential harms caused by AI in a balanced way. The authors proposed AI4People – a framework formed by the synthesis of existing sets of principles produced by various reputable, multi-stakeholder organisations and initiatives [13]. Their framework comprised of five principles- beneficence, non-maleficence, autonomy, justice and explicability [13]. These 5 principles have a major overlap with the principles found by Jobin et al. in their scoping review of AI ethics guidelines comprising 84 documents [14].

While there is a growing body of AI ethics guidelines and frameworks that can be found in literature [14, 10], these initiatives have primarily produced high-level ethical principles, tenets, values and abstract requirements for AI development and deployment [15]. This principle-based approach towards AI is criticised due to its inability to deal with the complexity of issues raised by AI [15, 16]. More specifically, the high-level ethical principles do not translate into practice automatically with the tools presently available to developers [17]. With the high number of abstract guidelines proposed, 'ethics washing' is on the rise by technology companies [18]. 'Ethics washing' occurs when technical companies define ethical policies to maintain outward appearances without following the principles in practice [18]. A second reason for criticism stems from the principle-based approached being aimed at a range of stakeholders and are thereby often difficult to understand for specific groups of users [16].

Although the principle-based approach is criticized to be ineffective due to issues such as ethics washing, it forms a good first step towards defining an ethical framework. Thereby, we begin by defining the high-level ethical principles in this paper. As part of future work, we adopt a similar approach as [17], in which we plan to define requirements from ethical principles for AIED and map them to design-based research (DBR) process instead, as elaborated in section 5. Armstrong et al. define DBR in an educational setting as "a research approach that engages in iterative designs to develop knowledge that improves educational practices" [19]. As DBR brings educational research closer to everyday practice, this methodology is increasingly being used in designing educational research [20].

2.2. Ethics of educational technology

As AIED ethics needs to consider the ethics of educational technology, ethical policies for educational technology are reviewed here.

Pardo and Siemens identified four principles to categorize the issues derived from privacy in educational data: transparency, student control over the data, security, and accountability and assessment [21]. As Learning Analytics (LA) is a sub-field of AIED that uses educational data to optimize learning, the ethics of AIED should consider the ethics of LA. LA is defined in the proceedings of the 1st International Conference on Learning Analytics and Knowledge 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" [22]. Sclater developed a code of practice for LA that advises educational institutions on how to use LA ethically. The authors considered eight focus areas - ownership and control, consent, transparency, privacy, validity, access, action, adverse impact, stewardship [23]. In a recent literature review on the ethics of LA in higher education, Pargman and McGrath found that the top three ethical areas most in LA articles are transparency, privacy, and informed consent [24]. In the context of Dutch higher education, Engelfriet et al. developed a guide to LA that focuses on the protection of personal student data. Drachsler and Greller developed an eight point checklist named DELICATE that can serve as a reflection aid for ethical and privacy-supported LA. The DELICATE checklist comprises 8 checkpoints- "Determination, Explain, Legitimate, Involve, Consent, Anonymise, Technical and External" as a quality checklist to make stakeholders aware and guide them through the process.

The ethics of educational technology contains issues that are relevant to the domain of education. Issues relating to student autonomy and control over their data can have long-term effects on the future of students. There needs to be regulations regarding informed consent and privacy of students, interpretation and management of student data. There is a clear overlap between ethics of educational technology and ethics of AIED - suggesting that ethics of AIED should draw inspiration from the ethics of educational technology and should build on top of frameworks for ethics of educational technology.

2.3. Ethics of AIED

This section looks at existing frameworks and guidelines for AIED ethics.

The conversation revolving around ethics for AIED was started over 20 years ago by Aiken and Epstein with an aim to raise awareness of researchers while designing educational systems [27]. The authors set down 10 principles for AIED systems based on "The Golden Rule for Computers in Education: Teach others as you would like to be taught" [27].

The first ethical framework for AIED was developed by The Institute for Ethical AI in Education that involves designers and developers for AIED and sets down guidelines for them [9]. However, this framework is aimed at the decision makers during the process of procurement and the application of AIED. This framework focuses on defining high-level ethical principles without any implementation guidelines that are relatable to developers during the design of AIED systems. It contains the downsides of the principle-based approach to AI ethics in the form of a lack of translation into practice for developers.

Holmes et al. conducted a survey with 17 domain experts comprising 10 open questions to gauge expert opinion about ethics of AIED [2]. They examined the various aspects of ethics of AIED and concluded that "the ethics of AIED cannot be reduced to questions about data or computational approaches alone" [2] and needs to account for the ethics of education – including, but not limited to – the purpose of learning, choice of pedagogy, role of technology with respect to teachers and access to education [2]. The authors created a 'strawman draft' framework, shown in Figure 1, that identified three areas of focus: "the ethics of data, computational approaches and education" and emphasized the overlaps between these foci. The authors identified 3 levels of overlap in their 'strawman draft' framework . The first level comprised of three foci: "the ethics of data, computational approaches and education" while the second level comprised of the overlap between each pair of foci. These 2 layers form the 'known unknowns' while the overlap between these 3 foci formed the 'unknown unknowns' [2].

3. Methods

This paper aims to answer the following research question: "Can Ethical FRAPPE be used to construct an ex-

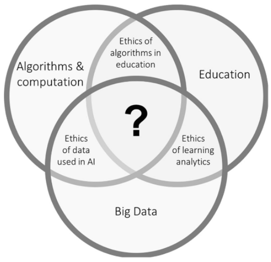

Figure 1: The 'strawman' draft framework for the ethics of AIED developed by [2]

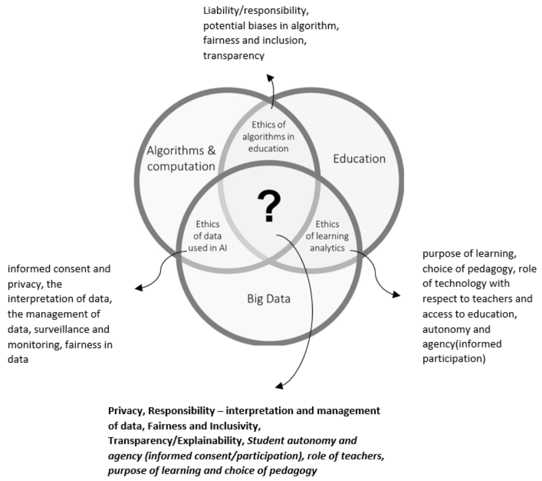

Figure 2: Revised version for the 'strawman' draft framework for the ethics of AIED adopted from [2]

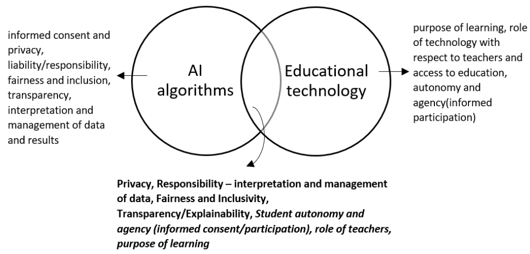

Figure 3: Adapted version for the 'strawman' draft framework for the ethics of AIED adopted from [2]

haustive ethical framework for AIED?"

In order to answer this research question, the draft framework by Holmes et al. was selected as a foundational framework. This is because this 'strawman draft' framework is well-informed by experts in the domain of AIED and considers a template model for the essential aspects of ethical AIED. However, it only forms a skeleton model and does not contain the ethical principles involved in these domains. After making a few modifications, we fill in this gap in the 'strawman draft' framework by Holmes et al. by examining existing literature in the domains of both AI ethics, ethics of educational technology and ethics of AIED. High-level ethical principles are identified from literature and incorporated into this framework.

We proposed two modifications to the 'strawman draft' framework by Holmes et al.. Firstly, we elaborated on and defined the aspects in the intersection of these foci with an aim to throw light upon the 'known unknowns' and the 'unknown unknowns' stated by Holmes et al.. The 'strawman draft' framework defines the domains involved in ethics of AIED but does not elaborate on the ethical aspects of these domains. Thereby, we identified the ethical aspects involved in each of these foci based on literature, as shown in Figure 2.

Secondly, a huge overlap was noticed in the ethical aspects mapped to the foci of 'ethics of data' and 'ethics of computational approaches', as can be seen in Figure 2. Data and computational approaches were seen to be tightly coupled as any changes in one of them leads to changes in the other. For example, bias in data can lead to bias in the computational algorithm. Similarly, the interpretation and management of the data can have a direct effect on the privacy of the computational approach in the form of exposing sensitive attributes. Due to this tight coupling between the ethics of data and the ethics of computational approaches, they cannot be separated into 2 separate foci. Hence, we decided to combine them into a single focus. The revised and adapted version of our framework draft can be seen in Figure 3. It contains 2 focal areas: ethics of AI algorithms and ethics of educational technology, each containing corresponding ethical principles. The intersection of these 2 foci contains the ethical principles that form our theoretical framework.

Following these modifications, literature in the domains of AI ethics, ethics of educational technology and ethics of AIED were reviewed. This was then used to obtain the ethical principles relevant to an ethical framework for AIED, abbreviated as the Ethical FRAPPE. The list of articles reviewed is grouped using the adapted draft framework as shown in Figure 3 into the ethics of AI, ethics of educational technology and ethics of AIED. Table 1 contains the list of selected articles that were reviewed to form the Ethical FRAPPE in order of year of publication. Following the adapted draft framework, these articles are grouped into the domains of ethics of AI, ethics of educational technology (EdTech) and ethics of AIED.

The high-level ethical principles seen in the literature

Table 1
List of selected articles

Year	Author(s)	Domain
2000	Aiken and Epstein [27]	Ethics of AIED
2014	Pardo and Siemens [21]	Ethics of EdTech
2016	Drachsler and Greller [26]	Ethics of EdTech
2016	Sclater [23]	Ethics of EdTech
2017	Engelfriet et al. [25]	Ethics of EdTech
2017	Prinsloo and Slade [29]	Ethics of EdTech
2017	Boddington [30]	Ethics of AI
2018	Floridi et al. [13]	Ethics of AI
2018	Whittaker et al. [7]	Ethics of AI
2019	Mittelstadt [15]	Ethics of AI
2019	Dignum [31]	Ethics of AI
2019	Jobin et al. [14]	Ethics of AI
2019	Crawford et al. [32]	Ethics of AI
2019	Kitto and Knight [33]	Ethics of EdTech
2019	Commission et al. [12]	Ethics of AI
2020	Morley et al. [17]	Ethics of AI
2020	Hagendorff [34]	Ethics of AI
2020	AlgorithmWatch [10]	Ethics of AI
2020	Vincent-Lancrin and van der Vlies [35]	Ethics of AIED
2021	Ryan and Stahl [16]	Ethics of AI
2021	Li et al. [36]	Ethics of AI
2021	Commission et al. [1]	Ethics of AI
2021	Miao et al. [4]	Ethics of AIED
2021	The Institute for Ethical AI in Education [9]	Ethics of AIED
2021	Holmes et al. [2]	Ethics of AIED
2021	Schiff [8]	Ethics of AIED
2021	Baker and Hawn [37]	Ethics of EdTech
2021	Pargman and McGrath [24]	Ethics of EdTech

were compared to each other. The ethical principles seen in a majority of the articles in each domain are identified and consolidated to create the Ethical FRAPPE. The 6 ethical principles identified as part of the Ethical FRAPPE are:

1. Fairness
2. Responsibility
3. Autonomy
4. Privacy
5. Purpose of learning
6. Explainability

Despite the presence of a large body of ethical guidelines, these guidelines rely on context-specific keywords and there exist multiple definitions of the ethical principles and technical terms involved [28]. This makes it challenging to interpret and operationalize these ethical values [28]. Keeping in mind the need for a common vocabulary to avoid misinterpretation of the ethical principles [28], we define the ethical aspects and explain them in the light of AIED ethics as below.

3.1. Fairness

Fairness, or Freedom from Bias is defined as - ”Systematic unfairness perpetrated on individuals or groups, including pre-existing social bias, technical bias, and emergent social bias” [38]. AIED systems should not be designed such that the algorithms develop historically unfair prejudices by ensuring fair data that is inclusive, representative of the target population and without inaccuracies [16]. Any conscious or unconscious biases that are incorporated into AI algorithms through the data analysis can have a negative impact on the rights of individual students [4]. AIED should strive towards equitable access to AI technologies for all, keeping in line with SDG 4 set down by the UNESCO - “ensure inclusive and equitable quality education and promote lifelong learning opportunities for all” [4].

3.2. Responsibility

“Responsible AI is concerned with the fact that decisions and actions taken by intelligent autonomous systems have consequences that can be seen as being of an ethical nature” [31]. In [31], the author states that Responsible

AI should follow 3 ethical principles:

1. Accountability: refers to the ability of the AI system to explain and justify its decisions
2. Responsibility: refers to the role of people with regards to the AI system
3. Transparency: refers to the capability of AI systems to "describe, inspect and reproduce the mechanisms through which AI systems make decisions" [31]

In the light of AIED, responsibility is required to ensure accountability of decisions, responsibility of the developers and maintainers of AI towards its users and transparency of data and purpose of the system.

3.3. Autonomy

Human autonomy is defined as "Refers to people's ability to decide, plan, and act in ways that they believe will help them to achieve their goals" [38]. Autonomy, also called 'agency' in some ethical guidelines, ensures that the users of the systems are informed actors and have full control over their own decisions when they interact with the AIED system [16]. In AIED systems, student autonomy is important to ensure that students understand the purpose of the system and have complete control over their personal data, including the right to opt out of such systems without negative consequences. Students should be informed about the data being collected about them and should be involved in any decisions made using such data. Teacher autonomy is equally important to ensure that the role of teachers is highlighted in the form of the human-in-the-loop in the AIED system. By allowing teachers to review and act upon the decisions made by autonomous AIED systems, teacher autonomy can be ensured and unfair decisions by the AIED system can be reduced. It is also important that the data collected about the teachers should not have an adverse effect on their role in the classroom. This can be ensured by ensuring both teacher and student autonomy in AIED systems.

3.4. Privacy

While there are multiple definitions of the term Privacy, we choose the consolidated definition from [38] - "a claim, an entitlement, or a right of an individual to determine what information about himself or herself can be communicated to others" [38]. In the context of AIED systems, privacy pertains to the sharing of private and confidential data with others. A huge amount of digital data is stored about students through their online activity and there is a need to regulate the access and ownership of this data so that it is only accessible to the concerned parties and is not repurposed for other uses. There is a fear of educational institutions and employers using 'old' data and the usage of student data for commercial purposes [35].

3.5. Purpose of learning

At the moment, AIED systems are being used as an application use-case of AI instead of being motivated by learning goals. AIED is partly taking the form of data scientists looking for a context where predictive modelling and other AI techniques can be applied [39]. There is a need to criticially examine the purpose of learning and the performance measures that this purpose of learning is being reduced to. It is important to keep in mind that "theories of learning cannot, after all, be 'discovered' by algorithms" [39]

3.6. Explainability

Explainibility is "understanding how an AI model makes its decision" [40]. AIED systems should be actively monitored to ensure accurate and reproducible results that can be explained with the data and algorithmic functionality [16]. AIED models should be built to be explainable by design (using partially or fully explainable models) or post-hoc explainability methods should be used in the case of black-box models that are not inherently explainable [40]. In the case of AIED models, it is necessary to ensure that the decisions taken by the algorithm are explainable to humans in order to avoid negative harms to students.

4. Conclusion

This paper aimed to answer the research question: "Can Ethical FRAPPE be used to construct an exhaustive ethical framework for AIED?" In order to answer this question, this paper aims to identify the high-level ethical principles that can be used to construct an exhaustive ethical framework for AIED. The existing 'strawman draft' framework for ethical AIED by Holmes et al. was adapted by adding high-level ethical principles that were identified from existing literature in the domains of ethics of AI, ethics of educational technology and ethics of AIED. The six high-level ethical principles identified and consolidated from literature are abbreviated in the form of the Ethical FRAPPE for AIED: Fairness, Responsibility, Autonomy, Privacy, Purpose of learning and Explainability. The 6 ethical principles in the Ethical FRAPPE were defined in the context of AIED systems to form the first outline of our theoretical framework for AIED.

5. Future Work

The construction of our framework and implementation guidelines will be conducted in 3 phases: 'theoretical framework', 'evaluation framework' and 'instantiation'.

In the first phase, a theoretical ethical framework will be developed for AIED. In order to define an exhaustive ethical framework for AIED, it is first essential to look at what comprises a good ethical framework. To answer this, a literature review will be conducted. This paper describes the first part of the first phase where an existing draft model for AIED ethics was adapted by identifying high-level ethical principles from literature. In the future work, these high-level ethical principles will be converted to requirements and then be used to create the theoretical framework in the form of a checklist that contains practical guidelines for developers of AIED. The expected theoretical framework will be a checklist comprising definitions, requirements, formula and guidelines for ethical principles. This first draft of the framework will be evaluated by experts in the domain of AIED for face validity and content validity.

In the second phase, a methodology will be developed to quantify the ethics of AIED applications based on the theoretical framework. We refer to this methodology as the 'evaluation framework'. This evaluation framework will provide quantification tools for the ethical principles integrated in the form of a pipeline that can check existing AI systems for ethical soundness and provide recommendations for improvement. First, a subset of the ethical principles from the theoretical framework will be identified as 'focus' principles based on their prominence and relevance. Following this, various tools will be examined to identify suitable quantification tools for the focus ethical principles. Lastly, there will be an evaluation of different technologies for the architecture, followed by design and implementation of the evaluation pipeline. This evaluation pipeline will receive the trained AI algorithm, input data and output data as inputs and will give an ethical score as an output. This ethical score will be calculated as the sum of individual scores for each ethical principle. The individual score for each ethical principle will be based on the implementation of the guidelines from the theoretical framework and will also contain recommendations for improvements. If the ethical score for a majority of the ethical principles (exact threshold to be decided based on the number of ethical principles) is above 80%, the ethical evaluation will be passed. Such an ethical score allows for some trade-offs between principles in the event of conflicts between them, while ensuring that the system is ethical as a whole.

In the third phase, called 'instantiation', a proof of concept or instantiation of the evaluation framework will be developed. For this purpose, an AIED application will be developed which enables the identification of struggling students in a university, online, distance education setting. The main goal of this sample use case would be to improve teaching and learning processes on the whole and support teachers. The theoretical framework will be used to guide the design of this use case and the evaluation framework will be integrated into this AIED application to evaluate the ethics of this application. Additionally, the evaluation framework will be applied to some selected AIED models for evaluation such that they can cover various use cases. Finally, recommendations and guidelines will be provided for application of the theoretical and evaluation frameworks into other AIED applications. These recommendations will be developed for common challenges (such as biases or issues) seen in different classes/applications of reviewed AI algorithms from the literature. The applications of AI seen from literature will be grouped based on parameters such as the class of algorithms, coding language used and data type used.

References

[1] E. Commission, C. Directorate-General for Communications Networks, Technology, Proposal for a regulation of the european parliament and the council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts (2021). URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206.

[2] W. Holmes, K. Porayska-Pomsta, K. Holstein, E. Sutherland, T. Baker, S. B. Shum, O. C. Santos, M. T. Rodrigo, M. Cukurova, I. I. Bittencourt, K. R. Koedinger, Ethics of AI in Education: Towards a Community-Wide Framework, International Journal of Artificial Intelligence in Education (2021). doi:10.1007/S40593-021-00239-1.

[3] L. Chen, P. Chen, Z. Lin, Artificial intelligence in education: A review, IEEE Access 8 (2020) 75264–75278. doi:10.1109/ACCESS.2020.2988510.

[4] F. Miao, W. Holmes, R. Huang, H. Zhang, et al., AI and education: A guidance for policymakers, UNESCO Publishing, 2021.

[5] X. Chen, D. Zou, H. Xie, G. Cheng, C. Liu, Two decades of artificial intelligence in education, Educational Technology & Society 25 (2022) 28–47.

[6] P. Stewart, Ethics and problem solving 11 (2021).

[7] M. Whittaker, K. Crawford, R. Dobbe, G. Fried, E. Kaziunas, V. Mathur, S. M. West, R. Richardson, J. Schultz, O. Schwartz, AI now report 2018, AI Now Institute at New York University New York, 2018.

[8] D. Schiff, Education for AI, not AI for Education: The Role of Education and Ethics in National AI Policy Strategies, International Journal of Artificial

Intelligence in Education (2021) 1–37. doi:10.1007/S40593-021-00270-2/.
[9] The Institute for Ethical AI in Education, The Ethical Framework for AI in Education, Technical Report, The Institute for Ethical AI in Education, 2021. URL: https://www.buckingham.ac.uk/research-the-institute-for-ethical-ai-in-education/.
[10] AlgorithmWatch, AI Ethics Guidelines Global Inventory - AlgorithmWatch, 2020. URL: https://algorithmwatch.org/en/ai-ethics-guidelines-global-inventory/.
[11] Future of Life Institute, Asilomar AI Principles, 2017. URL: https://futureoflife.org/2017/08/11/ai-principles/.
[12] E. Commission, C. Directorate-General for Communications Networks, Technology, Ethics guidelines for trustworthy AI, Publications Office, 2019. doi:doi/10.2759/346720.
[13] L. Floridi, J. Cowls, M. Beltrametti, R. Chatila, P. Chazerand, V. Dignum, C. Luetge, R. Madelin, U. Pagallo, F. Rossi, B. Schafer, P. Valcke, E. Vayena, AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations, Minds and Machines 28 (2018) 689–707. doi:10.1007/S11023-018-9482-5/.
[14] A. Jobin, M. Ienca, E. Vayena, Artificial Intelligence: the global landscape of ethics guidelines, Nature Machine Intelligence 1 (2019) 389–399. doi:10.1038/s42256-019-0088-2.
[15] B. Mittelstadt, Ai ethics–too principled to fail, arXiv preprint arXiv:1906.06668 (2019).
[16] M. Ryan, B. C. Stahl, Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications, Journal of Information, Communication and Ethics in Society 19 (2021) 1477–996. URL: http://creativecommons.org/licences/by/4.0/legalcode. doi:10.1108/JICES-12-2019-0138.
[17] J. Morley, L. Floridi, L. Kinsey, A. Elhalal, From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices, Science and Engineering Ethics 26 (2020) 2141–2168. doi:10.1007/S11948-019-00165-5/.
[18] E. Bietti, From ethics washing to ethics bashing: A view on tech ethics from within moral philosophy, in: Proceedings of the 2020 conference on fairness, accountability, and transparency, 2020, pp. 210–219.
[19] M. Armstrong, C. Dopp, J. Welsh, Design-Based Research - The Students' Guide to Learning Design and Research, EdTech Books, 2018, pp. 40–45. URL: https://edtechbooks.org/studentguide/design-based_research.
[20] D.-B. R. Collective, Design-based research: An emerging paradigm for educational inquiry, Educational researcher 32 (2003) 5–8.
[21] A. Pardo, G. Siemens, Ethical and privacy principles for learning analytics, British Journal of Educational Technology 45 (2014) 438–450. doi:10.1111/BJET.12152.
[22] P. Long, G. Siemens, G. Conole, D. Gašević, LAK '11: Proceedings of the 1st International Conference on Learning Analytics and Knowledge, Association for Computing Machinery, New York, NY, USA, 2011.
[23] N. Sclater, Developing a code of practice for learning analytics, Journal of Learning Analytics 3 (2016) 16–42. URL: https://learning-analytics.info/index.php/JLA/article/view/4512. doi:10.18608/jla.2016.31.3.
[24] T. C. Pargman, C. McGrath, Mapping the terrain of ethics in learning analytics: A systematic literature review of empirical research, Journal of Learning Analytics (2021).
[25] A. Engelfriet, J. Manderveld, E. Jeunink, Learning analytics under the (dutch) personal data protection act. policy paper. surfnet, 2017. URL: https://www.surf.nl/en/guide-to-learning-analytics-under-the-personal-data-protection-act.
[26] H. Drachsler, W. Greller, Privacy and Analytics-it's a DELICATE Issue A Checklist for Trusted Learning Analytics, Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (2016) 89–98. doi:10.1145/2883851.
[27] R. M. Aiken, R. G. Epstein, Ethical Guidelines for AI in Education: Starting a Conversation, International Journal of Artificial Intelligence in Education 11 (2000) 163–176. URL: http://www.cs.wcupa.edu/~epstein/AIStories.html.
[28] A. Theodorou, V. Dignum, Towards ethical and socio-legal governance in ai, Nature Machine Intelligence 2 (2020) 10–12.
[29] P. Prinsloo, S. Slade, Ethics and Learning Analytics: Charting the (Un)Charted, Handbook of Learning Analytics (2017) 49–57. URL: http://oro.open.ac.uk/49456/. doi:10.18608/HLA17.004.
[30] P. Boddington, Towards a code of ethics for artificial intelligence, Springer, 2017.
[31] V. Dignum, Responsible artificial intelligence: how to develop and use AI in a responsible way, Springer Nature, 2019.
[32] K. Crawford, R. Dobbe, T. Dryer, G. Fried, B. Green, E. Kaziunas, A. Kak, V. Mathur, E. McElroy, A. N. Sánchez, et al., Ai now 2019 report, New York, NY: AI Now Institute (2019).
[33] K. Kitto, S. Knight, Practical ethics for building learning analytics, British Journal of Educational Technology 50 (2019) 2855–2870.
[34] T. Hagendorff, The Ethics of AI Ethics: An Evaluation of Guidelines, Minds and Machines 30 (2020) 99–120. doi:10.1007/S11023-020-09517-8/.

[35] S. Vincent-Lancrin, R. van der Vlies, Trustworthy artificial intelligence (ai) in education: Promises and challenges (2020).
[36] B. Li, P. Qi, B. Liu, S. Di, J. Liu, J. Pei, J. Yi, B. Zhou, Trustworthy ai: From principles to practices, arXiv preprint arXiv:2110.01167 (2021).
[37] R. S. Baker, A. Hawn, Algorithmic bias in education, International Journal of Artificial Intelligence in Education (2021) 1–41.
[38] B. Friedman, P. H. Kahn, A. Borning, A. Huldtgren, Value Sensitive Design and Information Systems, Springer Netherlands, Dordrecht, 2013, pp. 55–95. URL: https://doi.org/10.1007/978-94-007-7844-3_4. doi:10.1007/978-94-007-7844-3_4.
[39] C. Perrotta, N. Selwyn, Deep learning goes to school: toward a relational understanding of ai in education, Learning, Media and Technology 45 (2020) 251–269. doi:10.1080/17439884.2020.1686017.
[40] B. Li, P. Qi, B. Liu, S. Di, J. Liu, J. Pei, J. Yi, B. Zhou, Trustworthy ai: From principles to practices, arXiv preprint arXiv:2110.01167 (2021). URL: https://doi.org/10.48550/arXiv.2110.01167.

Trade-off model for supporting educators' digital competence assessment

Linda Helene Sillat, Kairit Tammets, and Mart Laanpere
Tallinn University, Narva Road 25, Tallinn, 10120, Estonia

Abstract
The majority of the efforts in assessing educators' digital competence over the past decade have been focused on developing evidence-based and scientifically reliable assessment instruments. These instruments are often created ad-hoc by research groups without deeper understanding of the educators' needs and expected benefits for digital competence assessment. That implies that although the instrument might give valid and reliable results for the researchers it disregards all other related stakeholders – educators, school leaders, educational technologist, teacher trainers etc. To understand and guide evidence-informed decision-making when developing, adapting or implementing digital competence assessment instruments it is important to accommodate all stakeholders to provide meaningful assessment results and data. To provide a solution for this problem we have designed a trade-off model which focuses on mapping the digital competence assessment instruments to stakeholder needs and expected benefits. Our research is divided into three main phases. First, we focused on understanding the concept and domain of educators' digital competence. For which we analysed the existing educators' digital competence frameworks, models and similar previous mappings from the literature. Secondly, to explain the alternative digital competence assessment approaches and instruments we mapped the underlying assessment processes and piloted alternative instrument with different educator groups. The third and final phase focused on designing, developing and validating the trade-off model. The following describes all three phases and provides an overview of the initial findings which are accompanied with suggestions for further research in the field of educators' digital competence assessment.

Keywords 1
Digital competence; assessment, instruments, educators, trade-off model.

1. Introduction

Using technologies in teaching and learning is not considered a novel practice any more but rather presented as a norm for quality education. Innovative and pedagogically reasonable ways to implement technologies on the other hand has presented difficult among teachers and thus the discussion on educators' digital competence has gained popularity. However, it is evident that not only mapping the needed digital competence of educators is needed but more importantly we need to understand the level of digital competence of educators to support meaningful professional development. Digital competence is considered as a goal oriented, confident and critical use of technologies for work, employability, learning, leisure and inclusive participation in society [1].

Educational assessment has been a central discussion for overall quality assurance in educational settings or trying to understand knowledge development [2]. Harlen & James [3] have stated that there are three general

Proceedings of the Doctoral Consortium of the Seventeenth European Conference on Technology Enhanced Learning, September 12-16, 2022, Toulouse, France
EMAIL: linda.sillat@tlu.ee; kairit.tammets@tlu.ee; mart.laanpere@tlu.ee
ORCID: 0000-0001-9012-6165; 0000-0003-2065-6552; 0000-0002-9853-9965

assessment approaches which also related to digital competence assessment - formative, summative and diagnostic assessment. Within these assessment approaches there is a variety of instruments, most notably self-assessment, knowledge-based tests and authentic assessment instruments like e-portfolios of reflective journals. It can be argued that for the past decade the efforts have mainly been towards developing self-assessment instruments which are cost-effective, mostly adaptable and cover variety of educators' groups (i.e. primary to higher and vocational education). However, research done piloting and implementing these self-assessment instruments proposes a question whether educators assess their digital competence or something else entirely. Benali et al. [4] propose that majority of educators often assess their self-confidence in integrating technologies to their pedagogical practice and fail to give suitable evidence of their current practices. It is also considered that many digital competence assessment instruments which are based on self-assessment do not cover digital competence but rather focus on low-order cognitive skills [5], [6].

Previous research has also revealed that knowledge-based testing and authentic assessment requires higher volume of resources, both financial and human capital and is difficult to monitor [7].

Regardless the form of assessment and type of used instruments it is concluded that there is a sustainability issue which implies that there is a contradiction between the number of digital competence frameworks and models and the number of corresponding instruments.

Another dimension in educators' digital competence assessment is the understanding of the related stakeholder groups who either require access to the assessment results or data. Adhering to these stakeholder group needs and expectations has proven to be a difficult task [8]. On one hand we lack a clear understanding of these stakeholder profiles but more importantly there is little research which describes the needs.

2. Research methodology

The doctoral research was done in three phases implementing design-based research methodology [9] – (1) domain analysis, (2) exploration of alternative assessment and (3) developing and validating the trade-off model. To better focus the research, we examined the research problem through three research questions:

[RQ1] What are the implications and alternative approaches of assessing educators' digital competence?

[RQ2] What are the stakeholder requirements and needs for educators' digital competence assessment?

[RQ3] How are the alternative assessment approaches established and sustained?

2.1. Research context

The doctoral research focuses on the Estonian educational setting and educators. Based on Lucas et al. [10] educators' digital competence is considered as a complex concept due to the set of factors which include personal characteristics, social, cultural, pedagogical and ethical considerations.

Estonia operates in a decentralized educational system which allows competition between schools but also provides school and educator autonomy [11]. Autonomy is considered educators collective right to determine the way they implement the schools' curriculum in their classes while choosing suitable pedagogical methods, tools, materials and also technologies [12]. Educators autonomy is closely linked to professionalism where after initial teacher training period any form of examination or testing is not expected or accepted by the educators. Although, teachers are required to regularly commit to professional development activities there is minimal monitoring or control mechanism.

3. Phase 1 - Educators digital competence

The first phase of the research was to understand and delineate the concept and domain of educators' digital competence and assessment. This phase was guided by the research question:

[RQ1] What are the implications and alternative approaches of assessing educators' digital competence?

We carried out a systematic literature review (SLR) [13] following the methodological example of Siddiq et al. [14]. The SLR database search was carried out during March 2018 to January 2019. For clear overview of the field we first identified the underlying synonyms and alternative phrases for database search. The used terms included – digital competence: digital competency, ICT literacy, digital literacy, ICT skills, digital skills, computer skills, technology literacies, digital competencies and 21st century skills. To get an overview of the instruments developed based on the frameworks and models we also limited the database search based on the terminology related to measurement – assessment, evaluation, testing, measuring, questionnaire. Literature screening resulted 40 suitable studies which made up the literature used in the SLR.

Based on the analysis the SLR provided four key results which helped to better define the concept of educators' digital competence. Additionally, the results provided the first insight to the implications related to the alternative assessment approaches and instrument.

First, the SLR confirmed that majority of the educators' digital competence assessment related research focuses on quantitative studies by implementing self-assessment instruments and there is a clear lack of qualitative research to accompany the results to explain the reliability and validity of the instruments.

Secondly, used self-assessment instruments are created ad-hoc often based on country specific framework and targeted specific group of educators (i.e. in-service teachers, student teachers etc.).

Third and considerably most fundamental result revealed that self-assessment is often one-dimensional, meaning that there is relatively low possibility to understand and explain why and how educators approach digital competence self-assessment. To this end it is important to embed alternative assessment approaches like testing or authentic assessment – including portfolios, reflective journals and observations to understand educators' perceptions of their competence and make sense of the evidence provided by the educators. Furthermore, alternative and combined competence assessment would potentially further the research if educators assess their digital competence r rather self-efficacy or self-confidence.

The final key result of the SLR presented the need for validated guidelines for the digital competence assessment processes. One of the proposed solutions was a large-scale participatory research which would focus on piloting alternative assessment instruments and approaches.

Based on the SLR results we concluded that the future research lines included following the DigCompEdu framework [15] for educators which covers EU level specifics of educators pedagogical practice and the derivatives or predecessors were presented in the majority of the analysed literature. The results also pulled focus on piloting and analyzing alternative assessment approaches to self-assessment to better understand the implications.

4. Phase 2 – Alternatives in digital competence assessment

The second and most extensive phase of the study focused on implementing alternative digital competence assessment instruments based on the DigCompEdu framework [15] which was the contextual basis of the for the following research. The second phase of the study followed two research questions:

RQ1] What are the implications and alternative approaches of assessing educators' digital competence?

[RQ2] What are the stakeholder requirements and needs for educators' digital competence assessment?

While the main focus of this phase was to identify the implications of alternative approaches, the research done also gave input to the related stakeholder groups and the respective needs.

During this phase four studies were conducted which included self-assessment instruments, knowledge-based testing and e-portfolio based digital competence assessment approaches. The focus of the four studies was the following:

Study 1 – In-service teachers' perceptions of digital competence during distance learning period.

Study 2 – Comparative multiple-case study of three combined self-assessment and knowledge-based testing digital competence assessment approaches.

Study 3 – SELFIE4Teachers [16] instrument based mixed methods study combining self-assessment and nominal group technique (NGT) [17] group interview.

Study 4 – Competence based LMS[2] focusing on e-portfolio based assessment of digital competence.

Table 1 describes the methodology, research instrument, samples and timeline of these studies.

Table 1

Second phase studies.

	Study 1	Study 2	Study 3	Study 4
Methodology	Quan	Quan	MM	Qual
Instrument	SA	SA&KB	SA&NGT	Auth.
Sample	1125	2248	18	84
Study time	2020	2019-2021	2022	2022

SA – Self-assessment.
KB – Knowledge-based test.
NGT – Nominal Group Technique group interview.
Auth. – Authentic assessment using e-portfolio.

Main results of the four studies can be described in the following key ideas. First, when implementing self-assessment instruments, on average, educators assess their digital competence as average technology users. In some cases, this describes the educators' inability of assessing their own competence and once again presents the question whether they assess digital competence or perceived self-confidence.

Second outcome of the studies revealed that educators are unable to provide appropriate evidence to describe their digital competence. As always there are exceptions, but the main issue lies in the fact that educators do not differentiate the different digital competence dimensions [15] (professional engagement; digital resources, teaching and learning, assessment, empowering learners and facilitating learners' digital competence) and provide low-level generic evidence.

The third result describes the educators' expectations towards the assessment instrument, stating that the used instruments often include hard to understand concepts and definitions. Simultaneously, the educators brought out issues with the instrument length, time spent on completion and the feedback report usability.

The final contribution of the four studies relates to the validity, reliability and sustainability of the used instruments. Based on the research we concluded that although there are a lot of efforts in designing and developing these assessment instruments they often lack in reliability. Additionally, as instrument validity is a multifaceted concept (i.e. face validity, construct validity etc.) it boils down to the stakeholder needs. The second phase of the doctoral research also confirmed that there is a continuous issue with digital competence assessment instrument sustainability where focus on re-designing and developing new instruments is considered of higher priority, rather than updating the excising instruments.

5. Phase 3 - Trade-offs in digital competence assessment

The third and final phase of the research focuses on identifying the stakeholder specific trade-offs in educators' digital competence assessment, developing and validating the trade-off model. This phase followed two research questions:

[RQ2] What are the stakeholder requirements and needs for educators' digital competence assessment?

[RQ3] How are the alternative assessment approaches established and sustained?

The third phase included two main studies where the first focused on identifying the stakeholder profiles (in-service teacher, student teacher, advanced teacher, teacher trainer, educational technologist, school leader, qualification examination assessment board member) and scenarios and on the stakeholder expectations and needs, resulting in the first version of the trade-off model. The study was a combined quantitative (N=1125) and qualitative (N=4) methodology.

The second and final study of the doctoral research included the validation of the stakeholder profiles and the trade-off model. The study was done following a Nominal Group Technique and included representatives of each stakeholder profile (N=6).

As this phase of the research is still underway the following describes initial outcomes. We consider noteworthy that all stakeholders consider the process of digital competence assessment valuable which helps to understand the professional development needs of educators. Furthermore, the inductive analysis of the differences in stakeholder needs

[2] https://edidaktikum.ee

gave us a clear indication that it is nearly impossible to provide a reliable and of high validity universal digital competence assessment instrument. This means that a trade-off model could provide a solution to adhere to the stakeholder needs. The results also provide deeper understanding on the stakeholder specific scope and dimension of educators' digital competence assessment expectations.

6. Conclusion

The doctoral research is currently in the final stages where our efforts are focused on publishing the results of finalized studies and formulating the analytical overview and main scientific contributions.

While digital competence assessment and more specifically educators' digital competence has been an ongoing discussion and research topic for more than 15 years our research provides a new dimension to understanding the assessment instruments, approaches and processes. This doctoral research can be described a metalevel research which aims to describe and provide solutions for the digital competence assessment through multiple stakeholder lens rather than trying to provide one universal solution to a multifaceted research problem.

7. Acknowledgements

The doctoral research has received funding from:
1. The European Union's Horizon 2020 research and innovation program under grant agreement No. 669074. Activity was supported through and according to Mobilitas Plus MOBEC001 CEITER action plan.
2. The European Union's Horizon 2020 research and innovation program under grant agreement No. 856954.

8. References

[1] Ferrari, A. Digital Competence in Practice: An Analysis of Frameworks. (2013). Joint Research Centre of the European Commission., 91. https://doi.org/10.2791/82116.

[2] Broadfoot, P., & Black, P. Redefining assessment? The first ten years of assessment in education. Assessment in Education: Principles, Policy and Practice, 11(1), 7–26. (2004). https://doi.org/10.1080/0969594042000208976

[3] Harlen, W., & James, M. Assessment and learning: Differences and relationships between formative and summative assessment. International Journal of Phytoremediation, 21(1), 365–379. (1997). https://doi.org/10.1080/0969594970040304

[4] Benali, M., Kaddouri, M. & Azzimani, T. Digital competence of Moroccan teachers of English. International Journal of Education and Development using ICT, 14(2). (2018). Open Campus, The University of the West Indies, West Indies. Retrieved May 6, 2022 from https://www.learntechlib.org/p/184691/.

[5] Kluzer, S., & Priego, L. P. DigComp into action: get inspired, make it happen. Publications Office of the European Union. JRC Science for Policy Report, EUR 29115 EN. Editors: S. Carretero, Y. Punie, R. Vuorikari, M. Cabrera, & W. O'Keefe. (2018). https://doi.org/10.2760/112945

[6] Siddiq, F., Hatlevik, O. E., Olsen, R. V., Throndsen, I., & Scherer, R. Taking a future perspective by learning from the past - A systematic review of assessment instruments that aim to measure primary and secondary school students' ICT literacy. Educational Research Review, 19, 58–84. (2016). https://doi.org/10.1016/j.edurev.2016.05.002

[7] Zenouzagh, M. Z. The effect of online summative and formative teacher assessment on teacher competences. Asia Pacific Education Review, 20(3), 343–359. (2019). https://doi.org/10.1007/s12564-018-9566-1

[8] Struyven, K., Blieck, Y., & De Roeck, V. The electronic portfolio as a tool to develop and assess pre-service student teaching competences: Challenges for quality. Studies in Educational Evaluation, 43, 40–54. (2014). https://doi.org/10.1016/j.stueduc.2014.06.001

[9] Edelson, D. C. Design Research: What We Learn When We Engage in Design. The Journal of the Learning Sciences, 11(1), 105–121. (2002). http://dx.doi.org/10.1207/S15327809JLS1101_4

[10] Lucas, M., Bem-Haja, P., Siddiq, F., Moreira, A., & Redecker, C. The relation between in-service teachers' digital competence and personal and contextual factors: What matters most? Computers and Education, 160 (March 2020). (2021). https://doi.org/10.1016/j.compedu.2020.104052

[11] Erss, M. Complete freedom to choose within limits'–teachers' views of curricular autonomy, agency and control in Estonia, Finland and Germany. Curriculum Journal, 29(2), 238–256. (2018). https://doi.org/10.1080/09585176.2018.1445514

[12] Lawson, T. Teacher autonomy: Power or control? Education 3-13, 32(3), 3–18. (2004).

[13] Sillat, L. H., Tammets, K., & Laanpere, M. Digital competence assessment methods in higher education: A systematic literature review. Education Sciences, 11(8). (2021). https://doi.org/10.3390/educsci11080402 Yur'yev

[14] Siddiq, F.; Hatlevik, O.E.; Olsen, R.V.; Throndsen, I.; Scherer, R. Taking a future perspective by learning from the past—A systematic review of assessment instruments that aim to measure primary and secondary school students' ICT literacy. Educ. Res. Rev. 19, 58–84. (2016).

[15] Redecker, C., & Punie, Y.: Digital Competence of Educators DigCompEdu. (2017).

[16] European Commission. Digital Education action Plan 2021-2027 Resetting education and training for the digital age. Commission Staff Working Document, (SWD (2020) 209 final), 1–103. (2020).

[17] Delbecq, A. L., & Van de Ven, A. H. A Group Process Model for Problem Identification and Program Planning. The Journal of Applied Behavioral Science. (1971).

Methods and perspectives for the automated analytic assessment of free-text responses in formative scenarios

Sebastian Gombert*

DIPF | Leibniz Institute for Research and Information in Education, Frankfurt am Main, Germany

Abstract

Assessment is the process of testing learners' skills and knowledge. Free-text response items are well suited for the assessment of learners' active knowledge and writing skills. However, the automatic assessment of respective responses is not trivial and requires the application of natural language processing. Accordingly, the automatic assessment of free-text responses is a widely researched topic in educational natural language processing. Most past work targets holistic scoring, the process of assigning overall scores or grades to responses. This is problematic in formative scenarios because learners require feedback rather than summative scores in such scenarios. Such feedback ideally targets specific aspects of responses, and, accordingly, automated systems which only predict holistic scores cannot be used as a basis for providing the same. What is instead needed are systems which implement analytic scoring approaches. Analytic scoring targets specific aspects of responses and scores them according to corresponding criteria. This requires different systems than addressed by the broad research on automated holistic scoring. In my PhD work which is outlined by this paper, I want to explore approaches for implementing analytic scoring systems by means of state-of-the-art natural language processing. These systems are targeted at providing a basis for feedback generation.

Keywords

Assessment, Automated Assessment, Analytic Assessment, Short Answer Grading, Essay Grading

1. Introduction

Educational assessment is the process of empirically measuring and documenting learners' skills and knowledge [1]. This is conducted through tests composed of various kinds of test items. Assessing learners' knowledge and skills is also the basis for providing them with appropriate content-related feedback in formative scenarios [2]. In the context of technology-based assessment, multiple-choice items have grown to be a popular choice to implement tests [3, 4]. This is mostly the case since evaluating multiple-choice items is rather trivial. Test creators simply need to define a set of responses out of whom they define one or more as the correct ones. When test-takers select respective responses during testing, the computer only needs to determine which of them were among the correct ones. Moreover, multiple-choice items take only a short time to answer which makes it possible to include many different of them within tests and test for a broad range of knowledge [4].

However, not every skill and every kind of knowledge can be assessed through multiple-choice items. "A multiple-choice test for history students can test their factual knowledge. It can also determine whether they can discriminate between correct and incorrect statements of the relationships between facts — but it cannot determine whether the students can write a well-reasoned essay on a historical question. [...] A multiple-choice test of writing ability can determine whether the test takers can discriminate between well written and badly written versions of a sentence — but it cannot determine whether they can organize their own thoughts into a logically structured communication in clear and appropriate language" [4]. Moreover, multiple-choice cannot test for active knowledge. A test-taker might simply conduct (informed) guessing and there is no guarantee that they would have been able to actively reproduce this knowledge.

2. Constructed Responses and their Automatic Assessment

To test skills such as the ones described by [4], constructed response items are needed instead multiple choice items. In their most common form, they require students to enter a free text as response into a text field. However, this drastically increases the complexity of assessing learners' responses in an automated fashion, as the computer-based analysis of human language is far from trivial. With natural language processing respectively computational linguistics, a whole interdisciplinary field of research building upon various methods and theories from linguistics, artificial intelligence, statis-

Proceedings of the Doctoral Consortium of Seventeenth European Conference on Technology Enhanced Learning, September 12–16, 2022, Toulouse, France.

*Corresponding author.
gombert@dipf.de (S. Gombert)
https://edutec.science/team/sebastian-gombert/ (S. Gombert)
0000-0001-5598-9547 (S. Gombert)

tics, logic, psychology, cognitive science, software engineering and philosophy is dedicated to this issue, and the automatic processing of many aspects of language remains open research. What makes the automatic analysis of free text difficult are the properties of language itself. Humans can generate an unlimited set of different linguistic utterances, and often, there are many ways to express the same or similar semantics, i.e., through different synonyms, the usage of passive vs. active constructions, or ways of paraphrasing. In past research, many different methods were applied to the automatic assessment of free-text responses. These range from simpler keyword, pattern and regular expression searches, and methods building upon distributional vector space semantics, to fully-fledged machine learning systems [5, 6].

Most recently, transformer language models such as BERT [7] were successfully applied to the problem of free-text assessment [8, 9, 10, 11]. The application of transformers to the assessment of constructed responses promises major advancements in the field, but nonetheless, most of the systems available are built to predict only holistic scores [5, 6], ergo scores aimed at denoting the overall quality of a response [4]. Most of the established datasets, especially the ones focused on short answers, also cater towards this approach [5, 9, 6]. While holistic scores reflect how well learners were able to overall solve a given task, they do not necessarily denote which aspects of their response were of good quality and in which regards they could improve. However, especially in formative scenarios, providing students with feedback is crucial, which puts the application of holistic scoring systems in formative scenarios into question.

There is a second scoring approach in constructed response assessment which can be seen as a better basis for providing detailed, personalized feedback: analytic scoring. In analytic scoring, rather than judging responses as a whole, they are assessed for multiple different aspects which need to be specifically defined in a coding rubric [4]. I.e., "[o]n a science question, the scorer may award two points for providing a correct explanation of a phenomenon, one point for correctly stating the general principle that it illustrates, and one point for providing another valid example of that principle in action" [4]. Drawing such distinctions and coding responses for multiple different aspects allows to provide more detailed and concise feedback as the same can specifically address these aspects.

3. Research Questions

The two most common types of free-text responses are short answers and essays. While short answers are used to test students' ability to explain phenomena or demonstrate their active knowledge, essays are used for analysing their writing skills of students, e.g., their skill to clearly and coherently discuss or communicate a given issue or argue against or in favour of an opinion. Accordingly, approaches for the analytic assessment of both text forms must inevitably differ. For short answers, it presumably should be sufficient to simply assess whole responses for the different aspects, as short answers are rather condensed texts. From a formal point of view, this can be interpreted as a (multi-label) text classification task [5].

On the other hand, for essays, the respective coding can require more varied approaches. Are the aspects coded related to content or writing style? Does a content-specific code apply to the whole text or to specific sections? These questions need to be addressed in order to come to appropriate operationalisations. E.g., if it is likely that each code corresponds to a specific part of an essay, one needs to first semantically segment it into the respective parts. One could then in a second step separately classify these parts for the actual codes. On the other hand, if a code corresponds to a whole essay, such separation is not needed.

I plan my PhD to be paper-based where the single papers are connected by the overarching topic of analytic constructed response coding. First and foremost, I want to explore what has been already done in past work and how my own work can benefit from these insights. The acquired knowledge is then to be used for the practical implementation of constructed response scoring systems in a range of case studies. For these case studies, I plan to leverage data sets from several research projects I am involved in. In the projects *AFLEK* and *ALICE*, I have access to a set of short answers to different science-related tasks with detailed coding rubrics focusing on scientific knowledge and argumentative skills. On the other hand, the project *HIKOF* provides a data set of essays in which students discuss learning tips from a YouTube video with respect to their grounding in educational psychology. Both data sets are coded in a way which allows for the implementation of automated analytic assessment systems.

Another important aspect of my work is the question how codes from response scoring systems can be transformed into concrete learner feedback. Feedback can be given on an item-specific level as well as on a more global one. It can focus the content or the form of concrete responses, and it can also target the overall domain knowledge of a student across multiple items. For the prior case, generative language models could be promising [9, 12]. For the latter case, a way of modeling learners' domain knowledge is required. A conceptual framework which goes into this direction was provided by [13] with their *expanded evidence-centred design* model, which adds multiple feedback-related aspects to the well-known *evidence-centred design* [14]. However, to my best knowledge, this conceptual framework was not operationalised into a concrete feedback-driven assessment system so far.

The last aspect I want to address is the one of explainability. Ethical frameworks in learning analytics and educational technology such as [15] often call for the application of transparent and explainable models where possible. It is likely that providing learners with simple explanations on why models made a given prediction, which, in turn, led to a particular feedback outcome, can increase their acceptance for respective systems. For natural language processing models, a wide range of methods for providing such explanations has been developed [16]. Research for making state-of-the-art methodology explainable also shows promising results, e.g. [17]. For this reason, I want to leverage this potential and explore, if providing learners with explanations for their feedback can increase trust.

To summarzie, I want to address the following research questions:

1. What were the main methods, characteristics and results of past work in constructed response scoring?
2. What techniques were applied for coding constructed responses in an analytic fashion in past work?
3. What machine learning-based pipelines and approaches are effective for the automated analytic assessment of constructed responses and to what extent can they be generalized?
4. How can the predictions of automated analytic assessment systems be transformed into useful learner feedback?
5. To what extent can explaining model outputs make learners trust in the provided feedback?

4. Design

From a technical perspective, the intention behind my PhD work is to implement and evaluate respective methods for the analytic assessment of free-text responses for exemplary use cases drawing from state-of-the-art NLP research. I plan to study and summarize what methods were applied to the assessment of free-text responses in past work via a literature review to address RQ1 and RQ2. For this literature review, I plan to draw from past reviews on the topic, in particular [5] for the text type of short answers and [6] for the text type of essays, but primarily with a focus on work which was not covered by them. The main goal behind the literature review is to provide a concise overview over the methods and features which can be successfully applied to the task.

The review by [5] is, thanks to its publication date, fairly outdated. Moreover, in my opinion, it fails to function as a lookup guide for possible techniques to use, and rather focuses on summarizing papers from past work. The review by [6], on the other hand, is well structured but also fairly short thanks to it being published in conference proceedings. The plan for my literature review is to primarily act as a guide for practitioners which they can refer to when they plan to build their own free-text assessment systems rather than as a pure overview over past work. It shall equip interested researchers with a clear plan on how they can approach their own free-text response assessment system in a structured manner.

The next papers deal with the implementation of respective systems themselves to address RQ3. The most recent achievements in holistic free-text response assessment, in line with the general developments in natural language processing, were achieved using transformer language models [8, 9, 10, 11]. For this reason, my plan is to also apply transformer language models to the task of analytic assessment. However, [5] and [6] document a wide range of methods from the pre-transformers era. It is an interesting question In this context, my plan is to implement and evaluate exemplary systems for assessing both short answers and essays in an analytic fashion.

In a first research paper, which is currently under review, I implemented and evaluated multiple systems aimed at assessing German middle school students' knowledge about energy physics. In particular, the systems classify if students mentioned certain concepts related to energy transformation, i.e., different manifestations of energy, indicators for the same, and if energy is transformed, in a meaningful manner. For this purpose, first data was collected and coded using a coding rubric which targeted the different categories of knowledge. I then implemented and evaluated multiple text classification systems trained to replicate the coding for the respective purpose, transformer- and feature-based. The systems are given the response, a provided sample solution and the item prompt. Moreover, using different methods for generating model explanations, I evaluated the descriptive accuracy of the implemented models. Overall, a transformer-based model based upon GBERT could achieve superior results. In subsequent research, I want to explore how well the predictions of such systems can be concretely translated into feedback.

In another research paper, I want to implement systems targeting essays. In particular, I aim to use a data set of essays collected throughout the *HIKOF* project. These essays discuss ten different learning tips presented in a YouTube video with respect to their grounding in educational and psychological research. For each tip, ten different codes were assigned. Moreover, it was coded which sentences within an essay correspond to which tips. This results in two problems which need to be solved.

First, unseen essays must be segmented into sections corresponding to the different tips. This can be approached as a sentence classification task. In a second step, the resulting sections must then be given to a second text classification system which classifies the sections with respect to the analytic codes corresponding to each tip.

In the next step, feedback needs to be generated from the predicted codes. For this purpose, I use content-related feedback templates which are assembled dynamically depending on the predicted codes. In particular, the predicted codes are matched with ground truth codes, and discrepancies between the two lead to The generated feedback will be tested within a university lecture in an AB setup. In a followup study, I plan to add aspects of explainability to this feedback. In particular, I plan to present learners with highlighted text of what exactly in their response led to a concrete feedback in an AB setup. This shall then be combined with questionnaires evaluating if showing these explanations to learners increases acceptance. For educational recommender systems, findings from [18] suggest that showing explanations to learners can increase the acceptance for respective systems. I want to find out if this is also the case for assessment-driven feedback systems.

5. Conclusion

In this document, I presented my PhD project which deals with systems for the automatic assessment of constructed responses in formative scenarios implemented through machine learning-based natural language processing. In particular, I explore the implementation and evaluation of respective systems for multiple use cases. Moreover, I plan to write a literature review on constructed response scoring in the form of a practitioner lookup guide. Finally, I then want to explore how codes predicted by automatic assessment systems can be translated into automatic actionable feedback, and if explaining the model predictions behind this feedback can contribute to the acceptance of these systems.

References

[1] R. Nelson, P. Dawson, A contribution to the history of assessment: how a conversation simulator redeems socratic method, Assessment & Evaluation in Higher Education 39 (2013) 195–204. URL: https://doi.org/10.1080%2F02602938.2013.798394. doi:10.1080/02602938.2013.798394.

[2] P. Black, D. Wiliam, Assessment and classroom learning, Assessment in Education: Principles, Policy & Practice 5 (1998) 7–74. URL: https://doi.org/10.1080/0969595980050102. doi:10.1080/0969595980050102. arXiv:https://doi.org/10.1080/0969595980050102.

[3] S. K. Mangal, S. Mangal, Assessment for Learning, PHI Learning, Delhi, India, 2019.

[4] S. A. Livingston, Constructed-response test questions: Why we use them; how we score them. r&d connections. number 11., Educational Testing Service (2009).

[5] S. Burrows, I. Gurevych, B. Stein, The eras and trends of automatic short answer grading, International Journal of Artificial Intelligence in Education 25 (2014) 60–117. URL: https://doi.org/10.1007%2Fs40593-014-0026-8. doi:10.1007/s40593-014-0026-8.

[6] Z. Ke, V. Ng, Automated essay scoring: A survey of the state of the art, in: Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, International Joint Conferences on Artificial Intelligence Organization, 2019. URL: https://doi.org/10.24963%2Fijcai.2019%2F879. doi:10.24963/ijcai.2019/879.

[7] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 4171–4186. URL: https://aclanthology.org/N19-1423. doi:10.18653/v1/N19-1423.

[8] L. Camus, A. Filighera, Investigating transformers for automatic short answer grading, in: Lecture Notes in Computer Science, Springer International Publishing, 2020, pp. 43–48. URL: https://doi.org/10.1007%2F978-3-030-52240-7_8. doi:10.1007/978-3-030-52240-7_8.

[9] A. Filighera, S. Parihar, T. Steuer, T. Meuser, S. Ochs, Your answer is incorrect... would you like to know why? introducing a bilingual short answer feedback dataset, in: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Dublin, Ireland, 2022, pp. 8577–8591. URL: https://aclanthology.org/2022.acl-long.587. doi:10.18653/v1/2022.acl-long.587.

[10] A. Poulton, S. Eliens, Explaining transformer-based models for automatic short answer grading, in: 2021 5th International Conference on Digital Technology in Education, ACM, 2021. URL: https://doi.org/10.1145%2F3488466.3488479. doi:10.1145/3488466.3488479.

[11] C. Sung, T. Dhamecha, S. Saha, T. Ma, V. Reddy, R. Arora, Pre-training BERT on domain resources for short answer grading, in: Proceed-

ings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Association for Computational Linguistics, Hong Kong, China, 2019, pp. 6071–6075. URL: https://aclanthology.org/D19-1628. doi:10.18653/v1/D19-1628.

[12] A. Filighera, J. Tschesche, T. Steuer, T. Tregel, L. Wernet, Towards generating counterfactual examples as automatic short answer feedback, in: Lecture Notes in Computer Science, Springer International Publishing, 2022, pp. 206–217. URL: https://doi.org/10.1007%2F978-3-031-11644-5_17. doi:10.1007/978-3-031-11644-5_17.

[13] M. Arieli-Attali, S. Ward, J. Thomas, B. Deonovic, A. A. von Davier, The expanded evidence-centered design (e-ECD) for learning and assessment systems: A framework for incorporating learning goals and processes within assessment design, Frontiers in Psychology 10 (2019). URL: https://doi.org/10.3389%2Ffpsyg.2019.00853. doi:10.3389/fpsyg.2019.00853.

[14] R. J. Mislevy, R. G. Almond, J. F. Lukas, A brief introduction to evidence-centred design, ETS Research Report Series 2003 (2003) i–29. URL: https://doi.org/10.1002%2Fj.2333-8504.2003.tb01908.x. doi:10.1002/j.2333-8504.2003.tb01908.x.

[15] S. Slade, A. Tait, Global guidelines: Ethics in learning analytics (2019).

[16] M. Danilevsky, K. Qian, R. Aharonov, Y. Katsis, B. Kawas, P. Sen, A survey of the state of explainable AI for natural language processing, in: Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing, Association for Computational Linguistics, Suzhou, China, 2020, pp. 447–459. URL: https://aclanthology.org/2020.aacl-main.46.

[17] H. Chefer, S. Gur, L. Wolf, Transformer interpretability beyond attention visualization, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 782–791.

[18] K. Takami, Y. Dai, B. Flanagan, H. Ogata, Educational explainable recommender usage and its effectiveness in high school summer vacation assignment, in: LAK22: 12th International Learning Analytics and Knowledge Conference, LAK22, Association for Computing Machinery, New York, NY, USA, 2022, p. 458–464. URL: https://doi.org/10.1145/3506860.3506882. doi:10.1145/3506860.3506882.

Supporting self-regulated learning in a blended learning environment using prompts and learning analytics

Sabina Rako[a,b], Diana Šimić[a] and Bart Rienties[c]

[a] University of Zagreb Faculty of organization and informatics, Pavlinska 2, Varazdin, 42000, Croatia
[b] University of Zagreb University Computing Centre, Josipa Marohnica 5, Zagreb, 10000, Croatia
[c] Open University, Milton Keynes MK7 6AA, United Kingdom

Abstract

Higher education institutions, teachers, and students face new difficulties and opportunities resulting from the introduction of modern technology into the learning process. The widespread of learning environments that integrate online learning and face-to-face learning may pose some opportunities as well as difficulties for some groups of students' self-regulation skills. Providing automated prompts may help to support those students with insufficient self-regulation skills. The use of learning analytics and multiple methods and data sources (data triangulation) may give better insight into the self-regulation process.

The objective of the proposed research is to explore the students' evaluation of the usefulness of prompts implemented in a blended learning environment. A secondary objective is to develop and evaluate a real-time dashboard designed to notify teachers of student responses to deployed prompts.

The research methodology will be grounded in action research and empirical research. The scientific contribution will be achieved through the development of artefacts and the performance of empirical research to advance understanding of the student's self-regulation in a blended learning environment.

Keywords

learning analytics, self-regulated learning, prompts, blended learning, dashboards, higher education

1. Introduction

In the past two decades, blended learning in higher education has been increasingly widespread [1]. The effectiveness of blended learning in relation to traditional learning is continuously reviewed [2,3]. Recently, Müller and Mildenberger [4] conducted a meta-analysis of scientific papers published from 2008 to 2019 and found that identical learning outcomes were achieved in blended learning as in a conventional classroom setting, with a reduction of time spent in physical space by 30 to 79% (division according to Allen et al. [5]). This research also revealed that it is not yet possible to identify for which specific competencies (or disciplines) a blended learning format is most appropriate.

Several teachers and institutions strive to develop personalised learning approaches in an effort to meet the needs of each student to the greatest extent possible. To be able to customise the approach, it is necessary to examine the views and habits of students. For example, information systems deployed in the teaching and learning process are sources of valuable educational data that may be used to monitor and assess the teaching and learning process

Proceedings of the Doctoral Consortium of Seventeenth European Conference on Technology Enhanced Learning, September 12–16, 2022, Toulouse, France
EMAIL: sabina.rako@srce.hr (A. 1); diana.simic@foi.unizg.hr (A. 2); bart.rienties@open.ac.uk (A. 3)
ORCID: 0000-0002-8457-3089 (A. 1); 0000-0002-6721-7250 (A. 2); 0000-0003-3749-9629 (A. 3)

[6], and play a vital part in the development of personalised solutions.

Learning analytics as a research area is focused on 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" [7]. The implementation of learning analytics is a complex process that requires capability building and certain specific competencies of stakeholders in the education system. In practice, learning analytics examples can be found at several levels (e.g., students, courses, programmes, institutions, and consortiums of institutions) [8]. When applying learning analytics, technology should be used wisely taking into account existing educational concepts and research knowledge [9].

Tsai et al. [10] provided an overview of trends and limits in the deployment of learning analytics in the European higher education system. According to their research, teachers and teaching staff are the primary users of learning analytics, and there is limited evidence of active engagement with students and the use of learning analytics to improve self-regulated learning skills.

Self-regulated learning includes cognitive, metacognitive, behavioural, motivational, and emotional aspects of learning. This area has been extensively researched in the field of educational psychology, and among the best known and most applied models is the Zimmerman's model of self-regulated learning, that consists of three main phases: (a) forethought, (b) performance, and (c) self-reflection [11]. Wong et al. [12] in a systematic review of self-regulated learning in an online environment and massive open online courses (MOOCs) demonstrated the need for further research of self-regulated learning in an online environment, particularly through an empirical approach. Furthermore, Viberg et al. [13] examined empirical research in which learning analytics were used to improve self-regulated learning and concluded that few studies related to the self-reflection phase of the Zimmerman model, and that the majority of research focused on measuring self-regulated learning and less on support.

In previous research, feedback and prompts have been identified as the most important elements that encourage self-regulated learning [12]. Prompts are "visual, textual, or spoken elements that the teacher uses to encourage understanding and are most often in a form of questions, although they can also be formulated in the form of advice or instructions" [14]. Another definition of prompts is "short hints or questions presented to students in order to activate knowledge, strategies or skills that students have already available but do not use" [15]. Additionally, students do not usually manifest self-regulated behaviour spontaneously without guidance [16]. Despite the fact that the research revealed a number of potential advantages of prompts for self-regulated learning, Schumacher and Ifenthaler [17] reported that learning analytics approaches have not been thoroughly examined during prompt implementation, and that future studies should also focus on the student's responses to prompts.

The proposed research will also consider learning design as an important element in educational interventions.

Specifically, these research questions will drive the proposed research.

RQ1: To what extent are students aware of self-regulation elements, such as metacognitive activities before/during/after learning, environmental structuring, help seeking, and time management in the blended learning environment?

RQ2: In a blended learning environment, which types of prompts (cognitive, metacognitive, motivational, or content-related) do groups of students find most useful?

RQ3: Is there a difference in perceived usefulness of the same type of prompt based on the mode of learning (online and face-to-face)?

RQ4: How does the implementation of specific prompts affect

(a) student's engagement

(b) results achieved in formative assessment

(c) overall learning satisfaction?

What distinctions exist amongst student groups?

RQ5: Which components of the real-time dashboard for displaying student feedback on prompt implementation are important to students and/or teachers?

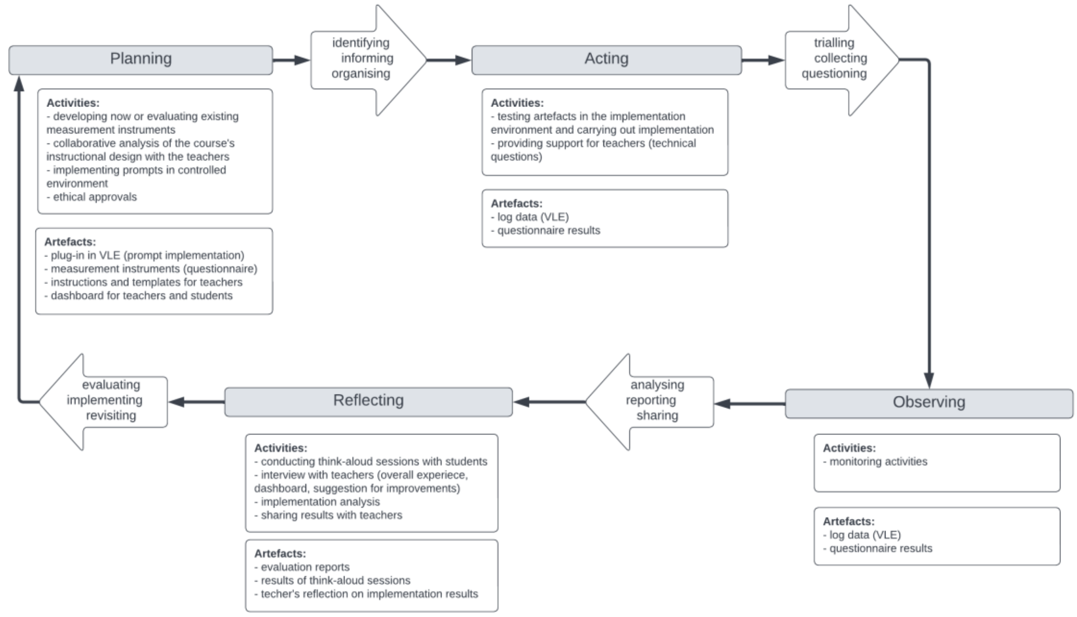

Figure 1: Proposed activities and key artefacts based on steps in Somekh's action research process (Source: Author)

2. Methodology

This proposed research will utilise a mixed-method practical action research design. According to Creswell [18], action research is used to address specific, practical issues that seek solutions to a problem, and both quantitative and qualitative methods may be employed. Somekh [19] proposes a four-step process for action research: planning, acting, observing, and reflecting. The proposed activities in each action research step and key artefacts are shown in Figure 1. Several research methods, including descriptive statistics, natural language processing methods (open-ended questions), statistical analysis, and nonparametric tests, will be utilised for data analysis. For statistical analysis, the statistical programming language R [20] will be used.

2.1. Planning

The initial literature review showed the research gap in the area of learning analytics approaches in investigating prompts for supporting students' self-regulation. During the preparation phase, an additional literature review will be conducted to synthesise the findings of prior research, identify appropriate measurement instruments, and provide an overview of the outcomes of prior empirical interventions.

The intervention will be designed as an iterative process, with a pilot trial followed by the main study. The interventions are intended to be implemented at two higher education institutions in Croatia, aiming to target around 340 students and 3 teachers. Ethical approval from participating higher education institutions will be obtained.

Teachers will be closely involved in preparations for implementation (analysis of current learning design of a course, defining specific goals of prompt implementation, finding appropriate learning types, and defining prompts based on selected models).

During this phase, the appropriate measurement instruments will be evaluated (linguistic evaluation) or, if necessary, a new measurement instrument will be developed.

2.2. Acting

This activity is a key component of the research proposal. During this phase, the developed artefacts will be used in the real environment.

The dominant research method used will be pretest-posttest nonequivalent groups design, a type of quasi-experimental design. One group of students will be exposed to an intervention, while the other group will not. The two groups will then be compared. According to previous research [21], in order to eliminate confounding variables, the duration of exposure should not be excessively long (preferably 2 - 4 weeks).

Before the intervention, a priori statistical power analysis will be conducted to determine the required number of outcome observations.

During this stage, the measurement instruments will be evaluated in a real environment.

2.3. Observing

In this phase, monitoring activities and providing teachers with adequate technical support will be the primary activities. Data will be collected via system logs, measurement instruments and prompt feedback.

To monitor student progress, teachers will have access to a real-time dashboard with visualisations of student responses.

2.4. Reflecting

Teachers will receive the intervention results during the phase of reflection. In addition, they will assess the real-time dashboard that was accessible during the observing phase.

In addition, a think-aloud protocol [22] will be implemented to collect specific information about students' and teachers' experiences with prompt implementations.

3. Current results

A literature review with the focus on available measurement instruments (self-regulated learning, engagement, satisfaction and other relevant constructs) is currently in progress.

Based upon the initial reading of the literature and good practice identified, a prototype of plug-in for prompt implementation has been developed in Moodle LMS Platform (Figure 2). The plug-in makes it possible to embed prompts wherever an HTML editor is available.

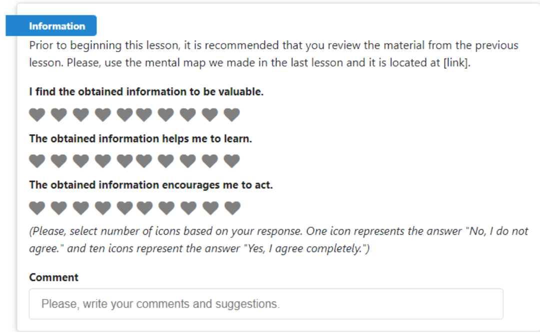

Figure 2: Prompt prototype. Students could rate prompts and give textual feedback (Source: Author)

Prototype of teacher's dashboard has been also developed (Figure 3).

Figure 3: Prototype of teachers' dashboard providing real-time monitoring of student's responses (Source: Author)

In order to test the feasibility of the proposed study, pre-pilot study has been conducted. 38 students gave consent to participate in the pre-pilot study. The students were second-year students of the informatology programme at the Faculty of Humanities and Social Sciences. 36 out of 38 students were female, while two were male.

Lessons learned from the pre-pilot study:

- the suggested plug-in is appropriate for prompt implementation and gives considerable design flexibility with respect to learning design
- students are more likely to rate prompts during face-to-face meetings than during online sessions
- the teacher acknowledged the advantages of monitoring student responses, and the input gained could be useful for designing course improvements
- think-aloud sessions conducted with two students gave valuable insights into the perception of implemented prompts
- adjustment of rating scale should be considered (10 or 7-level scale)

- it would be useful to collect additional demographic information in order to better understand behavioural differences among students.

4. Contribution to TEL domain

The expected contributions of the proposed research to the Technology Enhanced Learning (TEL) domain are:

- synthesis of empirical interventions and the results on supporting self-regulated learning with prompts using learning analytics in a blended learning environment
- development and evaluation of artefacts related to prompt implementation in real environment
- better understanding of students' self-regulation in blended learning environment using prompts
- results of empirical research on supporting self-regulated learning in blended learning environment using prompts and learning analytics. After completing experimental part of the proposed research, differences across student groups can be expected in terms of student engagement, formative assessment outcomes, and overall learning satisfaction. The combination of accessible students' demographic information with their responses and system data will provide insight into students' self-regulation practises and awareness.

5. Acknowledgments

This work has been fully supported by the Croatian Science Foundation under the project IP-2020-02-5071.

6. References

[1] M. Lundin, A. Bergviken Rensfeldt, T. Hillman, A. Lantz-Andersson, L. Peterson, Higher education dominance and siloed knowledge: a systematic review of flipped classroom research, International Journal of Educational Technology in Higher Education 20 (2018). doi: 10.1186/s41239-018-0101-6

[2] R. M. Bernard, E. Borokhovski, R.F. Schmid, R. M., Tamim, P.C. Abrami, A meta-analysis of blended learning and technology use in higher education: From the general to the applied, Journal of Computing in Higher Education, 26(1) (2014), 87-122. doi:10.1007/s12528-013-9077-3

[3] B. Anthony Jr., A. Kamaludin, A. Romli, A.F.M. Raffei, D. Nincarean A/L Eh Phon, A. Abdullah, G.L. Ming, N.A. Shukor, M.S. Nordin, S. Baba, Exploring the role of blended learning for teaching and learning effectiveness in institutions of higher learning: An empirical investigation, Education and Information Technologies, vol. 24, no. 6 (2019) 3433-3466. doi: 10.1007/s10639-019-09941-z

[4] C. Müller, T. Mildenberger, Facilitating flexible learning by replacing classroom time with an online learning environment: A systematic review of blended learning in higher education, Educational Research Review, Volume 34 (2021), ISSN 1747-938X. doi: 10.1016/j.edurev.2021.100394

[5] I.E. Allen, J. Seaman, R. Garrett, Blending in: The extent and promise of blended education in the United States, Newsburyport, MA: Sloan Consortium (2007)

[6] G. Siemens, Learning Analytics: The Emergence of a Discipline, American Behavioral Scientist, 57(10) (2013) 1380–1400. doi: 10.1177/0002764213498851

[7] Society for Learning Analytics Research, What is Learning Analytics?, Society for Learning Analytics, [Online]. Available: https://www.solaresearch.org/about/what-is-learning-analytics/ [Accessed, June 29, 2021]

[8] Tyton, Learning analytics strategy toolkit, [Online], Available: https://www.everylearnereverywhere.org/resources/learning-analytics-strategy-toolkit/ [Accessed, June 29, 2021]

[9] D. Gašević, S. Dawson, G. Siemens, Let's not forget: Learning analytics are about learning, Techtrends, vol. 59 (2015) 64-71. doi: 10.1007/s11528-014-0822-x

[10] Y. Tsai, D. Rates, P. M. Moreno-Marcos, P. J. Muñoz-Merino, I. Jivet, M. Scheffel, H. Drachsler, C. D. Kloos, D. Gašević, Learning analytics in European higher education-Trends and barriers, Computers & Education, Volume 155 (2020), 103933.

ISSN 0360-1315. doi: 10.1016/j.compedu.2020.103933
[11] E. Panadero, A review of self-regulated learning: Six models and four directions for research, Frontiers in Psychology 8, Article 422 (2017)
[12] J. Wong, M. Baars, D. Davis, T. Van Der Zee, G. Houben, F. Paas, Supporting Self-Regulated Learning in Online Learning Environments and MOOCs: A Systematic Review, International Journal of Human-Computer Interaction 35:4-5 (2018) 356-373
[13] O. Viberg, M. Khalil, M. Baars, Self-Regulated Learning and Learning Analytics in Online Learning Environments: A Review of Empirical Research, in: Proceedings of the 10th International Learning Analytics and Knowledge Conference, LAK'20, Association for Computing Machinery, New York, NY, 2020, pp. 173–186. ISBN: 978-1-4503-7712-6
[14] British Council, Prompts, British Council,[Online]. Available: https://www.teachingenglish.org.uk/article/prompts [Accessed, Jun. 27, 2021]
[15] J. Wirth, Promoting self-regulated learning through prompts, Zeitschrift für Pädagogische Psychologie, 23(2) (2009) 91-94
[16] C. Sonnenberg, M. Bannert, Evaluating the impact of instructional support using data mining and process mining: A micro-level analysis of the effectiveness of metacognitive prompts. Journal of Educational Data Mining, 8(2) (2016) 51-83
[17] C. Schumacher, D. Ifenthaler, Investigating prompts for supporting students' self-regulation - A remaining challenge for learning analytics approaches?, The Internet and Higher Education, Volume 49 (2021). doi: 10.1016/j.iheduc.2020.100791
[18] J. W. Creswell, Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research, fourth edition, Pearson, 2012
[19] B. Somekh, Action Research: A Methodology for Change and Development, 1st edition, Open University Press, 2005
[20] R Core Team, R: A language and environment for statistical computing, 2022. URL: https://www.R-project.org/
[21] L. Zheng, The effectiveness of self-regulated learning scaffolds on academic performance in computer-based learning environments: a meta-analysis, Asia Pacific Educ. Rev. 17 (2016) 187–202. doi: 10.1007/s12564-016-9426
[22] M. E. Fonteyn, B. Kuipers, S. J. Grobe, A description of think aloud method and protocol analysis, Qualitative Health Research, vol. 3, issue 4 (1993) 430-441, doi: 10.1177/104973239300300403

Towards a comprehensive framework for situated collaborative learning tools

Sebastian Simon, Iza Marfisi-Schottman, Sébastien George

LIUM, Le Mans Université, 72085 Le Mans, Cedex 9, France

Abstract
CSCL (Computer Supported Collaborative Learning) is a dynamic field that has considerably evolved in recent years. The result is a myriad of tools and theories that have emerged from numerous studies. While different studies shed light on different aspects of collaboration, a comprehensive connection between tool functionalities, learning activities and the collaboration processes they support has not been established yet. This PhD aims at providing a joint conceptual framework and environment to achieve this objective.
Keywords 1
Computer Supported Collaborative Learning, framework, collaborative processes, tool

1. Introduction

The field of CSCL (Computer Supported Collaborative Learning) aims at analyzing and improving collaborative learning activities through digital tools. Collaboration has become especially prominent with the rise of learning theories such as Social Constructivism and has been found to be a key property of learning [1]. The research focus has therefore shifted from the individual to the group, as unit of analysis [2]. Researchers argue that the process of learning in groups becomes more explicit since individuals have to communicate intentions, knowledge and actions – which, in turn, allow researchers to capture parts of learning that would remain invisible if only the individual was studied [3]. However, groups also add complexity to investigate learning since they form complex systems in which individuals influence each other in various ways.

CSCL tries to address this by providing digital **tools** that help analyze and improve collaboration. Studies have proven superiority of digital tools over traditional means to support collaboration [4]. Nevertheless, detailing which functionality has which impact on collaboration has proven difficult [5]. Multiuser systems are especially hard to conceive since they have to take into account not only interactions between the system and a user but also interactions happening between users that may lead to conflicts [6]. Lately, new technologies have led to new possibilities of analysis and support of collaboration. Interactive tables for instance, while still rarely found in classroom settings, are one of the main device types used for collaborative research (figure 1).

Figure 1: Example of an interactive table with tangible tokens for collaboration

Proceedings of the Doctoral Consortium of the Seventeenth European Conference on Technology Enhanced Learning, September 12-16, 2022, Toulouse, France
EMAIL: Sebastian.simon@univ-lemans.fr (A. 1); iza.marfisi@univ-lemans.fr (A. 2); sebastien.george@univ-lemans.fr (A. 3)
ORCID: 0000-0003-3218-2032 (A. 1); 0000-0003-3218-2032 (A. 2); 0000-0003-0812-0712 (A. 3)

CSCL has seen numerous theoretical frameworks emerge on the nature of collaboration these past years. Indeed, CSCL is a cross-domain discipline drawing on concepts and theories from Psychology, Computer Science, Education and Sociology [7] and is in close neighborhood to CSCW (Computer Supported Collaborative Work). Consequently, overlapping concepts and varying views from researchers across disciplines have resulted in a variety of frameworks. Even the definition of collaboration itself is not unique and has evolved over time [8]. The first challenge is therefore to establish a **unified conceptual framework of collaboration**. If this challenge can be mastered, a second challenge would be to **identify links between collaborative processes and tool functionalities**. Indeed, even though studies have proven that digital tools can provide better assistance for collaboration then traditional means in many aspects [5], there is no clear link between activity, low-level functionalities and collaboration.

The objective of our Computer Science PhD thesis is to help overcome the aforementioned challenges. In the next section of this paper, we first present the related work on different aspects of collaboration. In section 3, we then propose a conceptual framework that combines key insights of previous work and provide an overall vision of collaboration on a process level (challenge 1). In section 4, we build on this conceptual framework to provide a tool framework in order to identify links between the high-level collaboration processes and the low-level functional parts of digital tools (challenge 2). In section 5, we present the work that has already been done during this first year of PhD and finally, in section 6, we present the upcoming work to validate our propositions.

2. Related work

High-level definitions of *collaboration* mainly diverge when it comes to *cooperation*. It is disputed whether cooperation should be a part of collaboration or a separate concept. In our work, we settle with the vision of Roschelle *et al*. [9] and consider collaboration a distinct concept from cooperation. Collaboration requires group members to act as one, while cooperation splits a task into smaller parts. It seems important to make this distinction between cooperation and collaboration in the context of collaborative learning since "acting as one" requires members to agree on their vision of the task, yielding group behavior patterns beneficial to learning not present in cooperative tasks. When both collaboration and cooperation occur, we group them under the concept of *collective activity* [10]. As an example, the activity of brainstorming is a collective activity since participants may split up the mental work of idea generation (cooperative activity) but organizing themselves involves joint planning and coordination (a collaborative activity).

In an attempt to detail the concept of collaboration further, two types of frameworks have emerged: on one hand, frameworks based on the notion of collaborative skills (*e.g.* [11]) and on the other hand, frameworks on the notion of collaborative processes (*e.g.* [4] [12]). We present three main frameworks and important work on peripheral concepts that will be the basis for our own proposition.

Meier *et al*. have identified five aspects of collaboration in their attempt of assessing the quality of computer supported collaboration processes [12]: *Communication, Joint Information Processing, Coordination, Interpersonal Relationship* and *Motivation* (figure 2).

Communication includes processes such as "grounding" to build a shared vision of concepts [13], *Joint information processing* refers to reaching consensus on decisions and processing available information collectively. To do so, members need to know what others know within the group and may use transactive memory systems [14]. *Coordination* concerns the organization of resources and monitoring critical subtask sequences while *interpersonal relationship* is characterized by Meier *et al*. by the absence of hierarchies where members have the same status, referring to Dillenbourg's notion of symmetrical relationships [15]. Finally, the *Motivation* category involves motivation by members to their individual contribution as well as to the group task result.

Collaborative Process Categories
Communication
Joint information processing
Coordination
Interpersonal relationship
Motivation

Figure 2: Five aspects of collaboration, colour coded for integration into our proposition Meier *et al.* (2007)

Mateescu *et al.* identified five dimensions of collaboration in their systematic review on collaborative studies [4]: *Workspace Awareness*, *Verbal and gestural communication*, *Participation*, *Coordination Flow*, *Artifact interaction* and *Level of Reasoning* (figure 3).

Workspace Awareness means understanding another person's interactions with the shared workspace. *Verbal and gestural communication* corresponds to the number of assertions, questions and answers. *Participation* is defined as a level of involvement by the participants in the problem solving process. Coordination flow embodies the strategies on how a group links or orchestrates individual contributions. *Artefact interaction* refers to the use of any object (e.g. tangible tokens). Finally, the *level of reasoning* is defined as "Measures that reflect the level of reasoning observed in or expressed by group members".

Collaborative Process Categories
Participation
Workspace awareness
Verbal and gestural communication
Coordination flow
Level of reasoning
Artefact interaction

Figure 3: Five dimensions of collaborative processes, colour coded for integration into our proposition, Mateescu *et al.* (2019)

Hesse *et al.* distinguish *conceptual skills* from *social skills* in their framework for teachable collaborative problem solving skills [11]. Social skills comprise *Participation*, *Perspective taking* and *Social regulation* whereas Conceptual skills concern *Learning and knowledge building* and *Task regulation*.

Hesse et al. describe participation skills as "observable action of engaging in discourse" and distinguish between action, interaction and task completion. *Perspective taking* is the capability to understand what other people think and know. *Social regulation* refers to the capacity of group members to be aware of and overcome biases (*e.g.* confirmation biases) so as to fully exploit the potential of the group's mental resources. *Task regulation* is a synonym for planning and coordination skills. *Learning and knowledge building* is a two-folded category in which knowledge building designate the "ability to take up ideas from collaborators to refine problem representations, plans, and monitoring activities" and learning as "the ability to identify and represent relationships, understand cause and effect, and develop hypotheses based on generalizations."

Collaborative Skill Categories
Cognitive skills
Task regulation
Learning & Knowledge building
Social skills
Participation
Perspective Taking
Social Regulation

Figure 4: Five collaboration skills, colour coded for integration into our proposition, Hesse *et al.* (2015)

Collaborative processes and skills are only a part of collaboration and how it emerges. As Dillenbourg notes, there is no guarantee collaborative learning will take place, but chances that it will occur can be increased by setting the right conditions [15]. The choice and design of activities are crucial to collaboration. The reason why collaboration is nothing natural is that it is not the most effective way to accomplish a task. Cooperation, in contrast, provides the advantage of task parallelization and a lower cognitive load per individual. Hierarchical structures further reduce cognitive load by limiting information spaces necessary for the execution of specialized subtasks. However, this intuitive *modus operandi* is counterproductive to learning since learning takes place in exchanges [1]. In order to make collaboration emerge in a team setting, Johnson & Johnson thus defined conditions for successful collaboration featuring *social skills*, *promotive interaction*, *positive interdependence*, *group processing* and *individual & group accountability* [14] (figure 5).

Social skills and *promotive interaction* refer to how individuals encourage and facilitate each other's efforts to complete tasks in order to reach the group's goals [16]. *Group processing* consists of multiple layers: self-reflection and regulation with respect to the needs and goals of the others in the group, co-reflection and regulation, and shared reflection and regulation (Kirshner *et al*). Such meta-cognitive skills require meta-cognitive evaluations: members must give feedback to each other and reflect on these to elicit which individual or group actions were helpful or unhelpful and to make decisions as to whether to continue or to change particular actions. *Positive Interdependence* links member of a team together so one cannot succeed unless all group members succeed [17]. This can be done for example through the design of the activity, by strategically providing knowledge for task completion among different members of a team. By doing so, members are constrained to collaborate and exchange. Finally, *group and individual accountability* in activities hold people responsible for their individual as well as the group performance. "When a person's performance affects the outcomes of collaborators, the person feels responsible for their welfare as well as his or her own (Matsui, Kakuyama, & Onglatco, 1987). Failing oneself is bad, but failing others as well is worse." [16]

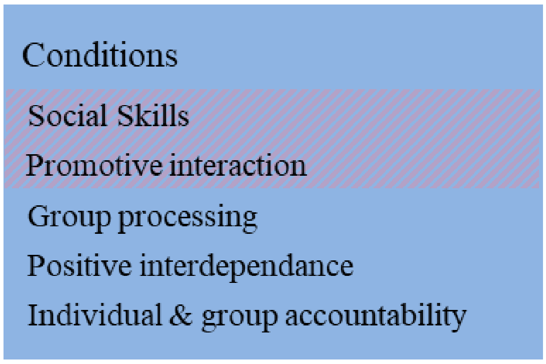

Figure 5: Necessary conditions for collaboration, color coded for integration into our proposition, Johnson & Johnson. (2004)

One last important concept related to collaboration is described in literature: cognitive artefacts. These are mental representations that help the group keep track of shared knowledge and a common representation of the task state. Since collaboration requires significantly more attention and cognitive resources than cooperation, groups organize and manage *transactive memory systems*. Such systems only require individuals to know what others know (meta-knowledge) to pool and process distributed knowledge within a group [18]. A *joint problem space* is established when members of a group successfully communicate a shared vision of the task or problem at hand. The notion of a joint problem space was first introduced by Roschelle et al. [19] .

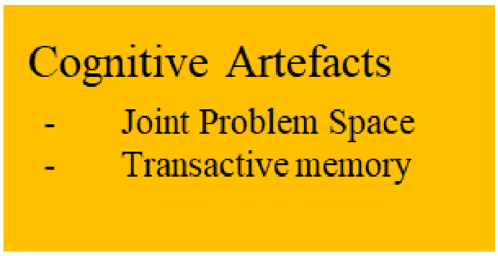

Figure 6: Cognitive Artefacts, Wegner, Roschelle *et al.* (1985, 1993)

3. **PhD thesis propositions**

As presented in the previous section, researchers have proposed various types and categories of collaborative processes, including related concepts such as skills, conditions and cognitive artefacts. The problem is, for the purpose of establishing links between processes and tools, to reunite these different visions under a common framework.

3.1 Conceptual framework

We attempt to provide a comprehensive conceptual framework that encompasses all of these views.

3.1.1 Process categories

We combine the collaborative process categories, proposed by Mateescu *et al*. [4] Hesse et al. [11] and Meier et al [12], into three main categories: ***Perception***, ***Participation*** and ***Coordination*** (the three categories are colored in shades of green throughout the presented frameworks in figure 2 – 4 and match our framework proposition in figure 7).

The **participation category** contains collaboration processes that Meier et al. grouped under communication and Mateescu *et al.* within *Verbal and Gestural communication*. We widen Hesse's definition of participation as an "observable action of engaging in discourse" into an observable action of engaging in communication. We further follow Hesse in his distinction of different levels of participative processes along actions, interactions and task completions. This category definition allows us to include processes considered by Mateescu *et al.* as *artefact interaction*. Examples of participative collaborative processes are grounding (the process of building a common vision by adapting individual knowledge to the other person's level of understanding), dialogue management, building on existing ideas, challenging arguments or managing transactive group memory (by creating and managing shared knowledge across group members).

The **awareness category** relates to knowledge about the environment, more specifically about cognitive awareness (what do I and other people know), behavioral awareness (what do other people do) and social awareness (emotional state of other group members [20]. As such, Hesse's social skill of *Perspective Taking* corresponds to a type of social awareness as well as Mateescu's workspace awareness to behavioral awareness in the presence of a shared tool. It also englobes Meier's interpersonal relationship category since it involves processes such as sensibility for hierarchical orders and potential conflicts that are a type of social awareness essential to maintain collaboration. Examples of awareness processes include self-evaluation (gaining awareness of personal strengths and weaknesses), pooling from transactive memory (gaining awareness of knowledge, strengths and weaknesses of others) or assuming responsibility for aspects of the activity itself. While those processes are not directly visible for an observer, they feed participative processes that reflect their presence within a group (such as taking part in an activity and informing others about its progress).

The **coordination category** relates to collaboration processes that coordinate how the task is resolved by the group. This category exists in all three frameworks (named task regulation in Hesse's framework). This category encompasses processes for resource management and planning (goal negotiation and expectations). Group processing is another important process which refers to the capability of a group to assess and evaluate their strategies for task completion and adapt them accordingly [21].

In addition, we propose to link several peripheral concepts to these three collaboration processes: conditions, skills and artefacts.

3.1.2 Preconditions, skills and cognitive artefacts

In order for collaborative processes to take place, we consider favorable **conditions**, such

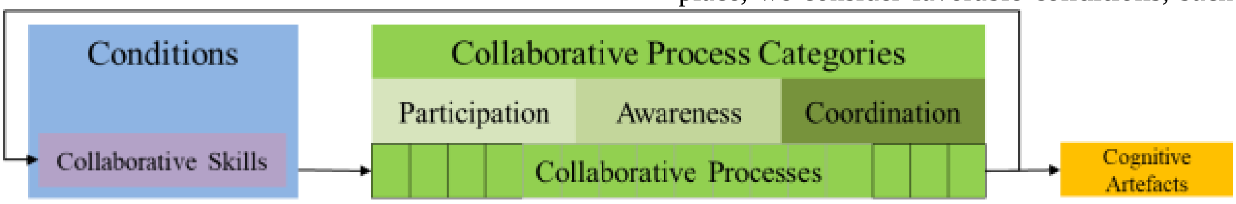

Figure 7: Proposition n°1: A Global Conceptual Framework

as format and design of the activity itself (providing rule sets to create forms of positive interdependence) and existing social and cognitive skills among team members.

In particular collaborative **Skills** can facilitate collaboration but can also be acquired and enhanced by engaging in collaboration, therefore being a reciprocal system in which processes act on skills and vice-versa.

Successful collaboration yields **cognitive artefacts** and group behavior patterns such as a joint problem space [9] (consisting of content and relational spaces [22]) and a shared group memory [18]. These cognitive artefacts can be detected and their quality measured for both analysis and tool support. This is the reason why Level of reasoning is colored in orange in figure 2: The use and quality of those cognitive artefacts allow us to assess the level of reasoning that participants deploy during collaboration.

In conclusion, the proposed conceptual model consists of a collaborative process hierarchy that groups different collaborative processes together under the following three main categories: *Participation, Awareness* and *Coordination*. When collaborative processes take place, they yield *cognitive artefacts* such as a joint problem space, shared group memory etc. For collaborative processes to take place, *preconditions* have to be met such as positive interdependence, accountability and, in particular, existing social *collaborative skills*.

3.2 Linking processes & tool functionalities

Previous studies on CSCL have mainly been concerned with providing evidence that digital tools provide advantages over more traditional means of collaboration, such as pen and paper. While this aspect is now widely accepted, studies are now starting to consider the impact of tools on the various collaboration processes. However, these tools are often composed of several functionalities, making it difficult to identify which of these functionalities, or a combination, is really supporting collaboration.

Prominent examples include Hwang et Su 2012: The study of surface computer supported cooperative work and its design efficiency and challenges, where a number of concepts such as territoriality and multiple gesture/action visualisations and have been condensed in a single tool. Caretta is another example of a tool that combines functionalities such as voting, shared and private screens, physical tokens, action visualisation and other functionality in one tool.

Having established a common framework on collaborative process level, the main question of our work is the following: *Can we link tool functionality to collaborative processes and if so, is there a combination that optimizes collaboration for a given activity and context?*

Investigating the potential existence of such links requires a notion of functionality that has the potential to be linked to one or more collaborative processes.

3.3 Functional bricks

We envision every tool to be a set of modular *functional bricks*, configured to work together. A functionality may be a shared mobile screen, or a widget to balance participation as demonstrated by Bachour *et al.* [23]. Another functional brick could be a shared mobile display to augment a static surface using a peephole approach. The tool presented in Figure 1, for example, has a functionality to filter the information presented on the shared screen and a functionality to interact with the screen by manipulating tangible tokens [24].

These functional bricks may directly impact certain collaborative processes or indirectly, by impacting related concepts. For example, a functional brick that manages positive resource interdependence helps at upholding conditions for collaboration. Another type of indirect functional bricks are those supporting cognitive artefacts, such as maintaining a joint problem space (*e.g.* by visualizing group findings).

A tool based on our framework is a mere aggregation of one or more functional bricks, each configured and orchestrated by a class of core bricks. The orchestration bricks allow for dynamic configuration of functional bricks included in the tool. Thereby, researchers can trigger the use of certain bricks at different moments of the experimentation or provide different groups with different functional bricks and information, effectively testing positive or negative impact of functional bricks (or variations thereof) on collaboration in an experimental manner (figure 8).

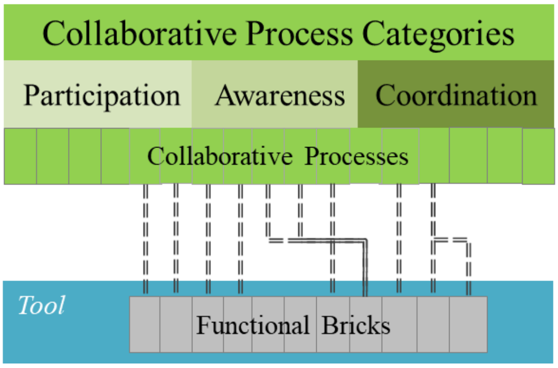

Figure 8: Investigating the link between tool functionality and the collaborative processes it supports

4. Conclusion and perspectives

During this first year of PhD, we have tried to form a comprehensive view of all the literature related to collaborative learning. We propose a conceptual framework that combines the important concepts and show the relations between them. In particular, this framework groups the collaboration processes into three main categories: participation, awareness and coordination. Our objective is now to build on this conceptual framework to identify links between the functional bricks, found in digital tools, and the collaborative processes they support. Understanding these links between functionalities and collaborative processes will be a significant breakthrough in CSCL because it will allow designers to implement only the necessary functionality to support the type of collaborative activity they want to create. However, there is still a long way to go before we can identify the effect of functional bricks.

To start with, we intend to analyze previous studies on collaborative tools. This will provide insight on the possible effects of the functional bricks on the collaborative processes. However, this will not be very precise, as systemic reviews are limited in depth and explanatory power due to heterogeneity of study parameters such as activity design, domain context, experimental parameters such as group size and composition but also tool design.

Ideally, more studies should be led with all the existing functionalities to help measure their impact on collaboration. Our intention is not to do this ourselves (which would be impossible within the given time of a PhD) but rather to provide a **framework** on which the community can build on. We also intend on providing an **open-source software architecture** to facilitate the implementation of these functional bricks and there orchestration. We plan on developing the core orchestration module and two functional bricks as a proof of concept. These functionalities and there combinations will be tested in 2023, during three experimentations planed in diverse contexts: a field trip in geography with master students, an orienteering race with disabled students in secondary school and a history-geography field trip with novice primary school teachers. The design of learning activities will be based on the MoCoGa model developed by Marfisi-Schottman *et al*. [26].

We believe that using a modular approach, under a common framework, allows for a better comparability and reproducibility of studies and strengthening identified links between functionalities and collaboration. In addition, developing tools takes up a significant amount of available resources. Sharing development efforts in a collaborative matter has the potential to liberate resources that can be used elsewhere. In the medium term, data and results from the scientific community using this framework for further experiments will validate modules and combinations that cannot be tested during this project and provide insights to enhance the interaction model that our experimentations will yield. In implementing the before mentioned methodology, we hope to also address the ongoing reproducibility crisis which is not exclusive to domains such as psychology or medicine [25].

While approaches like open data or pre-registrations can improve reproducibility, the variety of tools (and their limited availability for replication studies) used in CSCL make it near to impossible for other researchers to validate results. Not only may software not be available to other researchers but software is usually built for specific hardware (e.g. interactive tabletop), further limiting reproducibility and comparability. The latter is especially important in CSCL since study group sizes are small. Large size studies on situated collaborative learning are uncommon and thus, generalizing results is difficult [7].

5. Acknowledgements

The research published in this article was carried out for the SituLearn project, financed by the French *Agence National de la Recherche* (ANR-20-CE38-0012).

References

[1] J. Lave and E. Wenger, Situated Learning: Legitimate Peripheral Participation, 2nd ed. Cambridge University Press, Cambridge, UK, 1991.

[2] G. Stahl, T. Koschmann, and D. Suthers, 'Computer-supported collaborative learning: An historical perspective', Cambridge handbook of the learning sciences, p. 20, 2006.

[3] G. Stahl, Group Cognition: Computer Support for Building Collaborative Knowledge a book by Gerry Stahl. 2006.

[4] M. Mateescu, C. Pimmer, C. Zahn, D. Klinkhammer, and H. Reiterer, 'Collaboration on large interactive displays: a systematic review', Human-Computer Interaction, vol. 36, pp. 1–35, Dec. 2019, doi: 10.1080/07370024.2019.1697697.

[5] P. Dillenbourg, Collaborative Learning: Cognitive and Computational Approaches. Advances in Learning and Instruction Series. Elsevier Science, Inc, New York, NY, 1999.

[6] M. R. Morris, A. Cassanego, A. Paepcke, T. Winograd, A. M. Piper, and A. Huang, 'Mediating Group Dynamics through Tabletop Interface Design', IEEE Computer Graphics and Applications, vol. 26, no. 5, pp. 65–73, Sep. 2006, doi: 10.1109/MCG.2006.114.

[7] C. E. Hmelo-Silver, E. Chernobilsky, O. Mastov, C. Chinn, A. O'Donnell, and G. Erkens, 'Analyzing collaborative learning: Multiple approaches to understanding processes and outcomes', ICLS, 2006.

[8] L. Tong, 'Designing and Analyzing Collaborative Activities in Multi-Surface Environments', Ph.D. thesis, Université de Lyon, Lyon, 2017.

[9] J. Roschelle and S. D. Teasley, 'The Construction of Shared Knowledge in Collaborative Problem Solving', in Computer Supported Collaborative Learning, Berlin, Heidelberg, 1995, pp. 69–97. doi: 10.1007/978-3-642-85098-1_5.

[10] S. George, 'Apprentissage collectif à distance. SPLACH : un environnement informatique support d'une pédagogie de projet', Ph.D. thesis, Université du Maine, Le Mans, 2001.

[11] F. Hesse, E. Care, J. Buder, K. Sassenberg, and P. Griffin, 'A Framework for Teachable Collaborative Problem Solving Skills', 2015. doi: 10.1007/978-94-017-9395-7_2.

[12] A. Meier, H. Spada, and N. Rummel, 'A rating scheme for assessing the quality of computer-supported collaboration processes', International Journal of Computer-Supported Collaborative Learning, vol. 2, no. 1, pp. 63–86, 2007.

[13] H. H. Clark and S. E. Brennan, 'Grounding in communication', in Perspectives on socially shared cognition, American Psychological Association, Washington, DC, US, 1991, pp. 127–149. doi: 10.1037/10096-006.

[14] D. W. Johnson and R. T. Johnson, 'Cooperation and the Use of Technology', in Handbook of research on educational communications and technology, 2nd ed, Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US, 2004, pp. 785–811.

[15] P. Dillenbourg, 'What do you mean by collaborative learning', undefined, 1999, Accessed: Jul. 14, 2022. [Online]. Available: https://www.semanticscholar.org/paper/What-do-you-mean-by-collaborative-learning-Dillenbourg/9874b4cfd9e8ef89fd0b753af18c14cbc7c42744

[16] D. W. Johnson, 'Social Interdependence: Interrelationships Among Theory, Research, and Practice', American Psychologist, p. 12, 2003.

[17] D. Johnson, R. Johnson, and K. Smith, 'Active Learning: Cooperation in the College Classroom', The Annual Report of Educational Psychology in Japan, vol. 47, Jan. 1998, doi: 10.5926/arepj1962.47.0_29.

[18] B. Biancardi, L. Maisonnave-Couterou, M. Mancini, and G. Varni, ‘Tu sais qui sait quoi ? Suggestions pour l’étude du système de mémoire transactive dans un groupe à partir des patterns comportementaux et conversationnels.’, Saint Pierre d’Oléron, France, Jun. 2020. Accessed: May 24, 2022. [Online]. Available: https://hal.inria.fr/hal-02934517

[19] J. Smith, A. Disessa, and J. Roschelle, ‘Misconceptions Reconceived: A Constructivist Analysis of Knowledge in Transition’, Journal of the Learning Sciences, vol. 3, Nov. 1993, doi: 10.1207/s15327809jls0302_1.

[20] J. Fransen, P. A. Kirschner, and G. Erkens, ‘Mediating team effectiveness in the context of collaborative learning: The importance of team and task awareness’, Computers in Human Behavior, vol. 27, no. 3, pp. 1103–1113, May 2011, doi: 10.1016/j.chb.2010.05.017.

[21] D. W. Johnson, R. T. Johnson, M. B. Stanne, and A. Garibaldi, ‘Impact of Group Processing on Achievement in Cooperative Groups’, The Journal of Social Psychology, vol. 130, no. 4, pp. 507–516, Aug. 1990, doi: 10.1080/00224545.1990.9924613.

[22] B. Barron, ‘When Smart Groups Fail’, Journal of the Learning Sciences, vol. 12, no. 3, pp. 307–359, Jul. 2003, doi: 10.1207/S15327809JLS1203_1.

[23] K. Bachour, F. Kaplan, and P. Dillenbourg, ‘An Interactive Table for Supporting Participation Balance in Face-to-Face Collaborative Learning’, IEEE Trans. Learning Technol., vol. 3, no. 3, pp. 203–213, Jul. 2010, doi: 10.1109/TLT.2010.18.

[24] H.-C. Jetter, J. Gerken, M. Zöllner, H. Reiterer, and N. Milic-Frayling, ‘Materializing the query with facet-streams: a hybrid surface for collaborative search on tabletops’, in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, New York, NY, USA, mai 2011, pp. 3013–3022. doi: 10.1145/1978942.1979390.

[25] J. Janssen and I. Kollar, ‘Experimental and Quasi-Experimental Research in CSCL’, in International Handbook of Computer-Supported Collaborative Learning, U. Cress, C. Rosé, A. F. Wise, and J. Oshima, Eds. Springer International Publishing, Cham, 2021, pp. 497–515. doi: 10.1007/978-3-030-65291-3_27.

[26] I. Marfisi-Schottmann, A. Laine, and P. Laforcade, ‘Towards an Authoring Tool to Help Teachers Create Mobile Collaborative Learning Games for Field Trips. In: Proceedings of EC-TEL 2022, In press’, Toulouse, France, Jul. 2022.

The mediators of the effect of meaningful classroom digital technology integration on students' subject-specific learning outcomes in basic education

Doris K. Raave, Margus Pedaste and Katrin Saks

University of Tartu, Jakobi 5, Tartu, 51005, Estonia

Abstract

Despite the prominence of classroom digital technology integration (CDTI) in contemporary education, controversy remains on its effects on learning. Hence, previous research suggests concentrating not on the essentially transient effect of CDTI but rather on what mediates its effect on teaching-learning processes. The PhD study introduced in this paper aims to identify mediators of the effect of meaningful CDTI on students' subject-specific learning outcomes in basic education. For that, data were collected from 93 basic education teachers, 984 students, and their parents through interviews, in-class observations, tests, surveys, and questionnaires on CDTI practices, students' subject-specific and general competencies, and students' background information such as personality, mental capacity, school satisfaction, and relationship with teachers. Collected data are processed through clustering with cross-tabulation to identify teacher CDTI profiles, latent profile analysis to identify student subject-specific achievement profiles, and nested multi-group SEM analysis to detect possible mediators of CDTI's effect on student learning outcomes. The results help understand what mediates the effect of meaningful CDTI on students' subject-specific outcomes, which contributes to giving recommendations on how to personalise the teaching-learning processes. Stakeholders such as teachers, students, and developers benefit from this knowledge to plan, design, implement, evaluate, and reflect on meaningful CDTI.

Keywords 1

classroom digital technology integration, technology-mediated learning, basic education

1. Introduction

The use of digital technology in the teaching-learning processes is a salient feature of modern education, rendered more prominent by the COVID-19 pandemic. The pandemic embodied a disruption in diverse sections, including education. Many researchers in the field are hence spurred by sense-making of the changes derived from this disruption. As one example, the pandemic provided a chance for educational innovations that had been initiated but not completely implemented before, mainly regarding the use of digital technology [1].

For several years, the potential learning benefits of digital technology have been explored, leading to digital technology integration in education being encouraged by education policies [see, e.g., 2]. Making use of the learning affordances expects a meaningful use of digital technology, resulting in technology-enhanced teaching and learning. For example, the latter has been deconstructed as improvements in practicality, understanding and engagement [3].

Proceedings of the Doctoral Consortium Seventeenth European Conference on Technology Enhanced Learning, September 12–16, 2022, Toulouse, France
EMAIL: doris.raave@ut.ee (A. 1); margus.pedaste@ut.ee(A. 2); katrin.saks@ut.ee (A. 3)
ORCID: 0000-0003-4779-0006 (A. 1); 0000-0002-5087-9637 (A. 2); 0000-0002-4084-8765 (A. 3)

However, the term technology-enhanced learning includes an inherent bias [4, 5], an issue further so, as there is a noteworthy dispute between the researchers regarding the effectiveness of digital technology integration [see, e.g., 6, 7]. Regarding the main agents in determining the outcome, approaches to the field tend to be mainly divided into two, technology-led or pedagogy-led. The former invites compelling, rethinking and reevaluating pedagogical practices to incorporate technology's affordances [see, e.g., 8, 9, 10]. Opposing is the pedagogy-led approach, where pedagogical stances determine the technology integration [see, e.g., 11, 12, 13]; thus, pedagogy is seen as the main agent in determining the outcomes of the technology integration practices [14].

Recent research posits taking a step further from the technology-pedagogy dichotomy towards a consideration of entangled pedagogy [see 14], recognising that pedagogy and technology work in tandem not only with each other, but in interaction with the context and the different relations within these contexts on micro, meso, and macro levels, e.g., considering teachers, students, and the environment, such as the institution [4, 14]. Thus, the use of digital technology in education is regarded as "[…] complex, situated, and social in their constitution, their form, and their purpose, and as ungeneralisable in their effects as the choice of paintbrush is to the production of great art" [15], acknowledging that this recognition implies that the integration of technology into the teaching-learning processes may have different effects depending, for example, on the student [5].

Consequently, previous research suggests shifting focus from measuring the effects of digital technology integration on the teaching-learning processes, which are essentially transient, and concentrating rather on what mediates the effect of the technology integrations on these processes [5], considering and evaluating the relations of different elements [16]. Nevertheless, research alike is still scarce, possibly due to the complexity of the research design and process emanating from the numerous interacting and intertwined variables.

The PhD study described in this paper undertakes to gain insight into what affects the effect of technology-mediated learning on subject-specific learning outcomes aiming to identify some mediators of the effect of classroom digital technology integration. More specifically, we focus on i) understanding the practices of classroom digital technology integration in terms of how and why technology is integrated, ii) how these practices impact technology-mediated learning in basic education, and iii) what role do teacher, student and context-specific characteristics, such as attitudes, general competencies and personality, subject and institutional support, play in mediating the effect of classroom digital technology integration on subject-specific learning outcomes.

The context of the study is Estonia, where considering the effect of digital technology integration on learning outcomes and what affects the effect might be meaningful. The latter is due to two reasons; first, education in Estonia is considered one of the top-performing [17] and second, the use of digital technologies for learning and teaching is fairly widespread [18]. The latter is not only expected from the teachers [19], but teachers are also relatively well-prepared for it [20].

The following research questions thus guide the study:

RQ1: What are the teachers' classroom digital technology integration practices in Estonian basic education schools?

RQ2: What are the students' subject-specific learning outcomes in technology-mediated learning in Estonian basic education?

RQ3: What are the associations between classroom digital technology integration and subject-specific learning outcomes in Estonian basic education?

RQ4: What mediates the associations between classroom digital technology integration and subject-specific learning outcomes in Estonian basic education?

The aim of the described PhD study research is a contribution toward considering not only how technology affords learning and how to utilise these affordances to support pedagogical underpinnings but how to personalise education through evaluating the interactions of the technology, pedagogy, and the context.

2. Methodology

The PhD study described in this paper is a part of a larger research project, Digiefekt, running from May 2020 to April 2023. The

Human Research Ethics Committee of the University of Tartu, Estonia, approved the DigiEfekt project's research activities in December 2020 and again in September 2021 for a follow-up application that further developed the main study's plan in the light of the pilot studies' findings. The research project underwent two piloting studies to develop validated and reliable data collection instruments. The first piloting took place in April-May 2021, and the second piloting was in September-October 2021. The main study's data collection started in November 2021 and was completed in May 2022. The collected data will be analysed between June 2022 and March 2023. The main results of the project will be obtained by April 2023.

2.1. Sample

Purposeful sampling was used in the research project. We recruited schools with different profiles, considering different combinations of the following: i) academic achievement, ii) digital competence and iii) school satisfaction. More specifically, schools' performance was regarded in terms of students' achievement on academic tests. Digital competence was self-assessed by the teachers and the students. School satisfaction was reported in a survey conducted among teachers, students, and parents.

As participants, we selected Estonian, mathematics, and natural science teachers and their students from the end grades of each basic education level in Estonia, i.e., third (9–10 y/o), sixth grade (12–13 y/o), and ninth (15–16 y/o) grades. The participation of the schools, teachers and students was voluntary. The end sample consisted of 93 teachers and 984 students from 14 different schools across Estonia. Included were urban, suburban, and rural schools.

2.2. Data Collection and Analysis

To support the reader in following the research flow, the methodology will be described by the three sub-studies, which will make up the discussed PhD study. The first sub-study aims to identify teachers' classroom digital technology integration practices in their use and reasoning. To that end, data were collected by in-class observations to get an overview of how teachers integrate technology into the classroom. Further, interviews were conducted with the teachers to get an insight into the reasoning behind the specific use of digital technology. Following a content analysis of the collected data, a non-latent cluster analysis was conducted to identify profiles of teachers in terms of their digital technology integration practices. Moreover, data on teachers' background and demographics, e.g., age, years of service, self-efficacy, agency, and attitudes towards digital technology integration, were collected via questionnaires and will be used as control variables in cross-tabulations to support describing and explaining the identified clusters considering the relationships between the variables. Further, member checking will be conducted to validate the identified profiles.

For the second sub-study to identify students' profiles regarding subject-specific learning outcomes while also considering categorical latent variables, the following data were collected: students' results in digital competence and subject-specific competence tests, i.e., Estonian, mathematical and natural science competence, measured twice in the frame of one year, and students agency, learning anxiety and learning competence, measured once with self-report questionnaires in the frame of each subject, validated by in-class observations, as well as a test on students' mental capacity. These data will be analysed with latent profile analysis. Identified profiles will be further described and explained in the light of additional control data collected from and on students, such as students' socioemotional skills, personality and school satisfaction, analysed in cross-tabulations to explore relationships between the profiles and the control variables.

The third sub-study aims to discover associations between the profiles of teachers (profiling according to the classroom digital technology integration, identified in the first sub-study) and students (profiling according to the subject-specific learning outcomes, identified in the second sub-study) while considering the aforementioned background variables describing learners and teachers as well as a students' self-reported relationship with the teachers, and a nested multi-group SEM analysis will be conducted.

3. First Results

The data from the in-class observations on 167 lessons shows that digital technology is integrated into 82% of the lessons. These lessons included 269 different learning activities with the use of digital technology. In 59% of these activities, the technology was used only by the teachers. The activities used digital technology mainly as a substitute for a non-technological alternative, without making use of any functional improvement afforded by the technology (61% of the 269 activities). On approximately one-third of the occasions, technology was used for augmentation, relying on its affordances to provide functional improvements to the learning activities (34% of the activities). The rest of the 5% of the detected activities with digital technology integration made use of its affordances to revise and redesign the teaching-learning process (2% of the activities) or to adopt new teaching-learning practices, such as hybrid learning (3% of the activities) [see more 18].

Digital technology was integrated mainly to improve the practicality of the teaching and learning processes (42% of the activities), and the focus was more on facilitating teaching than learning. Besides, teachers adopted CDTI more commonly for its affordances to increase engagement (30% of the activities) than its affordance to facilitate deeper understanding (26% of the activities). In addition, teachers chose CDTI because they consider it more sustainable than non-technological alternatives (2% of the activities) [see more 21].

Regarding the teachers' classroom digital technology integration practices in terms of both the use and its purpose, we identified four profiles: introducers, facilitators, motivators, and deepeners. Introducers, facilitators, and motivators use technology mostly, although with different regularities, as a substitute, but the purposes for the substitution differ among these profiles.

More specifically, introducers integrate digital technology seldom to the classroom, and when doing so, there is no specific aspect of enhancement in mind. Facilitators stand out from the other profiles due to their main pedagogical reasonings for digital technology integration lying in its practicality and affordances to improve understanding, i.e., facilitating the teaching-learning processes for more in-depth learning. Conversely, Motivators focus mainly on digital technology's affordances to engage and motivate students.

The fourth identified profile, deepener, integrates technology both as a direct substitute and for the augmentation of the learning activities. Deepeners' aim in digital technology integrations is to facilitate understanding and augment learning gain. What lies behind these profiles, e.g., what would contribute to the sense-making and thus predict the occurrence of the specific profiles, is, however, still in the process of analysing.

4. Discussion and Conclusion

The PhD study described seeks to generate a holistic understanding of technology-mediated learning, i.e., what mediates the effect of technology integration on students learning, by adopting a relatively diverse and vast sample. Considered are the interactions between teachers' CDTI practices and students' subject-specific learning outcomes while scrutinising numerous student, teacher, and context-specific characteristics, acknowledging the agency of the stakeholders and environment in determining the effectiveness of the CDTI [5, 14, 15].

Although the results of this PhD study are still being processed, it emerged that in regard to teachers' classroom digital technology practices, student-centred objectives predominated over teacher-centred objectives, suggesting a focus on students in the pedagogical stances. These findings align with prior research showing a relationship between teachers' use of technology and the co-constructivist teaching approach, where learning is based on a conversation between teachers and students or peers [22].

Indeed, in the context where this study has been conducted, students are increasingly considered in the dialogue of creating educational experiences under the predominant paradigm of contemporary learning. In this paradigm, students are placed at the centre of learning design and instruction to scaffold their agency, as in the quickly changing, unpredictable environment, there is a need for autonomous, self-regulated learners [23]. Hence, lending to the aspirational digital technology integration, which is guided by the context and the combined purposes of the stakeholders [14].

The PhD study contributes to understanding how stakeholders and context interact in technology-mediated learning, which is necessary for planning, implementing, and reflecting on meaningful CDTI practices and supporting the personalisation of education. This study is done in the context of one country, having thus the predominant learning paradigm acting as a constant variable. Hence, similar research in diverse contexts would be desirable since the practised learning paradigm can be considered as one of the essential mediating factors to evaluate the effect of the CDTI.

5. Acknowledgements

The research project Digiefekt refers to the DIGIVARA5 project, formally titled "The Effect of Using Digital Learning Materials for Learning and Teaching in the Context of Estonian Basic School (1.05.2020–30.04.2023)," funded by the Estonian Ministry of Education and Research (EMER). The described PhD study is made possible by the collaboration of the research team of more than 20 researchers from the University of Tartu and the University of Belgrade, as well as the research assistants at the University of Tartu.

6. References

[1] Y. Zhao, J. Watterston, The changes we need: Education post COVID-19, Journal of Educational Change 22(1) (2021) 3–12.

[2] F. Gabriel, R. Marrone, Y. van Sebille, V. Kovanovic, M. de Laat, Digital education strategies around the world: practices and policies, Irish Educational Studies 41(1) (2022) 85–106. doi: 10.1007/s10833-021-09417-3

[3] A. Kirkwood, L. Price, Technology-enhanced learning and teaching in higher education: what is 'enhanced' and how do we know? A critical literature review, Learning, Media and Technology 39(1) (2014) 6–36. doi: 10.1080/17439884.2013.770404

[4] S. Bayne, What's the matter with 'technology-enhanced learning'?, Learning, Media and Technology 40(1) (2015) 5–20. doi: 10.1080/17439884.2014.915851

[5] M. Bower, Technology-mediated learning theory, British Journal of Educational Technology 50 (3) (2019) 1035–1048. doi: 10.1111/bjet.12771

[6] L. Daniela, D. Kalniņa, R. Strods, An overview on effectiveness of technology enhanced learning (TEL), International Journal of Knowledge Society Research 8(1) (2017) 79–91. doi: 10.4018/IJKSR.2017010105

[7] J. W. M. Lai, M. Bower, Evaluation of technology use in education: Findings from a critical analysis of systematic literature reviews, Journal of Computer Assisted Learning 36(3) (2020) 241–259. doi: 10.1111/jcal.12412

[8] M. Webb, M. Cox, A review of pedagogy related to information and communications technology, Technology, Pedagogy and Education 13(3) (2004) 235–286. doi: 10.1080/14759390400200183

[9] P. C. Beukman, A pedagogical framework for the adoption of Technology Enhanced Active Learning (TEAL) with the consideration of learning spaces in a Higher Education context, Ph.D. thesis, University of Portsmouth, Portsmouth, 2021.

[10] D. A. Cohen, Contextual issues of technology integration in teacher practice, Ph.D. thesis, RMIT University, Melbourne, 2020.

[11] J. Tondeur, J. Van Braak, P. A. Ertmer, A. Ottenbreit-Leftwich, Understanding the relationship between teachers' pedagogical beliefs and technology use in education: A systematic review of qualitative evidence, Educational technology research and development 65(3) (2017) 555–575. doi: 10.1007/s11423-016-9481-2

[12] D. Baneres, D. Whitelock, E. Ras, A. Karadeniz, A. E. Guerrero-Roldán, M. E. Rodríguez, Technology enhanced learning or learning driven by technology, International Journal of Educational Technology in Higher Education 16(5) (2019) 26–40.

[13] F. Aubrey-Smith, An exploration of the relationship between teachers' pedagogical stance and the use of ICT in their classroom practice, Ph.D. thesis, Open University, Milton Keynes, 2021.

[14] T. Fawns, An entangled pedagogy: Looking beyond the pedagogy—

technology dichotomy, Postdigital Science and Education (2022) 1–18. doi: 10.1007/s42438-022-00302-7
[15] J. Dron, Educational technology: What it is and how it works, AI & Society 37(1) (2022) 155–166. doi: 10.1007/s00146-021-01195-z
[16] P. Goodyear, L. Carvalho, The analysis of complex learning environments, in: H. Beetham, R. Sharpe (Eds.), Rethinking Pedagogy for a Digital Age, 3rd, Routledge, 2019, pp. 49–65.
[17] OECD, PISA 2018 results (volume V): Effective policies, successful schools, OECD Publishing, 2020.
[18] M. Pedaste, Ä. Leijen, K. Kallas, D. K. Raave, How to increase the potential of digital learning in achieving both cognitive and non-cognitive learning outcomes?, CO:RE Short Report Series on Key Topics, Hamburg (2022). doi: 10.21241/ssoar.79415
[19] M. Pedaste, Ä. Leijen, K. Poom-Valickis, E. Eisenschmidt, Teacher professional standards to support teacher quality in Estonia, European Journal of Education 54(3) (2019) 389–399. doi: 10.1111/ejed.12346
[20] M. L. Schmitz, C. Antonietti, A. Cattaneo, P. Gonon, D. Petko, When barriers are not an issue: Tracing the relationship between hindering factors and technology use in secondary schools across Europe, Computers & Education 179 (2022). doi: 10.1016/j.compedu.2021.104411
[21] D. K. Raave, E. Roldan Roa, M. Pedaste, K. Saks, Classroom digital technology integration – A double-edged sword? Engaging and practical yet harmful, in: International Conference on Innovative Technologies and Learning, Springer, Cham, 2022, pp. 241–251. doi: 10.1007/978-3-031-15273-3_27
[22] W. Admiraal, M. Louws, D. Lockhorst, T. Paas, M. Buynsters, A. Cviko, C. Jannsen, M. de Jonge, S. Nouwens, L. Post, F. van der Ven, L. Kester, Teachers in school-based technology innovations: A typology of their beliefs on teaching and technology, Computers & Education 114 (2017) 57–68. doi: 10.1016/j.compedu.2017.06.013
[23] C. Sum, I. Chan, H. Wong, Ready to learn in an uncertain future: Ways to support student engagement, Accounting Research Journal 34(2) (2021) 169–183. doi: 10.1108/ARJ-08-2020-0220

Tracking learners' knowledge and skills development

Andrea Zanellati[1,*], Maurizio Gabbrielli[1] and Olivia Levrini[2]

[1]Department of Computer Science and Engineering, University of Bologna, Mura Anteo Zamboni 7, Bologna, Italy
[2]Department of Physics and Astronomy, University of Bologna, Viale Berti Pichat 6/2, Bologna, Italy

Abstract
This Ph.D. research proposal aims to investigate how exploiting educational data to track the learners' development of knowledge and skills, thus embedding this information in automated tools designed to enhance teaching and learning. The encoding of learners' knowledge and skills is a crucial issue which can be exploited in addressing several tasks, such as underachievement prediction and personalized learning. However, some challenges characterized how to design the encoding and include it in automated tools: dealing with several formats of data (among which also text, video, images, and audio recording), tackling the strong dependence of educational data from the context where they are collected, and consider ethical issues related to explainability and fairness. With this position paper, we introduce the research questions which lead the project, a brief state of the art about techniques used to tackle the students' knowledge and skills encoding, the methodology and the expected results. Specifically, we aim to investigate which data can be used to fulfill our main purpose, test our encoding solutions in two case studies (underachievement prediction and knowledge tracing), and assess the contribution of our encoding to tackle them. As for the methodology, we want to explore strategies of Informed Machine Learning, that is to say incorporating an external knowledge source in the machine learning pipeline, which can improve the explainability and fairness of the models and handle the influence of the external context on the educational data.

Keywords
knowledge tracing, skill development, educational computing, informed machine learning

1. Introduction

This position paper presents the research proposal for a Ph.D. in Educational Data Science discussed at a Doctoral Consortium, with a specific focus on the problem of how to track the development of students' knowledge and skills. The paper was presented a year and a half after the start of the Ph.D. and contains the conceptual and motivational framework, the expected development steps, and a summary presentation of the preliminary results of the work done so far.

The paper is organized as follows. In the next section, we describe the background for the research, focusing on some challenges, motivating the research questions for the proposal, and the rational for our methodological choices. The third section is dedicated to the methodology. We introduce Informed Machine Learning (IML) as a reference methodological approach and we outline some attention for its application to the educational context. Section 4 is for the introduction of two case studies which mainly serve to exemplify the application of an IML approach. For the first case study we also briefly discuss some preliminary results; the second case study is presented as future work. Finally, we conclude with a description of the expected results and some final remarks.

2. Background

2.1. Datafication in the Educational field

In the last decade, the process of datafication in society has become increasingly pervasive, also affecting the educational field [1]. We assisted in a growing and varied interest in the application of artificial intelligence and data science techniques in this sector, with the rise of new research fields such as Learning Analytics (LA) and Educational Data Mining (EDM) [2]. Despite some differences, especially in the analysis techniques most commonly used by the two research communities, LA and EDM share the goal of extracting knowledge of interest for educational stakeholders –policy-makers, didactic coordinators, teachers, parents, and students– and using the extracted knowledge to improve the learning process in some way.

In this Ph.D. research proposal, in a broad perspective, we consider a key issue which is transversal to many educational situations: tracking the learners' development of knowledge and skills, thus embedding this informa-

Proceedings of the Doctoral Consortium of Seventeenth European Conference on Technology Enhanced Learning, September 12–16, 2022, Toulouse, France.
*Corresponding author.
andrea.zanellati2@unibo.it (A. Zanellati); maurizio.gabbrielli@unibo.it (M. Gabbrielli); olivia.levrini2@unibo.it (O. Levrini)
https://www.unibo.it/sitoweb/andrea.zanellati2/en (A. Zanellati); https://www.unibo.it/sitoweb/maurizio.gabbrielli/en (M. Gabbrielli); //www.unibo.it/sitoweb/olivia.levrini2/en (O. Levrini)
0000-0003-1632-3800 (A. Zanellati); 0000-0003-0609-8662 (M. Gabbrielli); 0000-0003-1632-3800 (O. Levrini)

tion in automated tools designed to enhance teaching and learning. There are several tasks which can benefit from the encoding of students' knowledge and skills development, e.g. low achievement prediction models [3, 4] or automated feedback system for personalized learning [5]. As main objective, we aim to tackle the problem of how encoding the students' knowledge and skills development, identifying valuable data resources for its representation, and testing our solutions effectiveness in addressing the tasks listed above. There are three main starting considerations which motivate our proposal and lead to design our research questions.

2.2. Three challenges to address for automated tracking of knowledge and skills development

Firstly, the datafication process in the educational field is characterized by several types of data, namely product data, process data, and background data. Product data are related to what students produced, and how they show their learning. They can be collected while students are learning, e.g. personal notes during classes, production of diagrams, and concept maps, questions answering, resolutions, and formative and summative assessments. Process data deals with how students are learning a specific content or how they behave during their performance assessment. The possibility to gain process data increased in the last years due to a spread in the use of digital technology in education, e.g Learning Management Systems (LMS), Massive Open Online Courses (MOOCs), and computer-based tests (CBt), both at school and university. This was furtherly accelerated by the recent COVID-19 pandemic. These technologies allow to track individual students' learning processes and collect data such as their mouse clicks, scrolling behavior, or time spent on different tasks or content resources. Also face-to-face classes allow the collection of process data, although it is often challenging and more time-consuming. An example is the data collected through the Think-Aloud protocol [6]. As for background data, they usually contain student demographics (parents' education, family income, household registration), curriculum plans, teachers' quality and style, and student performance evaluation. The previous list shows the variety of formats for educational data referable to a single learning activity: numerical, categorical and boolean variables are enriched by other formats e.g. texts, images, videos, and voice recording. This leads to the problem of how multimodal data fusion can be conducted in learning analytics [7, 8]. We refer to this challenge as the *multimodal data challenge*

A second issue concerns how these data are collected, organized, and labeled [9]. On the one hand, the increasing availability of educational data promotes the application of data science and machine learning techniques, to exploit data potential in enhancing learning and teaching. On the other hand, we have to consider that educational data are often highly context-dependent. There are few standardized large-scale educational datasets, i.e. data are very heterogeneous for different class groups. Furthermore, data labeling is not a common practice in classroom settings, because it is not one of the main objectives of teachers or other stakeholders traditionally involved in training processes. Moreover, educational data are often indicators of a learning competence or behavior, whose evaluation depends to some extent also on the evaluator. Let us consider, for example, the evaluation of task on creative thinking skills: different evaluators can result in different evaluation labels. Therefore, we are affirming that the collection of data on students cannot ignore the context in which this occurs and can hardly be considered free from pedagogical, psychological, and cognitive science theories, consolidated over years of research and assumed more or less explicitly by teachers, by those who design the context of learning or by who carries out the data collection. The research on Intelligent Tutoring Systems or Adaptive Educational Systems already considers the domain model and the pedagogical model together with the learner model [10, 11]. However, in these systems, they are often separate components, while we are suggesting that the domain model and the pedagogical model directly influence the learner model. This assumption relies on the issue already stated in the literature of the theory-ladenness in data-intensive approaches [12]. We can name this as the *theory-ladennes challenge*.

As a consequence, it is not easy to have robust and fair datasets on which to apply automatic data mining techniques, pointing out the third issue on ethics. An unbalanced or unrepresentative dataset may disadvantage students not sufficiently represented by the sample. The model –here intended as an automated detector for a commonly seen outcome or measure in LA and EDM, such as dropout, underachievement, affects, learning strategies, and disengaged behaviors– may be prone to overfitting the profile of well-represented students, resulting inflexible to new cases or changes that may occur in the school population. According to Baker [13], this is not just a technical challenge but it is a challenge for inclusion. In fact, a lot of the populations that we want to focus on, including historically underserved and underrepresented populations, are the ones it is harder to collect data for. This can be seen as a generalizability challenge for the models developed in LA and EDM. We can refer to this last point as the *ethical challenge*.

2.3. Research Questions

To sum up, the datafication process, which affected the educational field, is an opportunity to promote data-informed decisions for revising the learning designs and avoiding behaviors that lead to poor learning. In particular, one of the central problems is how to use data for the design of a learner model, an essential component for data-informed pedagogies and educational actions. In the development of this model, there are some challenges to be taken into consideration: the multimodal data challenge, the theory-ladennes challenge, and the problems of inclusiveness, fairness, and generalizability, summarized in the ethical challenge. The considerations in the previous subsection lead us to formulate the following research questions.

Most studies in this area have a purely or highly data-driven approach, which does not consider how context and several pedagogical assumptions can affect and be integrated into the machine-learning pipeline. This leads us to formulate the following research questions.

> **RQ1** How different educational data can be used for a reliable representation of learners' development of knowledge and skills?
>
> **RQ2** Is there any prior knowledge which can be integrated into AI tools used for tracking learners' development of knowledge and skills to improve their performance or their explainability?

The first research question is motivated by both multimodal data and ethical challenges, and also wants to suggest the need to reflect on what information is actually collected and expressed in the data. The second question emphasizes the need to consider other sources of information. The term *prior knowledge* here is intended in the perspective of Informed Machine Learning, chosen as methodological paradigm, that we describe in the next section. In our discussion, we can assume the domain and the pedagogical models as integrative knowledge source to data.

3. Methodology: Informed Machine Learning

According to von Rueden et al. [14] Informed Machine Learning describes "learning from a hybrid information source that consists of data and prior knowledge". It is not a purely data-driven approach due to the integration of an external and independent knowledge source into the machine learning pipeline.

With the term *knowledge* they assumed a computer science perspective, defining it as "validated information about relations between entities in certain contexts". There are three types of knowledge, several possible representations, and different forms of integration, as shown in Table 1. When dealing with the approach of informed machine learning in the educational field, the main source of prior knowledge to consider is the expert knowledge, often informal and validated through a group of experienced specialists. Also world knowledge could be a source of information to take into account, referring to facts from everyday life that are known to almost everyone, subsuming also linguistics.

Some forms of knowledge integration in LA models already exist; it almost occurs with the search for synergy with learning design, oriented to data-informed learning and teaching practice that preserve the agency of students and teachers [15, 16], overcoming purely data-driven approach. This way can be seen as integrating prior knowledge into the final step of the machine learning pipeline when its output is validated or benchmarked against existing knowledge through human mediation. However, there are other forms of knowledge integration in the machine learning pipeline –Training Data, Hypothesis Set, and Learning Algorithm– that could be investigated to face the challenge of the reconstruction of students' learning trajectories and students' competence development. In this research proposal we want to address the problem of developing knowledge and skills by investigating which supplementary knowledge sources can be used, how they can be represented and where they can be integrated into the machine learning pipeline (training data, hypothesis set, learning algorithm, and final hypothesis). To do this we consider two case studies, i.e two situations in which the problem of tracking the development of knowledge or skills is relevant and which we propose to approach from the perspective of informed machine learning. The first case study concerns a predictive model of underachievement and represents a study already started for which there are some preliminary results. In this first case, we present an example of feature engineering strongly driven by an explicit integration of a theoretical framework. The second case study concerns the problem of knowledge tracing. It represents a work direction still to be developed which also requires an in-depth analysis of what already exists in the literature as attempts at hybrid approaches in which a theory-laden is present. Therefore, we propose to investigate the RQs through two case studies that allow to use of different data (in the first case it is a static dataset and in the second dynamic) for learner modeling and to test prior knowledge integration strategies.

Table 1
Informed Machine Learning taxonomy introduced by von Rueden et. al. [14]

Source Which source of knowledge is integrated?	Representation How is the knowledge represented?	Integration Where is the knowledge integrated in the ML pipeline?
Scientific knowledge World knowledge Expert knowledge	Algebraic equations Differential equations Simulation results Spatial invariances Logic ruels Knowledge graphs Probabilistic relations Human feedback	Training data Hypothesis set Learning algorithms Final Hypothesis

4. Case studies

4.1. Low achievement prediction exploiting longitudinal large-scale assessment tests

4.1.1. Problem definition and State of the Art

Firstly, we examine data collected through national large-scale assessment tests. These tests are often used to support educational policy decisions [17] or in studies aiming to determine the relationship between socio-economic factors and school performances. Nevertheless, they are designed to measure students' knowledge and skills and often to track longitudinally the students' learning path [18]. These test design features enable the collection of data that can be useful for tracking the development of knowledge and skills and building predictive models for the risk of long-term *underachievement* or *dropout*. In [19], for example, the authors refer to data collected through the PISA international large-scale assessment tests to predict math proficiency.

Several machine learning techniques have been exploited to build predictive models for students' performance [4], including supervised learning, e.g., random forests, support vector machine and Bayesian network, unsupervised learning, e.g., k-means and hierarchical clustering, and recommender systems, e.g., collaborative filtering.

4.1.2. Specific objectives and outcomes

In [20] we present some preliminary results about maths low achievement prediction exploiting a very large italian dataset (more than 700000 students). Specifically, we exploit data collected through the INVALSI[1] large-scale assessment test to predict at grade 5 low achievement in math at the end of compulsory school at grade 10. We used three AI tools based on state-of-the-art machine learning models: random forest and two neural networks (categorical embedding neural network and feature tokenizer transformer). Finally, we presented a knowledge-based methods to encode students learning. Specifically, in the design of the learner model, we exploit features already present in the dataset regarding demographic information and the socio-cultural-economic context of the student, together with other features more related with the student's learning. This second set of features is obtained through engineering the boolean features that record the correctness of the student's responses to the individual items of the test. The new features are defined based on the framework used by INVALSI for classifying the items, in terms of areas, processes, and macro-processes. The rational for this choice is two-folds: firstly, this allows application to students from different cohorts who have taken different tests; secondly, they are directly related to students learning in terms of knowledge and skills, that are very important to design educational interventions to counteract the phenomenon of underachievement. The classification framework is shown in Table 2 This framework represents the source of integrative prior knowledge. Its representation is in the form of algebraic equations, with which we define the new features, i.e. for each student a correctness rate is computed for each area, process, or macro-process. The integration takes place into the train set.

Our results are summarized in table 3, which are promising. We aim to improve the research in three main directions. Firstly, we want to test the transferability to other disciplines such as Italian and English, which are tested by INVALSI, by using a similar representation or encoding for students learning. Secondly, we aim to improve the data quality by training and testing the model with students from different cohorts. This is possible by using at least four cohorts of students and may improve the transferability of the models to new cohorts. In fact, training the model on students' data from different school years could help in avoiding over-

[1]Italian National Institute for the Evaluation of the School System

Table 2
Maths INVALSI framework for question encoding.

Areas
(NU) Numbers (SF) Space and figures (DF) Data and forecasts (RF) Relations and functions
Process
(P1) Know and master the specific contents of mathematics (P2) Know and use algorithms and procedures (P3) Know different forms of representation and move from one to the other (P4) Solve problems using strategies in different fields (P5) Recognize the measurable nature of objects and phenomena in different contexts and measure quantities (P6) Progressively acquire typical forms of mathematical thought (P7) Use tools, models and representations in quantitative treatment information in the scientific, technological, economic and social fields (P8) Recognize shapes in space and use them for problem solving
Macro-process
(MP1) Formulating (MP2) Interpreting (MP3) Employing

fitting patterns to a specific test. Furthermore, we can try different student modeling approach, which is not driven by the Invalsi theoretical framework but which take into account other contextual information, e.g the items difficulty or the items embedding based on their texts. A last point of development concerns the interpretability which can be improved by comparing the feature importance analysis of the random forest model with the weights which define our neural networks.

To sum up, With this case study we want to investigate the potential of educational data collected through longitudinal large-scale assessment tests for the representation of the development of knowledge and skills, and look for other prior knowledge resources that can improve the performance of the model.

Table 3
Performance on test set

Models	Accuracy	Precision	Recall
Random Forest	0.77	0.62	0.67
CE neural network	0.76	0.76	0.76
FTT neural network	0.78	0.77	0.78

4.2. Knowledge tracing for personalized learning

4.2.1. Problem definition and State of the Art

As a second case study, let us consider an instructional unit provided through a learning management system (LMS). This is usual for MOOCs courses, it has also been the case for many students and teachers during COVID-19 pandemic [21] and potentially it may also be exploited in face-to-face classes, as a tool to organize teaching materials and manage different activities. As students work with the LMS they produce a wealth of data including product data (e.g. an explanation written in an electronic journal, or a video recorded through a mobile app) and process data (e.g. the number of edits made in the writing of this explanation, or log data). This data can be exploited for the well-known problem of *knowledge tracing* [22], which can be described as monitoring students' changing knowledge states during the learning process and accurately predicting their performance in future exercises. This information can be further applied to pursue *personalized learning* in order to maximize students' learning efficiency.

The most common machine learning techniques to handle knowledge tracing are Bayesian Network [22] and Dynamic Bayesian Network [23], to build probabilistic models. Another frequent approach is that of logistic models, such as learning factor analysis [24], performance factor analysis [25] and knowledge tracing machines [26]. In recent years, it has been explored also the use of deep neural networks [27], which outperform more traditional techniques, named Deep Knowledge Tracing (DKT).

4.2.2. Dataset and goals

For this case study we will consider data collected by the ALICE project (Learning Progression Analytics - Analyzing Learning for Individualized Competence development in mathematics and science Education), led by IPN Kiel, with the cooperation of DIPF Frankfurt and Ruhr-University Bochum. ALICE aims to exploit data from students' interactions with digital technologies in STEM –Science, Technology, Engineer, and Mathematics– classroom learning, both to predict the productivity of students' learning trajectories for their competence development and to identify underlying causes of unproductivity. The data is collected through the implementation of some instructional units in face-to-face classes using an LMS as a teaching aid. In this context, we want to investigate which useful prior knowledge related to ALICE educational context can be modeled and how. Furthermore, we aim to explore where they can be integrated in the ML pipeline to improve the learning trajectories analysis.

Our hypothesis is that the analysis of log data for the knowledge tracing can benefit from information on the face-to-face context, such as the choices of the teacher in the exposition of the unit contents and teaching times, relationships peer-to-peer, or the didactic model on which the unit itself is designed. We want to investigate the possibility of representing one or more of these sources of prior knowledge through a graph or a bayesian network, that can be used as input for a DKT network together with the log data collected on the student's interactions with the learning management system.

4.3. Remarks

In both cases, we refer to data collected about students' learning to build a *learner model*. However, we have to consider that the learning dynamics are strongly influenced by the *domain model* – understood as physical space, social-relational space, disciplinary space–, as well as by tutors/teachers and the *pedagogical model* they assumed. Domain model and pedagogical model may be considered as prior knowledge, here intended as a separate source with respect to data about students' learning behaviors or performances, which can be integrated into the machine learning pipeline, following the paradigm of Informed Machine Learning[14].

5. Expected results

On the one hand, the brief background presentation in the previous section demonstrates how crucial and transversal the proposed RQs are in the educational field. On the other, it highlights their complexity and the need to address them focusing on case studies, which may be very different from each other, although they share the need to identify suitable data to represent the learners' development of knowledge and skills. The comparative analysis of the results obtained in different case studies can bring out good practices or scalable solutions.

Therefore, in this research project, we aim to focus on different educational issues, referable to those presented in the previous section: underachievement and knowledge tracing. For each case study, we are going to identify or build a dataset useful for defining a learner model, intended as a representation of learners' development of knowledge and skills, thus contributing to RQ1. To handle RQ2, the proposed representations will be used to test several state-of-the-art solutions of machine learning, which can tackle the educational problem that motivates each case study. Furthermore, we aim to identify significant sources of prior knowledge (domain model and pedagogical model) and investigate how to integrate them into the machine learning tools. Hence we will evaluate their effectiveness, with respect to conventional machine learning solutions, by considering models' performances and their explainability, trying to come up with the main goal of RQ2.

References

[1] A. Klašnja-Milićević, M. Ivanović, Z. Budimac, Data science in education: Big data and learning analytics, Computer Applications in Engineering Education 25 (2017) 1066–1078.

[2] C. Romero, S. Ventura, Educational data mining and learning analytics: An updated survey, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 10 (2020) e1355.

[3] B. Albreiki, N. Zaki, H. Alashwal, A systematic literature review of student performance prediction using machine learning techniques, Education Sciences 11 (2021) 552.

[4] J. L. Rastrollo-Guerrero, J. A. Gomez-Pulido, A. Durán-Domínguez, Analyzing and predicting students' performance by means of machine learning: A review, Applied sciences 10 (2020) 1042.

[5] N. S. Raj, V. Renumol, A systematic literature review on adaptive content recommenders in personalized learning environments from 2015 to 2020, Journal of Computers in Education (2021) 1–36.

[6] R. Jääskeläinen, Think-aloud protocol, Handbook of translation studies 1 (2010) 371–374.

[7] S. Mu, M. Cui, X. Huang, Multimodal data fusion in learning analytics: A systematic review, Sensors 20 (2020) 6856.

[8] T. Baltrušaitis, C. Ahuja, L.-P. Morency, Multimodal machine learning: A survey and taxonomy, IEEE transactions on pattern analysis and machine intelligence 41 (2018) 423–443.

[9] W. Wang, G. Xu, W. Ding, Y. Huang, G. Li, J. Tang, Z. Liu, Representation learning from limited educational data with crowdsourced labels, IEEE Transactions on Knowledge and Data Engineering (2020).

[10] R. Nkambou, R. Mizoguchi, J. Bourdeau, Advances in intelligent tutoring systems, volume 308, Springer Science & Business Media, 2010.

[11] V. J. Shute, D. Zapata-Rivera, Adaptive educational systems, Adaptive technologies for training and education 7 (2012) 1–35.

[12] W. Pietsch, Aspects of theory-ladenness in data-intensive science, Philosophy of Science 82 (2015) 905–916.

[13] R. Baker, Challenges for the future of educational data mining: The baker learning analytics prizes (2019).

[14] L. von Rueden, S. Mayer, K. Beckh, B. Georgiev, S. Giesselbach, R. Heese, B. Kirsch, M. Walczak, J. Pfrommer, A. Pick, R. Ramamurthy, J. Garcke, C. Bauckhage, J. Schuecker, Informed machine learning - a taxonomy and survey of integrating prior knowledge into learning systems, IEEE Transactions on Knowledge and Data Engineering (2021) 1–1. doi:10.1109/TKDE.2021.3079836.

[15] M. Blumenstein, Synergies of learning analytics and learning design: A systematic review of student outcomes., Journal of Learning Analytics 7 (2020) 13–32.

[16] H. Shen, L. Liang, N. Law, E. Hemberg, U.-M. O'Reilly, Understanding learner behavior through learning design informed learning analytics, in: Proceedings of the Seventh ACM Conference on Learning@ Scale, 2020, pp. 135–145.

[17] G. E. Fischman, A. M. Topper, I. Silova, J. Goebel, J. L. Holloway, Examining the influence of international large-scale assessments on national education policies, Journal of education policy 34 (2019) 470–499.

[18] L. Branchetti, F. Ferretti, A. Lemmo, A. Maffia, F. Martignone, M. Matteucci, S. Mignani, A longitudinal analysis of the italian national standardized mathematics tests, in: CERME 9-Ninth Congress of the European Society for Research in Mathematics Education, 2015, pp. 1695–1701.

[19] A. Pejić, P. S. Molcer, K. Gulači, Math proficiency prediction in computer-based international large-scale assessments using a multi-class machine learning model, in: 2021 IEEE 19th International Symposium on Intelligent Systems and Informatics (SISY), IEEE, 2021, pp. 49–54.

[20] A. Zanellati, S. P. Zingaro, M. Gabbrielli, Student low achievement prediction, Lecture Notes in Computer Science 13355 (2022).

[21] Q. Liu, S. Shen, Z. Huang, E. Chen, Y. Zheng, A survey of knowledge tracing, arXiv preprint arXiv:2105.15106 (2021).

[22] A. T. Corbett, J. R. Anderson, Knowledge tracing: Modeling the acquisition of procedural knowledge, User modeling and user-adapted interaction 4 (1994) 253–278.

[23] T. Käser, S. Klingler, A. G. Schwing, M. Gross, Dynamic bayesian networks for student modeling, IEEE Transactions on Learning Technologies 10 (2017) 450–462.

[24] H. Cen, K. Koedinger, B. Junker, Learning factors analysis–a general method for cognitive model evaluation and improvement, in: International conference on intelligent tutoring systems, Springer, 2006, pp. 164–175.

[25] P. I. Pavlik Jr, H. Cen, K. R. Koedinger, Performance factors analysis–a new alternative to knowledge tracing., Online Submission (2009).

[26] J.-J. Vie, H. Kashima, Knowledge tracing machines: Factorization machines for knowledge tracing, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 33, 2019, pp. 750–757.

[27] C. Piech, J. Bassen, J. Huang, S. Ganguli, M. Sahami, L. J. Guibas, J. Sohl-Dickstein, Deep knowledge tracing, Advances in neural information processing systems 28 (2015).