

Single-case learning analytics: Feasibility of a human-centered analytics approach to support doctoral education

Luis P. Prieto

(Tallinn University, Tallinn, Estonia)

 <https://orcid.org/0000-0002-0057-0682>, lprian@tlu.ee

Gerti Pishtari

(University for Continuing Education, Krems, Austria)

 <https://orcid.org/0000-0001-9451-6881>, gerti.pishtari@donau-uni.ac.at

Yannis Dimitriadis

(Universidad de Valladolid, Valladolid, Spain)

 <https://orcid.org/0000-0001-7275-2242>, yannis@tel.uva.es

María Jesús Rodríguez-Triana

(Tallinn University, Tallinn, Estonia)

 <https://orcid.org/0000-0001-8639-1257>, mjrt@tlu.ee

Tobias Ley

(University for Continuing Education, Krems, Austria)

 <https://orcid.org/0000-0002-2804-2304>, tobias.ley@donau-uni.ac.at

Paula Odriozola-González

(Universidad de Cantabria, Santander, Spain)

 <https://orcid.org/0000-0001-8632-3174>, paula.odriozola@unican.es

Abstract: Recent advances in machine learning and natural language processing have the potential to transform human activity in many domains. The field of learning analytics has applied these techniques successfully to many areas of education but has not been able to permeate others, such as doctoral education. Indeed, doctoral education remains an under-researched area with widespread problems (high dropout rates, low mental well-being) and lacks technological support beyond very specialized tasks. The inherent uniqueness of the doctoral journey may help explain the lack of generalized solutions (technological or otherwise) to these challenges. We propose a novel approach to apply the aforementioned advances in computation to support doctoral education. Single-case learning analytics defines a process in which doctoral students, researchers, and computational elements collaborate to extract insights about a single (doctoral) learner's experience and learning process (e.g., contextual cues, behaviors and trends related to the doctoral student's sense of progress). The feasibility and added value of this approach are demonstrated using an authentic dataset collected by nine doctoral students over a period of at least two months. The insights from this feasibility study also serve to spark a research agenda for future technological support of doctoral education, which is aligned with recent calls for more human-centered approaches to designing and implementing learning analytics technologies.

Keywords: Technology-enhanced learning, Learning analytics, Human-centered learning analyt-

ics, Doctoral education, Human-AI teams, Design patterns, Analytics approaches

Categories: L.3.6, H.1.2, J.4, I.2.6, I.2.7

DOI: 10.3897/jucs.94067

1 Introduction

Recent advances in machine learning (ML), natural language processing (NLP) and, more generally, in the field of artificial intelligence (AI) have achieved generalized support for many tasks that were once the sole territory of humans (text classification, image captioning, and even image generation or multimodal reasoning [Fei et al., 2022]). For instance, what some authors call “foundation models”, pre-trained with huge amounts (i.e., web-scale) data, have the potential to transform human activity in many areas (e.g., law, healthcare, education) [Bommasani et al., 2021].

The widespread use of such AI technologies may pose dangers and challenges to human values and well-being in those areas [Bommasani et al., 2021], specifically in education [Blodgett and Madaio, 2021]. Indeed, such challenges have also been detected in the field of learning analytics (LA) [Siemens, 2012]: being oblivious to the privacy and ethics aspects of LA [Slade and Prinsloo, 2013] could lead to dystopic visions of our future becoming true, from 24/7 monitoring of our children to robots replacing teachers [Ferguson et al., 2016, Selwyn, 2019].

Aside from the body of work on ethics and LA [Tzimas and Demetriadis, 2021], recent years have also seen the rise of what has been termed “human-centered learning analytics” (HCLA) [Buckingham Shum et al., 2019] as a way to avoid such dystopias. HCLA’s focus has mainly been on design processes for LA technology using participatory and other user-involving design methods (e.g., [Martinez-Maldonado et al., 2015, Holstein et al., 2019]). However, human-centeredness *is* multi-faceted [Fitzpatrick, 2018], and not limited to *how* to design LA in a more humane manner. In the area of HCLA there has been comparatively little attention paid to *what* LA to design (i.e., what is the structure and features of an LA system that fosters human agency and well-being [Buckingham Shum et al., 2019]).

Against this technological backdrop, we should note that not all areas of education have been equally permeated by LA support (human-centered or not). Despite a greater focus of LA research on higher and continuing education (e.g., [Leony et al., 2015]), areas like doctoral education remain largely untouched by LA (see [Di Mitri et al., 2017] for a rare exception to this trend).

Indeed, even if it is acknowledged that doctoral education suffers from widespread problems like high dropout rates [Wollast et al., 2018, Hardre et al., 2019] and low student well-being [Evans et al., 2018, Milicev et al., 2021], support and interventions to address these problems (especially, technology-based) are rare [Mackie and Bates, 2019]. At the heart of these widespread problems and the scarcity of interventions is each doctoral process’s inherent uniqueness (i.e., the lack of meaningful cohorts from which to draw large enough amounts of data and meaningful comparisons), both topic-wise and in terms of personal and contextual factors.

This paper aims to address the lack of LA support for doctoral education from a human-centered perspective. More concretely, we focus on the unexplored area of *what* LA systems can be designed to support doctoral student agency and well-being. Our main contribution is to propose a structured process in which human (doctoral students, doctoral education researchers) and computational elements (both foundation and simpler, more interpretable ML models) collaborate to better understand each individual’s problems in

their everyday context and define personalized interventions to support doctoral progress and well-being.

Our “single-case learning analytics for doctoral education” (SCLA4DE) approach follows a process inspired by modern psychotherapy approaches [Hofmann, 2021] and uses both structured quantitative data and unstructured text data (currently, self-reported) – although the approach does not preclude other kinds of data. The approach also defines specific kinds of computational analyses of these mixed data for different tasks within the process. The main novelty of this LA proposal is that it is specifically tailored to problems (dropout, well-being) and particularities of doctoral education (e.g., the lack of meaningful cohorts).

To explore and validate the value of this approach to learning analytics for doctoral education, we follow an iterative design-based research (DBR) methodology [Wang and Hannafin, 2005]. As a first empirical iteration in such DBR process, the present paper explores the feasibility and added value of the SCLA4DE approach in a case study. A proof-of-concept of the main computational elements involved has been developed and tested in authentic settings with nine doctoral students of different disciplines and at different stages in their doctorate, in Spain and Estonia. The intervention, which lasted for two months in the Spring of 2020, focused on the students’ sense of progress as a key marker of doctoral well-being and eventual completion [Devos et al., 2017, Milicev et al., 2021], and gathered daily quantitative and qualitative self-report data about their well-being and progress, which was used to build both cohort-based and SCLA4DE models.

The paper is structured as follows: section 2 lays out the research on HCLA and the specific challenges of doctoral education and its (lack of) LA support. Section 3 describes the SCLA4DE approach to build LA systems, while section 4 details the context, methods and results of the aforementioned illustrative case study. The paper finishes with a discussion of lessons learned and limitations of the case study (section 5) and outlines a future research agenda for this new line of inquiry (section 6).

2 Related Work

2.1 Human-centered Learning Analytics

Research on human-centered learning analytics (HCLA) applies insights and processes from the area of human-computer interaction (HCI) –and, particularly, from human-centered design (HCD)– to the design, development and evaluation of learning analytics solutions [Buckingham Shum et al., 2019]. The emergence of this body of research mirrors recent calls for more human-centered artificial intelligence (AI) research [Shneiderman, 2020], and human-centered AI in education [Yang et al., 2021]. Similarly to these other areas, one main underlying goal of HCLA is to increase the adoption of LA by making the solutions more compatible with human values and capabilities [Kloos et al., 2022], while avoiding dystopic visions of LA application [Ferguson et al., 2019].

Although user involvement in both design and evaluation has been a feature of various LA research works (e.g., [Rodríguez-Triana et al., 2018]), most HCLA work so far has focused on the application of participatory and co-design methods (related to HCD) with the aim to increase stakeholder participation and eventually increase fitness and adoption of LA in different educational contexts. Examples of this dominant approach to HCLA include Martínez-Maldonado et al.’s “LATUX workflow” to design and deploy LA awareness tools [Martínez-Maldonado et al., 2015], which features iterative cycles

of prototype testing (of increasing fidelity) as well as classroom piloting (of increasing authenticity). Another classic example is Holstein and colleagues' co-design with teachers of Konscia, a wearable classroom awareness tool showing analytics for classrooms that use AI technology (in this case, intelligent tutoring systems) [Holstein et al., 2019], which followed a similar sequence of steps, adding LA-specific prototyping methods such as "teacher superpowers" or "replay enactments". Indeed, much of the focus of current HCLA research is on stakeholder (especially, teacher) engagement and participation methods, including long-term partnerships for the development of LA tools that fulfill local needs [Ahn et al., 2019], participatory semi-structured interviews [Dollinger et al., 2019] and workshops [Mavrikis et al., 2019, Prestigiacomo et al., 2020] or LA interface walkthroughs [Wise and Jung, 2019]. Other researchers have also tried to include students specifically in these LA design and deployment processes [Prieto-Alvarez et al., 2018].

However, human-centeredness is multifaceted and not *only* concerns design methods and shifts in users' power and control [Fitzpatrick, 2018], but also the impact (of LA, in this case) on working practices and ways in which values are built into the data models [Buckingham Shum, 2023]. Borrowing from the ethics guidelines for human-centered AI by the High-Level Expert Group on Artificial Intelligence (AI HLEG) [Weiser, 2019], we could define multiple dimensions of HCLA, from human agency and oversight, to transparency or societal and environmental well-being. More importantly, experts agree that these different dimensions should be *continuously evaluated and addressed throughout the (AI/LA) system's lifecycle* [Weiser, 2019].

As we can see, human (e.g., teacher or student) agency should be a central concern in HCLA. Several recent HCLA research proposals place a key importance in, e.g., teacher agency [Dimitriadis et al., 2021, Kloos et al., 2022] – especially, in the design phase of an LA system. Indeed, much of the HCLA work cited above has the intention of increasing teacher/student agency in an LA system by giving them a voice during the system design process. Yet, much less attention has been given to LA that aims to improve stakeholder agency *beyond* the LA solution design phase. Also, there is a scarcity of research work on patterns for how to structure HCLA solutions to maintain such agency, which is valid knowledge in HCI and other design disciplines [Buckingham Shum, 2023]. Dimitriadis and colleagues' provision of design guidelines [Dimitriadis et al., 2021] is one of the few exceptions to this gap, which the present paper also aims to overcome.

2.2 Doctoral education: Challenges and learning analytics support

Doctoral education (DE) suffers from several endemic problems that researchers have failed to find effective, generalizable interventions for (be them technology-mediated or otherwise). One of the most important ones is the high rate of dropout that many doctoral programs across the globe face [Wollast et al., 2018], which is an especially virulent version of the wider problem of dropout in higher education (HE) [European Commission. Directorate General for Education and Culture. et al., 2015, Wild and Schulze Heuling, 2020]. While a variety of educational and institutional factors have been found to relate to HE dropout (see, e.g., [De Silva et al., 2022]), the doctoral education literature specifically highlights the importance of personal and contextual factors, from candidate preparedness or marital status, to supervisor support [Bair and Haworth, 2004, Rigler et al., 2017, Wollast et al., 2018, Maher et al., 2020].

Another wicked problem at this educational level is the high prevalence of mental health issues like depression, stress or anxiety, which have led some authors to talk about a "mental health crisis" in graduate education [Evans et al., 2018]. DE research has uncovered a multitude of contextual factors influencing these problems, from economic

or family situations to daily habits and coping strategies, or the relationship with one's supervisor [Mackie and Bates, 2019, Sverdlik et al., 2018].

Recent research not only has determined the high prevalence of dropout and well-being problems in DE (ranging from 20% to 70% [Bair and Haworth, 2004, Terrell et al., 2012, Satinsky et al., 2021]). It has uncovered that these two problems not only overlap in their personal and contextual risk factors – they have also found that they are closely related to each other. An exploratory qualitative study found that low mental well-being was one of the key distinguishing factors of doctoral students that dropped out [Devos et al., 2017]. Indeed, a later multi-wave survey study of N=461 doctoral students in Belgium found mental health as a key antecedent of dropout intentions [De Clercq et al., 2021]. One's own perception of progress in the doctoral dissertation materials has also been found as a critical motivational factor affecting both emotional well-being and dropout in doctoral studies [Devos et al., 2017, De Clercq et al., 2021, Milicev et al., 2021]. This finding about the importance of one's perception of progress, which confirms earlier research in the more general area of knowledge work [Amabile and Kramer, 2011], suggests that progress (and practices to make it visible, like self-tracking [Avrahami et al., 2020]) could be an important construct to target in relation to those doctoral problems.

In contrast with the relative richness of research on factors and correlates of these problems of dropout and well-being, there exist few evidence-based *intervention* approaches to address them [Mackie and Bates, 2019, Jackman et al., 2021]. Indeed, the wide range of potential stressors in the doctoral environment and their complex interdependencies [Mackie and Bates, 2019] (again, pointing to the uniqueness of each doctoral student's challenges) makes the proposal of generalizable interventions difficult, and suggests targeting individual and contextual factors as a potential way forward [Jackman et al., 2021]. In relation to this, doctoral students' need to develop a sense of agency during the doctoral process [McAlpine and Amundsen, 2009] may be an important feature of such personalized interventions, as illustrated by initial work on coaching doctoral students [Godskesen and Kobayashi, 2016]. The importance of agency is also supported by findings in evidence-based psychotherapy, which highlight the person's own awareness of well-being problems and active role in the proposal of solutions that require behavior change [Strosahl et al., 2012, Hofmann, 2021]. There also exist initial work on training interventions focusing on the students' perception of progress (mentioned above), which have shown promising results [Prieto et al., 2022].

Even fewer doctoral interventions for well-being or dropout exploit the potential of technology. This is in line with a generalized lack of technology support to DE in general, aside from technologies used for specific research activities like data analysis or laboratory processes. Taking into account that well-being is one of the key dimensions of a human-centered LA system [Weiser, 2019, Kloos et al., 2022], it would seem especially fitting to use LA (and, especially, an HCLA approach/system) to address the well-being challenges of DE. However, there is a similar scarcity of research on LA that promotes well-being (at any educational level). Indeed, a recent literature review of this area found only six LA publications mentioning well-being [Hakami and Hernández Leo, 2020]. Most of these proposals addressed well-being only aspirationally (LA and educational institutions should care and improve well-being of students), or in very specific cases (e.g., accessibility of e-learning systems for disabled students). Further, none of these proposals focused specifically on doctoral education.

A notable exception applying LA to DE (which will help illustrate the challenges of doing so) is the work on Learning Pulse, an LA system that tried to model doctoral students' performance using multimodal data analytics within a self-regulated learning

perspective [Di Mitri et al., 2017]. Despite the length of data gathering in the study (8 weeks) and its comparatively fine granularity (every 5 minutes), the study's small sample size ($N=9$ students) highlights the difficulty of finding large cohorts of somewhat comparable doctoral students. In this work, a linear mixed-effects models trained on the whole "cohort" tried to provide personalized predictions but, according to authors themselves, the goodness of fit (measured via R-squared of the models) was limited (0.05-0.32 for different target performance variables), leading to "poor prediction accuracies" (p. 196). Among the potential causes cited by the authors are the data sources used, the variability of learner contexts, and the lack of a clearly defined learning task.

To the best of our knowledge, there has been no other LA-related effort in DE. This scarcity can be traced back to the unstructured nature and uniqueness of doctoral processes (and individual differences in personality, culture, or skill-sets of students), along with a lack of large datasets (and the fact that learning management systems are not typically used in DE), which makes it hard to build reliable AI or machine learning models (on which most LA solutions rely). This is precisely the set of challenges that our proposed HCLA approach (see next section) intends to tackle.

3 Single-case learning analytics for doctoral education as an HCLA approach

To tackle the challenges of supporting doctoral education with learning analytics illustrated above (i.e., the uniqueness of each doctoral student and learning process, the lack of meaningful cohorts on which to perform comparisons or modeling, and the small size of typical datasets), we turn to recent advances in evidence-based psychotherapy, such as third-wave cognitive-behavioral therapies [Hayes and Hofmann, 2021]. These approaches look at well-being problems of humans idiographically (i.e., focusing on the uniqueness of individuals and their experience) and in-context, engaging the person in a collaborative relationship and as an active participant to find concrete solutions to their mental health problems [Hofmann, 2021]. This collaboration is often aided by active data gathering by the participants themselves (e.g., [Fernández and Mairal, 2017, Sellés et al., 2015]). We hypothesize that computational models could help participants in unearthing and understanding contextual cues related to problems they experience in their unique everyday work. In the context of such collaboration between the student and an expert (be it a counselor/coach, supervisor or researcher), we hereby propose a new (human-centered) LA approach, a pattern for structuring LA solutions: single-case learning analytics for doctoral education (SCLA4DE).

SCLA4DE defines a structured process involving both human stakeholders and computational elements with specific missions, to provide personalized support to doctoral learners (see Figure 1). The SCLA4DE approach requires a toolbox of computational elements specifically designed to address some of the challenges and particularities of supporting DE. We can thus see SCLA4DE as fitting the philosophy (and providing a concrete instantiation) of human-AI teams [Wilder et al., 2020] (i.e., socio-technical systems specifically designed considering the distinct abilities of people and machines), which have recently been proposed as relevant to the field of education as well [Molenaar, 2021].

The starting point for the SCLA4DE process is the detection of a problem (e.g., lack of progress in the dissertation), oftentimes by the doctoral student themself (as "progress" is often subjectively evaluated on an everyday basis by students, see [Devos et al., 2017]), or flagged by other actors (e.g., a doctoral supervisor). An SCLA4DE-based solution

would thus be a socio-technical system that instantiates a process with several distinctive phases and activities performed by different kinds of human or computational actors (*italics denote the different process steps in Figure 1*) to help understand and find solutions to such a problem:

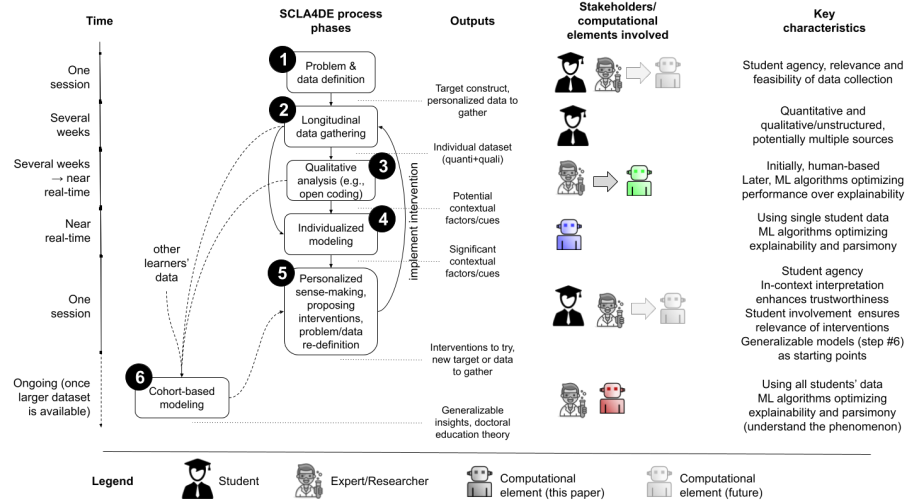


Figure 1: SCLA4DE process phases, their outputs, and main computational elements and stakeholders involved. Semi-transparent icons indicate stakeholder roles that could be supported or automated in the future

1. A *Problem and Data Definition* phase in which individual doctoral students are included in defining what exactly is the problem that the SCLA4DE system tries to help with (e.g., lack of perceived progress on thesis materials, cf. [Devos et al., 2017]). The researcher and the student also define the data that will be collected and how (is it self-report questionnaires, wearable devices, etc.), using the student's contextual knowledge of what data are feasible to collect, and what variables may be most relevant to track over time (including ones not anticipated by the researcher). This step also gives voice to the students and their concerns, thus increasing student agency in the deployment phase of the system.
2. A phase of *Longitudinal Data Gathering*, akin to a repeated measures design (i.e., data are gathered on more than one occasion) [Ellis, 1999]. This data gathering may occur across multiple data sources (i.e., would lead to multimodal learning analytics [Ochoa, 2017]) and would include both quantitative, structured variables (e.g., number of hours slept in a particular day) and unstructured variables more amenable to qualitative analysis (e.g., an open-text diary entry about the events of that day). Although this data collection can happen automatically (as with the wearable devices in the Learning Pulse study [Di Mitri et al., 2017]), it is likely that doctoral students will also have a role as active data collectors, and other human stakeholders like doctoral supervisors may also be involved in such data collection.

This collection of multiple time series variables aims to achieve enough data volume for LA modeling (even if we consider only one student), but it also will enable temporal analysis of the variables of interest (an under-utilized kind of analysis in the LA field [Knight et al., 2017]). Indeed, such active data collection by students can be already considered an intervention, as many psychotherapy approaches use self-tracking as a tool to focus individuals' attention to relevant, useful (or unhelpful) behaviors and phenomena [Fernández and Mairal, 2017, Sellés et al., 2015]. In terms of self-regulated learning, such active data collection can be considered a self-monitoring device, which is critical for learning and problem-solving [Zimmerman, 1990]. In addition, the collection of unstructured data by the students (e.g., in the form of open reflections or narratives) can be seen both as a generative activity (known to improve self-monitoring and self-regulation of learning [van Gog et al., 2020]) and giving additional voice to doctoral students during the operation of the system, and exploits their unique knowledge of the particular context and learning process, again increasing student agency in the system's operation.

3. Once a large-enough dataset has been gathered, a *Qualitative Analysis* phase ensues, initially performed manually by human researchers. This step mainly requires a (open/inductive or theory-based/deductive) coding of the unstructured data collected by the students over time, and is intended to adapt this part of the free-form dataset to a format more amenable to computational processing [Shaffer, 2017]. This step enables the SCLA4DE system to include in its analyses/modeling *new* factors and variables that the students have noticed, and which were not in the original data collection design (again, giving them a voice and agency in the LA process). The amount of data needed to perform this qualitative analysis and obtain useful insights is *a priori* unknown (and it may vary depending on the length of the student's narrative entries), although our current hypothesis is that a few weeks of data would be needed for the human coders to understand the context enough and to observe noticeable trends. Another issue of practical importance is the time needed to perform such analyses, as qualitative analysis of even a moderately-sized dataset can take weeks or months and insights derived from it may come in too late for the student originating the data. Thus, computational elements (either based on latest advances in ML/NLP [Pishtari, 2021] or other means like regular expressions [Cai et al., 2019]) will be needed to automate this coding process for near real-time extraction of timely insights for students.
4. Using both the quantitative time series and the data coming from the coding in the previous phase, *Individualized Modeling* is performed using only one individual learner's data. These models would thus include not only quantitative variables defined by the researchers in the initial problem definition phase, but also student-defined quantitative variables (also from the problem definition) as well as variables extracted from the students' unstructured data (i.e., coded by researchers) during the system's operation. These models would *not* be generalizable to other doctoral students – but that is not their aim, as they are intended only for in-context application by the student that originated the data (see [Shaffer and Serlin, 2004] for more on this issue). Furthermore, this individualized modeling should be performed using explainable, parsimonious models (as opposed to very complex, black-box models), as they are intended for use by students themselves (see following steps). Both the use of student-generated variables and explainable models are aimed at improving the agency of students in terms of deriving applicable insights from the analytics.

5. The individualized models from the previous phase are then used in a session of *Personalized Sense-Making*, in which those models are visualized and discussed with the aid of researchers or other stakeholders (e.g., the student's supervisor or a coach). In these (potentially periodic) meetings the stakeholders can *Propose New Interventions* (e.g., new behaviors to implement in order to improve the student outcomes of interest) and/or go through a phase of *Problem (or Data) Redefinition* for the next phase of system usage (hence, going back to step #2 above). To aid in this sense-making and proposal of interventions, experts/researchers can draw upon existing doctoral education literature (e.g., suggest strategies to manage the student's advisor, if their relationship seems to be an issue [Grover, 2007], or specific diagramming and discussion exercises if dissertation topic is unclear [Prieto et al., 2022]), or more general methodologies for well-being-related behavior change (e.g., identifying unhelpful behaviors used to procrastinate and more values-aligned alternative behaviors [Ruiz et al., 2016], or uncovering painful emotions that arise during paper writing and suggesting strategies to manage such emotions [Strosahl et al., 2012]). This step is aimed to increase student agency both in terms of interpretability of results (by including experts/researchers) and definition of the problem or the analytics.
6. In parallel to the workflow defined in steps #2–#5 above, once enough data are collected from multiple doctoral students, researchers can also develop *Cohort-based Models* using those students' data, so as to derive generalizable insights about the doctoral education experience and related phenomena (e.g., related to the problems of student dropout and well-being) – as is typical in traditional LA (and is the common goal of most researchers) –. This final phase also has the advantage of giving a voice to students in the generation of wider theories about doctoral education.

It is worth pointing out that different kinds of well-being and persistence problems may appear at different stages of the doctorate (as suggested by [Sverdlik and Hall, 2020]) and, consequently, different kinds of solutions/interventions may make sense at different stages (see [Grover, 2007, Ali and Gregg Kohun, 2007, J. Pifer and L. Baker, 2016] for examples in the doctoral education literature). However, the psychotherapeutic inspiration of our approach (which suggests that there exist common processes to act upon which are independent of the particular problem or diagnosis [Hayes and Hofmann, 2021]) and the SCLA4DE approach's emphasis on gathering unstructured and student-defined data, are precisely chosen to make it flexible enough to address these different kinds of challenges. It is also worth emphasizing SCLA4DE's requirement of *active* participation by doctoral students in data collection and interpretation. This requirement, also stemming from the psychotherapeutic origins of the approach (described above) and our own human-centered stance on learning analytics, could be seen as a limitation of the approach, as it requires a certain amount of time and effort (e.g., to report certain quantitative indicators and write down a short narrative entry, every day) by doctoral students, who are often time-deprived. This also limits the scope of application of the approach to problems of which the student is aware, and for which the student is willing to engage in the process described above). We argue, however, that active data collection is also a strength, which serves to circumvent the ethical concerns that gathering data from unaware students may entail. This issue of active participation nevertheless raises the question of whether it is feasible for doctoral students to engage in this process for extended periods of their everyday lives, or whether the amounts of data that they are

able to generate (e.g., length of the diary entries) is of sufficient value – thus requiring a feasibility study as a first step (see the next section).

The process outlined above defines at least three key computational elements (denoted by colored bot icons in Figure 1): 1) personalized ML/LA models of the phenomena of interest (blue bot); 2) more generalizable LA/ML models of DE phenomena across multiple students (red bot); and 3) computational means of reproducing human coding of unstructured data (green bot). It also defines several key stakeholders involved in the process (for now, mainly doctoral students and experts/researchers). Yet, it should be noted that further computational elements could be developed in the future to (partially or completely) automate some of the human stakeholder activities outlined above (see section 6).

This SCLA4DE approach should be understood as a human-computer interaction contribution, within an HCLA paradigm [Buckingham Shum, 2023]. Our approach provides a new design pattern to solve a common problem (in this case, specific problems of doctoral education like dropout and low well-being) and a description of value-adding technical capabilities [Zimmerman and Forlizzi, 2014]. To further illustrate the added value of implementing LA solutions using this approach, the next section describes an illustrative study in which initial prototypes of the aforementioned three computational elements were implemented and applied to an authentic dataset collected by doctoral students in two different countries (Estonia and Spain).

4 Illustrating SCLA4DE's computational elements using an authentic dataset

4.1 Goal and research questions

Our overall goal while proposing the SCLA4DE approach described above is to design and develop socio-technical interventions to support doctoral well-being and address problems of doctoral dropout. Within this overarching goal, the case study described in the rest of this section aimed at exploring the feasibility and usefulness of the three main computational elements of the SCLA4DE approach mentioned above (see the colored bots in Figure 1). We thus formulated three research questions (and several sub-questions, see Figure 2), each mapping directly to the intended outcomes of the computational elements defined in the SCLA4DE approach above:

- RQ1 (related to the red bot, in Figure 1): What across-learners insights can be derived from SCLA data?
 - RQ1.1 ... from the quantitative time series data?
 - RQ1.2 ... from adding qualitative data on top of the quantitative ones?
- RQ2 (blue bot): What is the added value of a single-learner analysis over the cohort-based one?
 - RQ2.1 ... in terms of (subjective) insight?
 - RQ2.2 ... in terms of (objective) model performance?
- RQ3 (green bot) Can computational techniques reproduce researcher-generated open coding reliably?

- RQ3.1 What computational technique produces most reliable results?
- RQ3.2 What factors influence model performance?

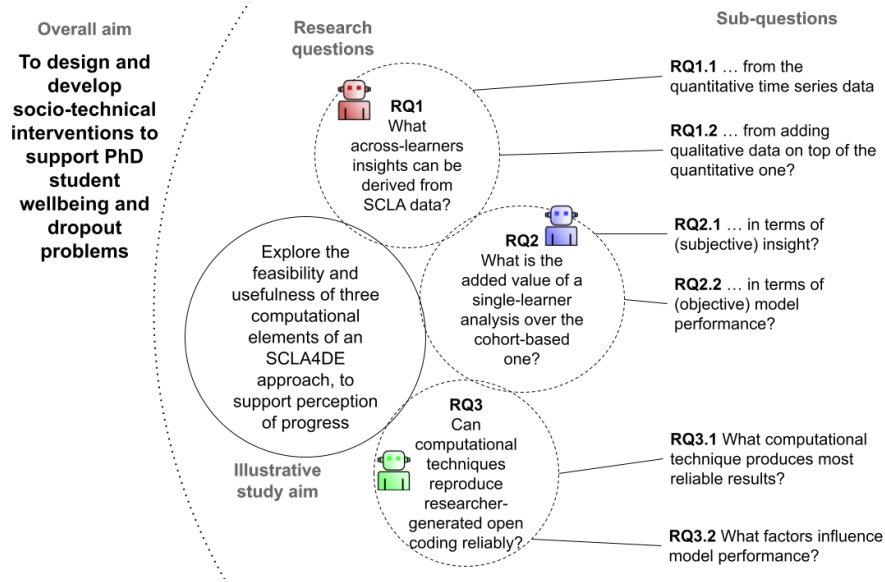


Figure 2: Research questions and sub-questions tackled by the illustrative case study

To provide an illustrative response to these research questions, we gathered an authentic SCLA4DE dataset using a process similar to the one presented in Figure 1. The next subsections describe where, when and how that dataset was gathered and analyzed.

4.2 Context and participants

Context. Progress is known as a critical motivational factor affecting both emotional well-being and dropouts of doctoral studies (see section 2.2). Taking such research into account, and the usefulness of self-tracking and diaries in many therapies for behavior change (see section 3), we developed a simple technology prototype (a web-based self-tracking and diary platform, see Figure 3) to gather personalized self-report data. Ethical approval for conducting the study was then obtained from the Ethics Committee at the Center of Excellence in Educational Innovation at Tallinn University (Estonia).

Participants. Volunteer doctoral students were recruited from the participants of previous progress-oriented training actions held at two public universities in Estonia and Spain [Prieto et al., 2022]. This ensured that all participants had a basic understanding of the well-being and dropout problems in doctoral education, the critical role of perceiving one's progress, and practices that could foster it. N=9 doctoral students (all female) from multiple disciplinary backgrounds (see Table 1) provided informed and voluntary consent to participate in the study by providing daily self-tracking/diary inputs. All data was introduced using an anonymous numerical code, to preserve participants' anonymity.

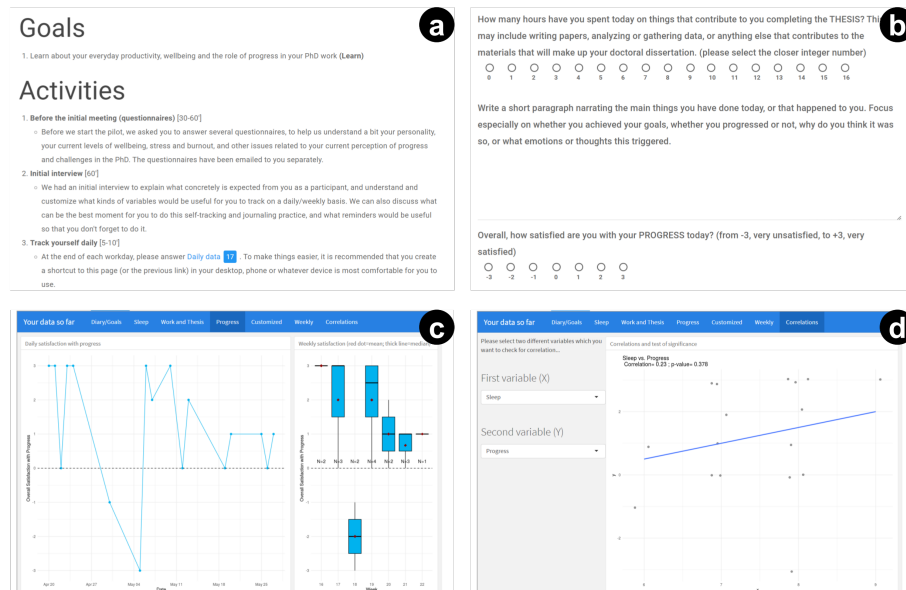


Figure 3: Screenshots of the web-based platform used by doctoral students to gather and visualize data: a) Description of the activities for students; b) Questionnaire for self-reporting of daily data; c) Dashboard visualizing a student's self-reported sense of progress (daily and aggregated per week); d) Dashboard visualizing a simple correlation model between two data variables

4.3 Methods

Dataset generation. The aforementioned web platform used for self-report data entry was configured so that each daily entry had several default quantitative indicators (see Figure 3a,b): hours slept, time spent working on the thesis materials, time working in other topics, and a self-assessment of the progress that day (a Likert scale from -3, very unsatisfactory, to +3, very satisfactory). The entry also included an open question asking students to narrate the main events of the day, especially regarding progress and goal achievement and emotions/thoughts this elicited. All doctoral students were interviewed (similar to step #1 in the SCLA4DE process, see section 3) before using the aforementioned prototype self-tracking and diary platform, to become familiar with the platform and to customize their data gathering by adding 1-5 additional quantitative indicators they themselves came up with as potentially relevant for their progress (see section 4.4 for concrete examples). The doctoral students then used the aforementioned web platform to self-track their progress and the aforementioned variables, for a period of 4-8 weeks in the Spring of 2020 (albeit they were free to decide not to log an entry for a day, or they could forget to do so). After the data input phase, another interview was done with students to show them their data and to discuss it (akin to step #5 in Figure 1). Our final dataset contained a grand total of 270 tracking and diary entries. Over this period of time, each participant logged between 14 and 51 entries (see Table 1 for further details about the data logged). Midway through this logging period, basic data

Name (fictional)	Country	Research discipline	Year into the PhD	Gender	Diary/ self-tracking entries	Mean(SD) words
Alice	ES	SOC	2	F	51	24(17)
Beth	ES	HEA	4	F	45	44(25)
Chris	EE	TEC	3	F	23	32(16)
Debbie	EE	TEC	2	F	17	40(35)
Emma	ES	HUM	2	F	46	40(26)
Fiona	ES	HEA	2	F	36	39(20)
Gloria	ES	TEC	1	F	20	75(40)
Heather	EE	SOC	4	F	18	15(21)
Irene	EE	HUM	3	F	14	5(8)

Table 1: Demographic profile and data gathered by the 9 doctoral students participating in the illustrative study. The last column provides the word count distribution of students' daily narrative (journaling) entries. ES=Spain, EE=Estonia. SOC=Social sciences, HEA=Health sciences, TEC=Technology, Architecture & Engineering, HUM=Humanities. F=Female.

visualizations of their own data and simple data models (e.g., correlations between two logged variables) were made available to the students (see Figure 3c,d).

Data analysis. In order to explore the answers to RQ1 (cohort analyses of SCLA-like data), we trained two kinds of exploratory, interpretable ML models. To understand the across-learner value of the quantitative data in the datasets (RQ1.1), a Gaussian graphical model (GGM) [Epskamp et al., 2018] has been created using the quantitative time series data provided by doctoral students (to understand temporal, contemporaneous and across-learner relationships between them). To further extract across-learner insights about what variables seem to be related with better/worse student progress, a stepwise linear regression model (that balances predictive power and simplicity, for better interpretability by researchers and learners) has been trained, using the students' self-assessed progress as the outcome variable, and using the other quantitative variables as predictors. To understand the *kinds of insights* that could be gleaned by adding (open-coded) qualitative analysis data to these models (RQ1.2), a single researcher performed an open coding of the diary entries in the dataset, using an inductive conventional content analysis [Hsieh and Shannon, 2005] at the diary sentence level. Within a naturalistic paradigm of research, we ensured the credibility of the coding process by prolonged engagement, persistent observation, and to a lesser extent member checking of data and method triangulation with the student interviews during the SCLA4DE process [Poduthase, 2015]. The resulting codes (see Table 2) were then transformed into binary variables using a process similar to that used in epistemic network analysis (ENA) [Shaffer and Ruis, 2017], i.e., whether a code was present or not on a certain date when the diary was written. Those additional variables from the open coding were used, along with the quantitative ones, as predictors in further GGM and stepwise linear models, thus eliciting contextual factors related to progress that could generalize across the whole dataset of doctoral students.

To explore the answers to RQ2 (single-learner analyses over cohort-based ones), we trained a similar set of models to those in RQ1, using only the data from one learner

at a time. We then compared these personalized models with the ones from RQ1, to understand in what ways they differed, qualitatively speaking (RQ2.1). To understand the objective added value of these single-learner models (RQ2.2), we have compared the stepwise linear models developed for RQ1 and RQ2 in terms of how well they predict the outcome variable (perceived progress). We have triangulated multiple evaluation metrics: root mean squared error (RMSE) and R-squared, both in-sample (i.e., predicting the outcome values in the training set) and out-of-sample (i.e., predicting outcome values of data points never seen by the model) using a 5-fold cross validation method.

To answer RQ3 (ML models to accurately imitate the open coding made by researchers), we have trained multiple natural language processing (NLP) models, using both classic machine learning (e.g., logistic regression, support vector classifiers) as well as state-of-the-art neural network models (including convolutional and recurrent networks, as well as Bidirectional Encoder Representations from Transformers [BERT], see [Devlin et al., 2019]). Each ML model tried to predict whether a certain code is present or not in a certain sentence. We compared the performance distribution of different families of ML models (RQ3.1), evaluated using Cohen's kappa (κ). The evaluation was performed using a 5-fold cross validation process, repeated five times to better understand the distribution of model performance, and then compared it with what is often considered "good agreement" ($\kappa > 0.65$) in qualitative analysis literature [Viera and Garrett, 2005]. Similarly, to understand the reasons for this performance (RQ3.2), we explored the best model's performance in the open-coding imitation task by looking at three dimensions of each code: its level of abstraction, the kind of entity the code represents, and the prevalence (i.e., the frequency of appearance) of the code in the dataset.

4.4 Results

4.4.1 RQ1 (red bot): What across-learners insights can be derived from SCLA data?

4.4.1.1 RQ1.1 From quantitative time series data

As a first step towards understanding what across-learners insights we can derive from a SCLA4DE approach, we developed a simple, interpretable across-learners model of progress. More concretely, the stepwise linear regression of perceived progress as a function of the other quantitative data are represented in Figure 4 below.

On average, the cohort of students' most reliable predictor of progress was to spend more time working on thesis-related tasks (including not only the dissertation itself, but also studies and publications to be included in it). Also, the students' amount of sleep was a positive predictor of progress, albeit with more variability.

We also developed a GGM [Epskamp et al., 2018], which not only looks at the relationship between progress and its quantitative predictors, but rather looks at the relationships between all pairs of variables, while controlling for all the other variables in the dataset. Further, such models can be used to look at contemporaneous relationships (what variables tend to be correlated on the same day), temporal relationships (i.e., correlations between the values of one day and the next) and between-subjects relationships (i.e., whether variables tend to be correlated across participants). The resulting model is represented in Figure 5.

The GGM further confirms the contemporaneous relationship of perceived progress and time dedicated to the thesis. We also uncovered a tension between thesis-related

Code	Definition
Effort	Doing a task that was explicitly labeled as hard, or exerting oneself effortfully in a certain activity
EmailCalls	Dedicating time to doing emails or making calls
Emotion.Positive	Mentions to positive emotional states like joy, pride, curiosity/interest, etc.
Exhaustion.Sickness	Being sick (incl. migraines or headaches) or feeling exhausted during the day
FreeTimeRest	Dedicating time to rest, sleeping more than usual, taking holidays or a day off
GenericTasks	Mentions to having done tasks, without mentioning their nature or whether they are thesis-related or not
HouseworkFamilyPersonal	Dedicating time to household or family-related tasks, or undefined personal matters
InterruptionsInterferences	Interruptions to one's work by other people, or other external interferences in one's work
LackingTime	Mentions to feelings of time pressure, time scarcity, or deadline-driven anxiety
Learning	Spending time learning new things, regardless of whether they are thesis-related or not
Meetings	Spending time in meetings, regardless of their relationship with thesis work
OtherWork	Mentions to tasks not related to the thesis
OtherWork.Teaching	Mentions to teaching tasks (also not related to the thesis)
ProductivityTechniques	Explicitly following a specific productivity technique or strategy (e.g., the Pomodoro method)
ThesisWork	Generic mentions to working on thesis-related tasks
ThesisWork.Reading	Spending time reading materials related to one's thesis topic
ThesisWork.Writing	Spending time writing thesis-related materials (e.g., papers, reports, etc.)

Table 2: Excerpt from the codebook describing the codes unearthed in the open coding of diary entries of doctoral students, mentioned in the case study results (section 4.4)

time and other work-related time (e.g., teaching or other research work not related to the students' dissertation topic). From the temporal relationships we can also see how the progress seems to go in streaks, that is, a high progress in one day seems to be correlated with high progress in the following day. Additionally, longer sleep often happens after a high-progress day. Looking at the between-subjects differences (Figure 5, right) we see that students that spend a lot of time in other work activities tend to spend less time on thesis-related products, and tend to report less progress. While these cohort-based trends are not very surprising, they provide initial ideas for theory-building and for potential interventions (e.g., focusing students to reduce the time they spend working on

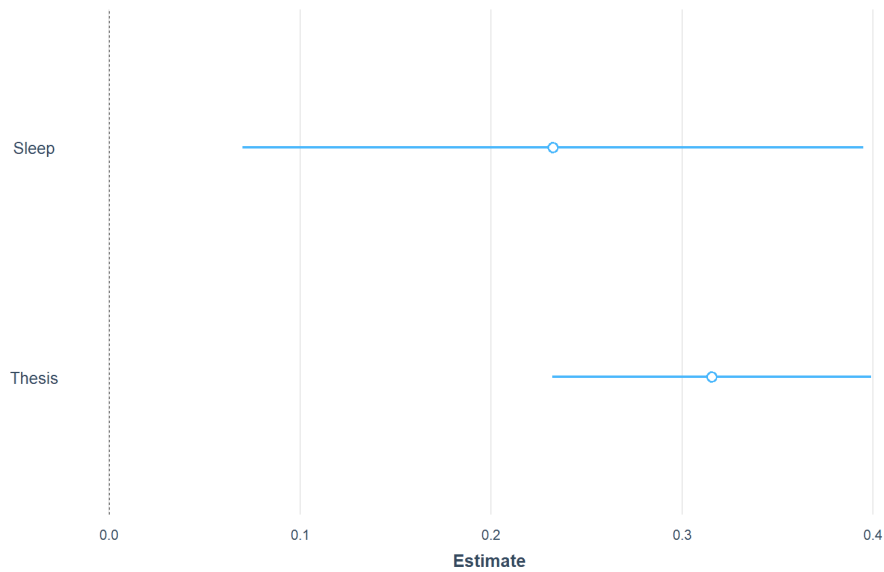


Figure 4: Graphical representation of the predictor coefficients and their confidence intervals, of a stepwise linear regression of perceived progress. Sleep: hours slept that day; Thesis: hours devoted to working on thesis-related materials

non-dissertation tasks).

4.4.1.2 RQ1.2 From adding qualitative data on top of the quantitative ones

To understand the added value of increasing the complexity of the data gathering with an unstructured diary entry, in getting across-learners insights, we used the open coding variables as additional predictors in another stepwise linear regression model of progress (see Figure 6).

By looking at the variables appearing in Figure 6 (i.e., those that the model selected as most predictive of progress), we can observe that there are effects consistent with existing literature (e.g., positive emotions being positively associated with progress, see [Amabile and Kramer, 2011]). There exist also other contextual variables (i.e., coming from the qualitative analysis of narrative diaries) seemingly associated with progress, which could provide ideas about potential interventions and behaviors to change. For instance, the negative impact of housework and family tasks could be linked with students' task/time management ability. This insight seems to be supported by the positive impact of explicitly using productivity techniques, or the negative effect of the feeling of time scarcity already, all of which suggest that interventions that allow students to better organize themselves, like doctoral workshops on time management and productivity (see [Prieto et al., 2022] for one such training), could be a way to overcome these challenges. Once detected as important, these variables could also be the target of direct quantitative measurement by the student in later iterations (see step #5 in Figure 1).

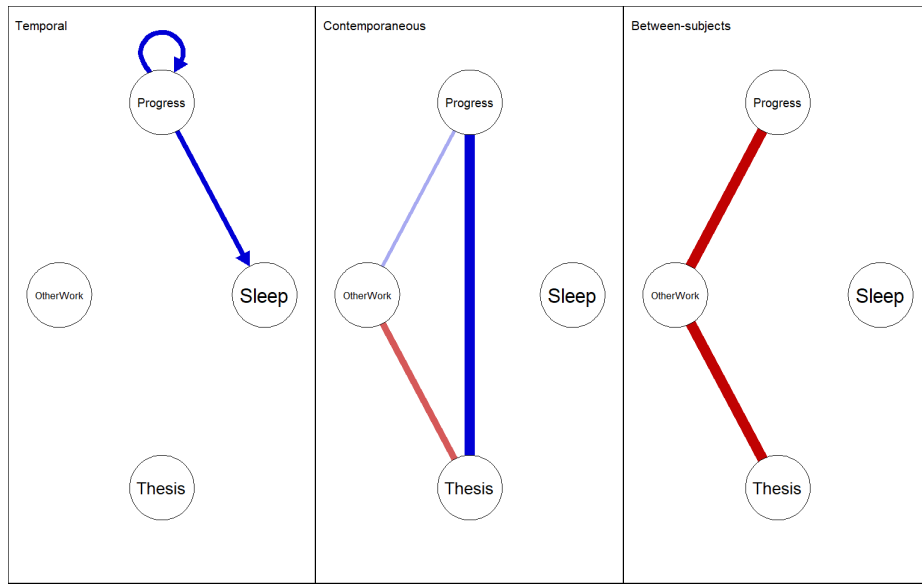


Figure 5: Graphical representation of the Gaussian graphical model (GGM) built using the common quantitative variables of the whole dataset ($N=9$ doctoral students). *OtherWork*: hours devoted to work not related with the thesis; *Progress*: self-evaluation of progress; *Sleep*: hours slept that day; *Thesis*: hours devoted to working on thesis-related materials. Blue lines/arrows indicate positive correlations, red ones indicate negative correlations

4.4.2 RQ2 (blue bot): What is the added value of a single-learner analysis over the cohort-based one?

4.4.2.1 RQ2.1 In terms of subjective insight

By training similar machine learning models to those of RQ1 on the data of each *single* doctoral student, we noticed how personalized models provide arguably more interesting insights, sometimes countering the cohort-based trends outlined above¹. For instance, participant Alice's linear regression model has sleep as a negative predictor of progress (maybe suggesting a tendency to oversleep), and uncovered predictors that the student herself had proposed in the problem/data definition phase. For instance, we find that negative thoughts about the thesis have a deleterious effect on progress, or that spending time on “non-essential” self-care activities (*NonEssentials*, see Figure 7, left) bear a positive effect on progress.

In contrast, for participant Beth (Figure 8), sleep is not a significant predictor of progress at all. However, time dedicated to learning new things (*Learning*), or (not) doing a long-postponed task (*PostponedTask_No*), were interesting predictors of progress. The GGM for Beth also uncovered certain indirect relationships to progress, like that of spending time finding a postdoctoral job (*JobSearch*) which was in tension with the spending time on non-thesis work tasks (*NonThesisWork*).

¹ When describing individual participants' data, fictional names are used throughout the paper (e.g., Alice, Beth).

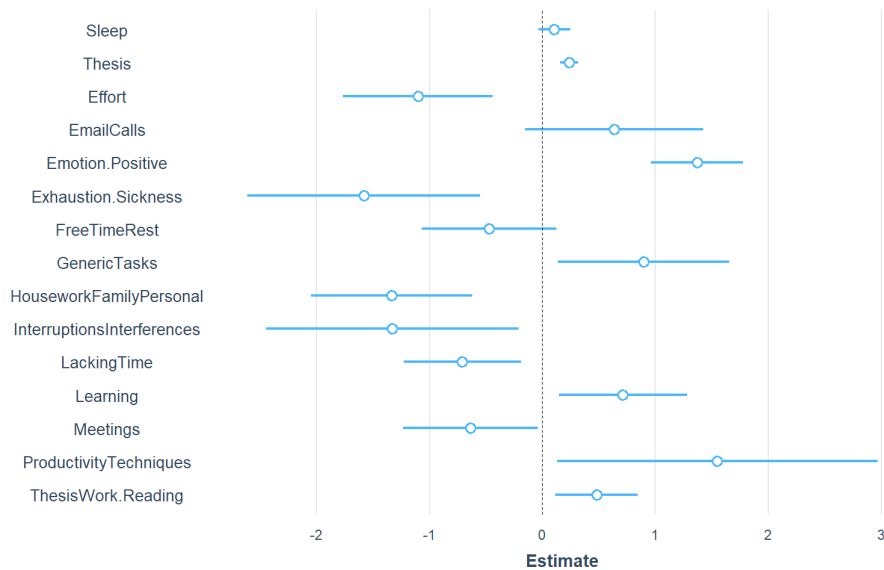


Figure 6: Graphical representation of the predictor coefficients and their confidence intervals, of a stepwise linear regression of perceived progress for the whole cohort ($N=9$ doctoral students), including both quantitative and open coding variables. Circles represent the coefficient estimate; lines represent the confidence interval of the coefficient. Sleep: hours slept that day; Thesis: hours devoted to working on thesis-related materials; for all other variables see the codes in Table 2

Once we add the data from the qualitative analysis, we find further (personalized) contextual correlates of progress, which are even more interesting and actionable. For instance, for participant Alice, the importance of household and family obligations was found an important hindrance to progress (Figure 9, left). Also, certain thesis activities (*ThesisWork.Reading*) and non-thesis ones (*OtherWork.Teaching*) seemed to play the largest positive and negative roles, respectively.

In turn, for participant Beth (Figure 10), the personalized, qualitative-enriched models suggest intriguing insights, like the negative relationship of meetings with progress. This could be due to the quality of the interpersonal relationship with the supervisor, unmet expectations about the outcome of the meetings, or the sheer amount of time spent in meetings impeding progress on the thesis materials. These conjectures could be discussed with the student herself for further contextualization and, eventually, also with her supervisor or lab head. Further, the fact that thesis writing (*ThesisWork.Writing*) seems associated with lower progress days suggests that potential interventions to support writing (e.g., doctoral writing groups or workshops) might be a good starting point.

Yet, these individual models also had limits, often related to lower data volumes. For instance, participant Chris (who had input only 23 diary entries) had stepwise linear regression models with low R-squared (i.e., the amount of variance in the dataset that is explained by the model variables).

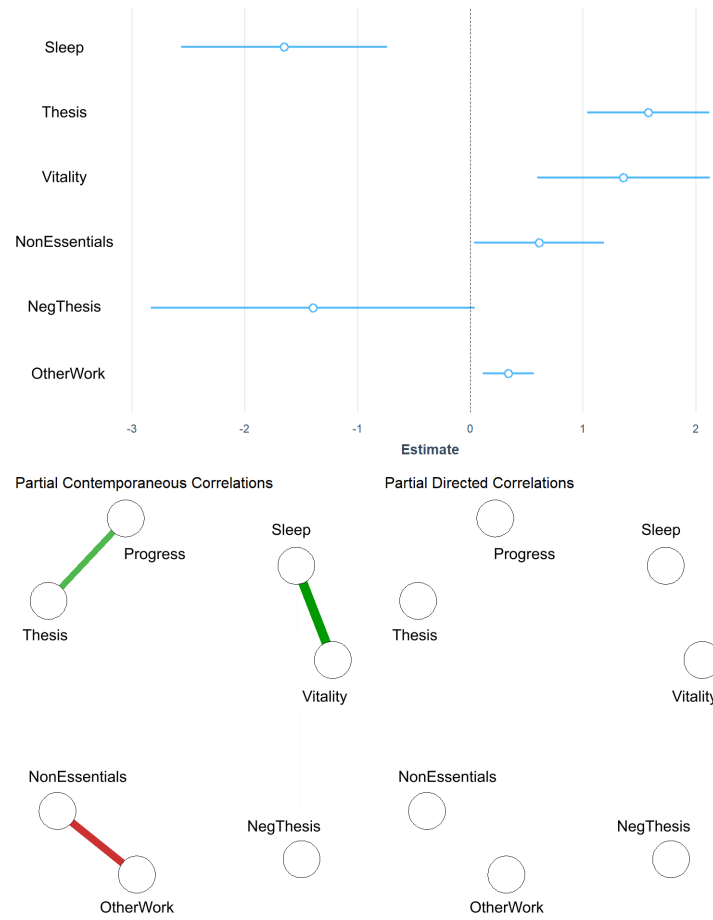


Figure 7: Top: Graphical representation of participant Alice's stepwise linear regression model of progress, using quantitative data only (including personalized variables). Bottom: Alice's GGM representation of contemporaneous relationships between variables. NegThesis: frequency of negative thoughts about the thesis; NonEssentials: hours devoted to non-essential self-care; OtherWork: hours devoted to work not related with the thesis; Progress: self-evaluation of progress; Sleep: hours slept that day; Thesis: hours devoted to working on thesis-related materials; Vitality: subjective feeling of vitality

4.4.2.2 RQ2.2 In terms of objective model performance

We evaluated the different models mentioned in the previous sections (both across-learners and personalized for a single learner) to get objective measures of model performance at predicting the actual progress perceived by a student for a particular day, using the other variables as predictors. Table 1 summarizes the results of such evaluation.

Even if the spread of the evaluation metrics suggests that some of the personalized models might be performing worse than the across-learner ones, the average performance

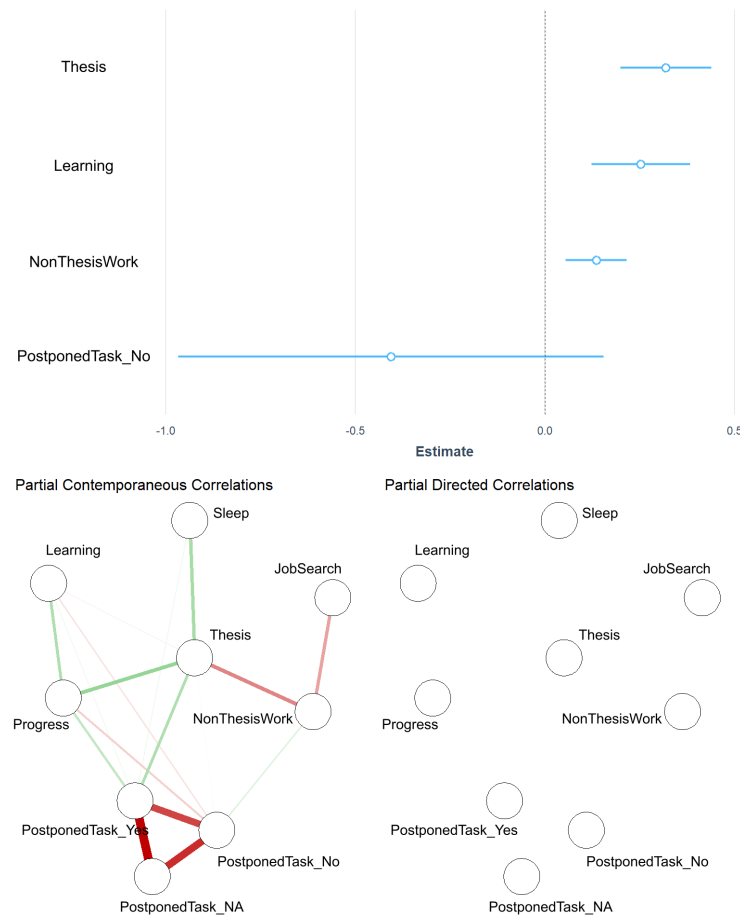


Figure 8: Top: Graphical representation of participant Beth's stepwise linear regression model of progress, using her (customized) quantitative data only. Bottom: Beth's GGM representation of contemporaneous relationships between such quantitative variables. JobSearch: Time spent finding a job for after the doctorate; Learning: hours dedicated to learning new things; NonThesisWork: hours dedicated to thesis-unrelated work; PostponedTask_xx: whether or not a long-postponed task was attempted that day; Progress: self-evaluation of progress; Sleep: hours slept that day; Thesis: hours devoted to working on thesis-related materials

of personalized models enriched with qualitative data vastly overcame the across-learner ones. To investigate the reasons for the few particular low-performing personalized models, we explored the possibility of *non-linear* relationships in the data as a potential cause for such performance. Using decision trees (a simple, easily interpretable non-linear machine learning model) on those students' data seemed to ameliorate the problem, giving again superior performance over the across-learner models, and suggesting potential alternative AI models and paths for future research (i.e., the investigation of non-linear

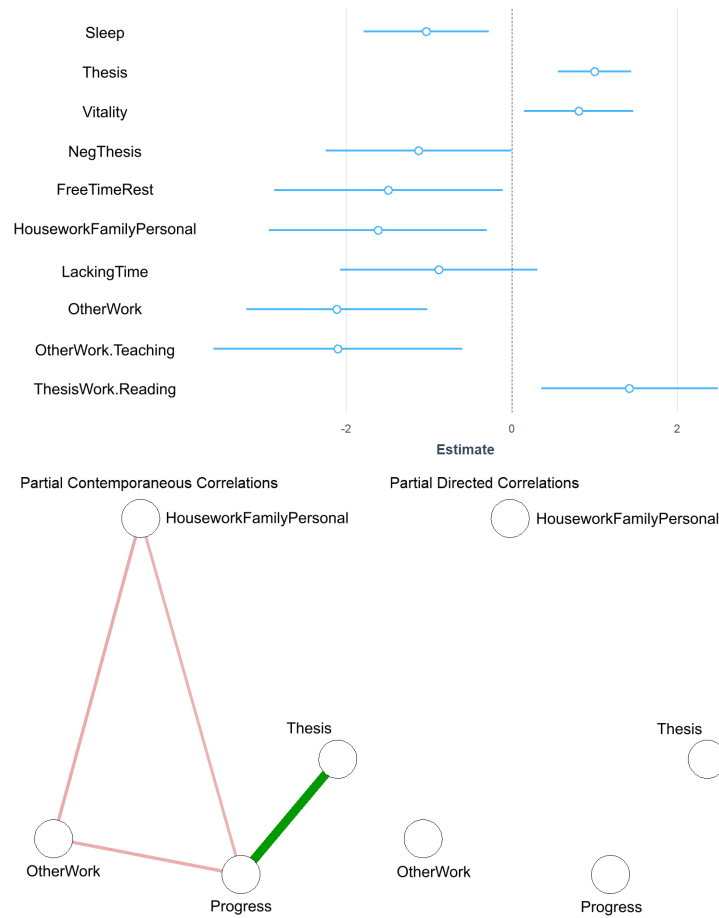


Figure 9: Top: Graphical representation of participant Alice's stepwise linear regression model of progress, once qualitative variables were added into the model. Bottom: Alice's GGM representation of contemporaneous relationships between variables (including qualitative ones). NegThesis: frequency of negative thoughts about the thesis; OtherWork: hours devoted to work not related with the thesis; Progress: self-evaluation of progress; Sleep: hours slept that day; Thesis: hours devoted to working on thesis-related materials; Vitality: subjective feeling of vitality; for all other variables, see the codes in Table 2

models as opposed to the simple linear ones portrayed in this paper). It is also worth noting that the across-learner models enriched with qualitative data outperformed the purely quantitative ones (suggesting that such enrichment of quantitative time series data with qualitative analyses may be beneficial, not only in the frame of a SCLA4DE approach, but also for other classic, cohort-based LA solutions).

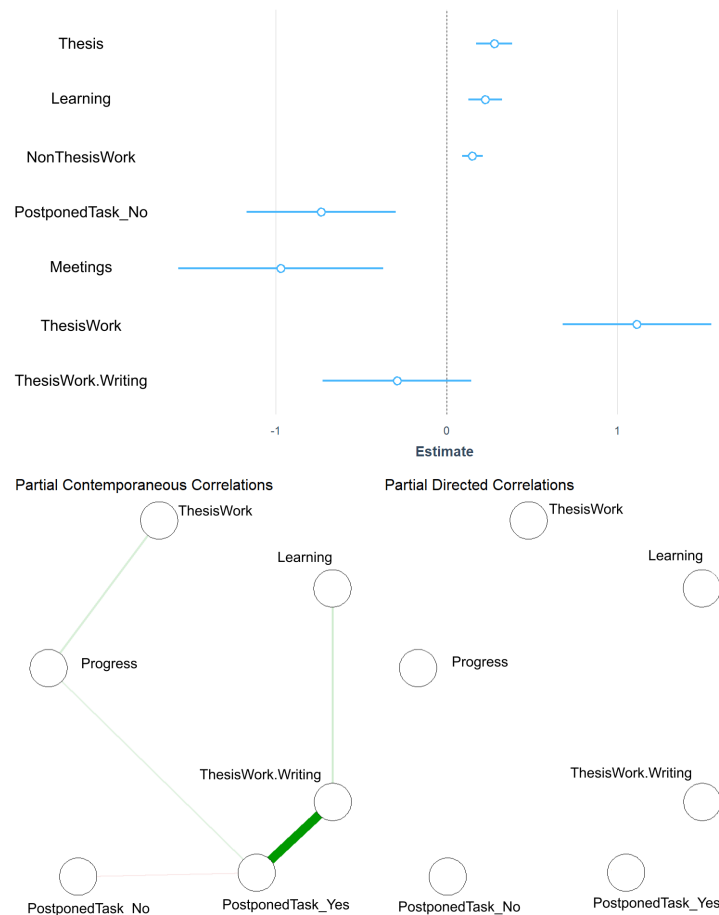


Figure 10: Top: Graphical representation of participant Beth's stepwise linear regression model of progress, once qualitative variables were added into the model. Bottom: Beth's GGM representation of contemporaneous relationships between variables (including qualitative ones). Learning: hours dedicated to learning new things; NonThesisWork: hours dedicated to thesis-unrelated work; PostponedTask_{xx}: whether or not a long-postponed task was attempted that day; Progress: self-evaluation of progress; Thesis: hours devoted to working on thesis-related materials; for all other variables, see the codes in Table 2

4.4.3 RQ3 (green bot) Can computational techniques reproduce researcher-generated open coding reliably?

4.4.3.1 RQ3.1 What AI technique produces most reliable results?

We applied different families of machine learning algorithms commonly used in NLP, to understand how well they can imitate the human-driven open coding. As we can see in Figure 11 below, BERT was the best-performing family, with a median $\kappa = 0.39$.

Model type	Data used	Mean(SD) in-sample RMSE	Mean(SD) in-sample R2	Mean(SD) out-of- sample RMSE	Mean(SD) out-of- sample R2
across- learners	quant	1.56	0.18	1.56	0.19
across- learners	quant+qual	1.28	0.22	1.33	0.21
within-learner (personalized)	quant (incl. cus- tomized)	0.85(0.42)	0.55(0.19)	1.08(0.85)	0.68(0.18)
within-learner (personalized)	quant+qual (incl. cus- tomized)	0.53(0.31)	0.78(0.15)	0.84(0.56)	0.73(0.10)

Table 3: Results of evaluating the stepwise linear regression models of progress, in-sample and out-of-sample (using a 5-fold cross validation scheme). Mean values and standard deviations (SD) reported. RMSE: root mean squared error (smaller is better); R2: R-squared (larger is better)

Other notably-performing families of algorithms were support vector classifiers (SVC, median $\kappa = 0.31$) and naïve Bayes algorithms (NB, median $\kappa = 0.21$). Yet, as we can see there, only a relatively small part of the distribution of coding tasks achieved the desired $\kappa > 0.65$.

4.4.3.2 RQ3.2 What factors influence model performance?

Several factors might be influencing the models' performance in imitating the human coding task: the code's level of abstraction (e.g., highly abstract notions like "mentions that progress was made", vs. very concrete notions like "I answered emails"), the type of entity the code represents (e.g., a description of an experience, like "feeling exhausted", vs. a mention of an activity like "I wrote for my thesis"), or the code's frequency in the dataset (as rare codes create imbalanced supervised learning tasks which often lead to lower-performing models). Figure 12 below shows comparisons of model performance for different code's supervised learning tasks, along these three dimensions.

We can observe that very concrete codes are imitated much better than very abstract ones (Figure 12, left), and indeed BERT's performance for the former category of codes is, on average, close to human-level standards (median kappa=0.65). We can also see (Figure 12, center) how codes describing the quality of a student's experience, or mentions to particular activities, were much more accurately detected than mentions to how the students worked (e.g., procrastination, long hours, etc.). Finally, there also seems to exist a trend whereby very rare codes (low number of appearances in the dataset) are not reliably detected, while more common codes are better detected by BERT (Figure 12, right), as expected. Yet, it is unclear whether this trend is actually linear.

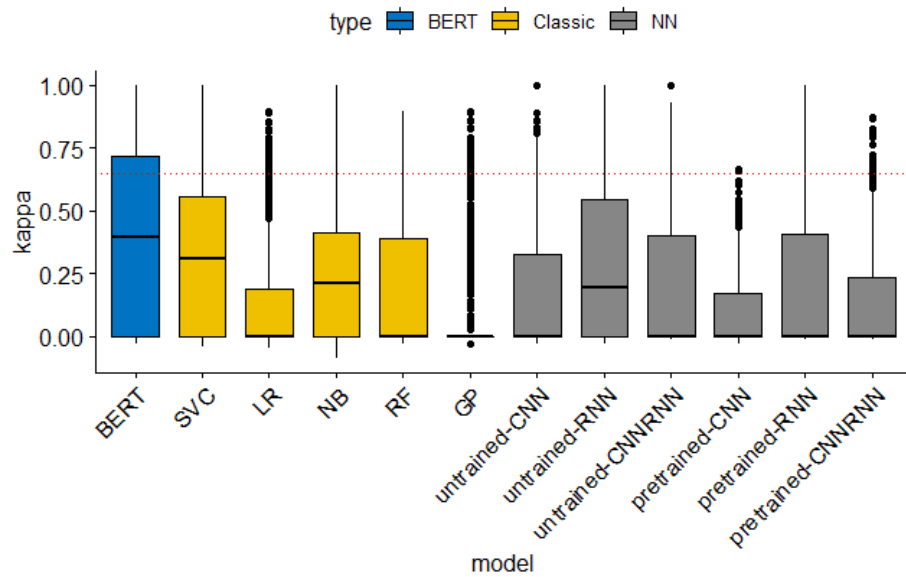


Figure 11: Distribution of inter-rater reliabilities of different kinds of ML models trying to mimic the open coding done by human researchers (Cohen's kappa). NN=neural networks; SVC: support vector classifier; LR=logistic regression; NB=naive Bayes; RF=random forests; GP=gaussian processes; CNN=convolutional NN; RNN=recurrent NN

5 Discussion

In response to the widespread dropout and well-being problems of doctoral education, we proposed SCLA4DE as a human-centered learning analytics process that is deeply centered on doctoral students and their agency, in line with doctoral education research [McAlpine and Amundsen, 2009] and more general educational theories like self-regulated learning [Zimmerman, 1990] and modern evidence-based frameworks of behavior change for improved well-being [Hofmann, 2021]. This student-centeredness contrasts with the dominant approach in higher education LA (which is often instructor-centered), but we suggest that the uniqueness of each doctoral learning process warrants this shift. The illustrative case study analyzing an authentic dataset from a small but highly varied sample of doctoral students helped in exploring the feasibility and potential of the computational elements in the SCLA4DE approach, but it also highlighted insights and limitations of this approach, along the six steps of its process, as laid out in section 3:

In terms of step 1. *Problem and Data Definition*, our illustrative case study focused on students' perception of progress as a key motivational factor unveiled by recent doctoral education research [Devos et al., 2017, De Clercq et al., 2021]. Yet, the lack of progress can take many forms, depending on multiple factors like research discipline or the stage within the doctorate the student is in (i.e., what is the current learning task). This initial step in the SCLA4DE process enabled our study's doctoral students to map their specific progress challenges with personalized data variables to be gathered by them daily. Yet,

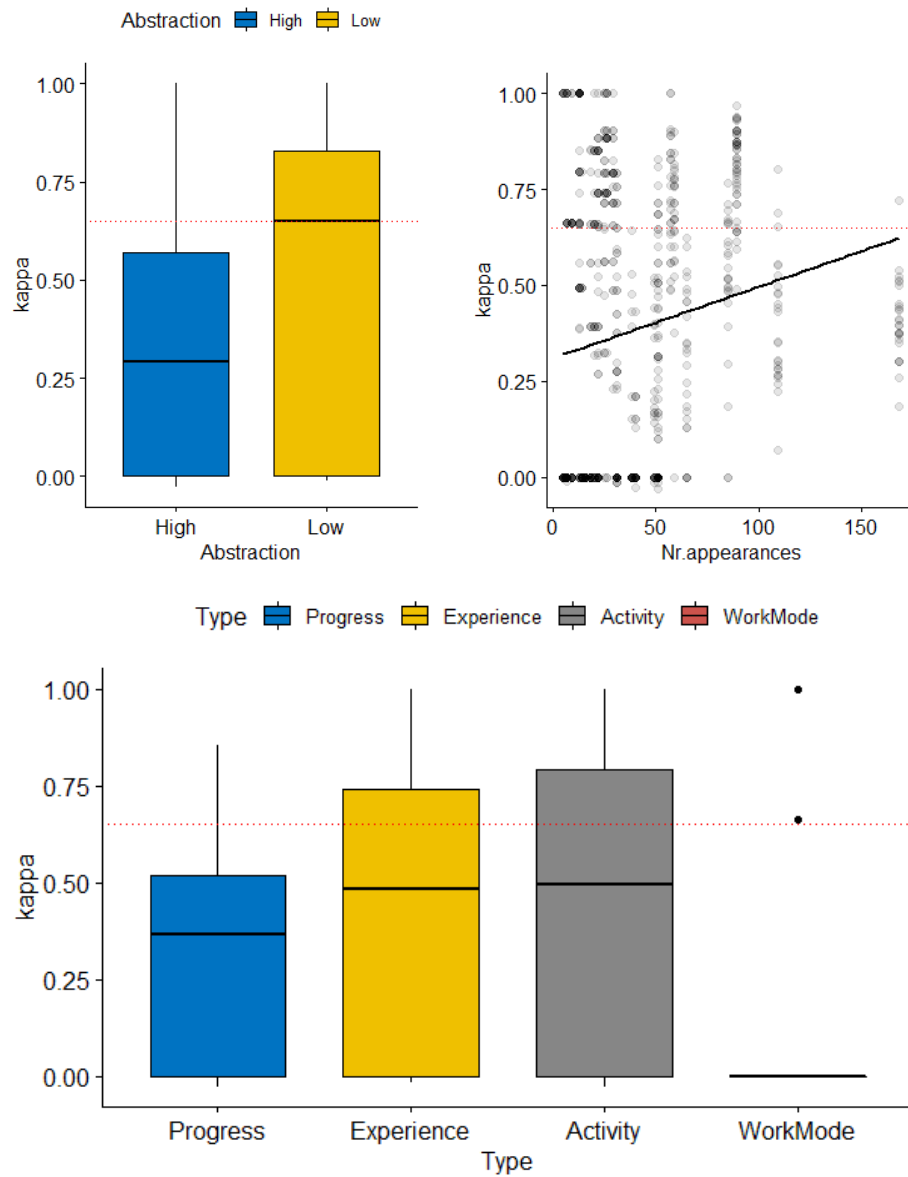


Figure 12: Distribution of inter-rater reliabilities of a multi-language BERT model trying to mimic the open coding done by human researchers (Cohen's kappa), as a function of several code and dataset characteristics. Top-left: level of abstraction of codes; Top-right: number of appearances of a code in the dataset; Bottom: type of entity represented by the code

this flexibility-enhancing step also hints at limits of the approach. For instance, not every kind of doctoral problem will be amenable to this approach: problems not clearly perceived by students themselves, or which do not depend on student behavior to be fixed, may benefit little from a SCLA4DE approach (although doctoral supervisors may help in detecting problems that the student is unable to see).

Regarding step 2. *Longitudinal Data Gathering*, we currently conceptualize it as requiring active participation by students. While this self-reporting serves as an attention-focusing device also used in many behavior change approaches [Fernández and Mairal, 2017, Sellés et al., 2015], and circumvents ethical concerns about gathering data from students unknowingly, our illustrative case study and the size of the dataset gathered (on average, 30 daily entries per participant, with diary entries averaging 35 words per entry) also hints at inherent limitations of this approach, given the feelings of time scarcity that many doctoral students experience. Issues like tracking fatigue are well known in the literature about quantified self and self-tracking [Choe et al., 2014], with no clear generalizable solution. Our current best guess to overcoming this hurdle is to provide useful insights (from the individualized analytics models) to students as soon as possible, thus outweighing the perceived cost of this data gathering. Our approach's current reliance on diaries also makes it share typical limitations of diary studies, such as self-selection bias, confirmability, or potential lack of depth in data [Mittelmeier et al., 2021]. These weaknesses could be ameliorated with the use of further mixed methods (e.g., Mittelmeier and colleagues' use of social network analysis and interviews aside from the self-tracking/diaries), or the use of multimodal learning analytics more generally (as already attempted by [Di Mitri et al., 2017]). Indeed, in a sense some of these weaknesses (like the self-reflection bias) are what the SCLA4DE approach exploits in order to filter the myriad of context pieces in the students' everyday experience that could be relevant, into a more manageable set. Therefore this is a key HCLA mechanism by which SCLA4DE adapts the analytics models to the stakeholders, increasing their agency [Chen and Zhu, 2019].

Step 3. *Qualitative Analysis of Unstructured Data* was the focus of our third research question (on whether the qualitative coding of narrative diary entries can be automated reliably). However, even before that automation can be attempted, a large enough dataset of such diary entries has to be manually coded by human researchers. This raises concerns about the timeliness of this step (especially considering the need to provide value-adding insights as soon as possible, as mentioned above), and the human effort involved in this step. Collaboration between multiple research groups (maybe specializing in different kinds of coding or doctoral education theories) could be a way forward in this sense, but probably will need to be combined with other techniques such as the use of semi-automated coding (e.g., aided by regular expressions or other means, [Cai et al., 2019]), or the use of reinforcement learning by having students themselves revise a sample of their data (i.e., a form of student member checking, often recommended in qualitative analysis). Indeed, one of the main limitations of this step in our illustrative case study was that the open coding of narrative diary entries used to enrich the quantitative dataset was performed by a single researcher. While we used several techniques to ensure the credibility of the analysis (prolonged engagement, persistent observation, and to a lesser extent member checking of data and method triangulation with the student interviews during the SCLA4DE process), we did not use the others (e.g., peer debriefing, negative case analysis, referential adequacy, and more complete forms of member checking and triangulation) [Poduthase, 2015]. Further, the use of multiple coders (and evaluating their inter-rater reliability as a marker of validity) would be indeed very important in order to claim generalizability about the cohort-based models (see step #6). Yet, given our

present goal of merely *illustrating* the *kinds of insights* that could be gleaned from adding such a qualitative analysis to the analytics process (and not the generalizability of those particular models extracted from our limited dataset), we considered this weakness non-critical. This issue of inter-rater reliability is indeed critical for the analytics approach we propose, as it relies on qualitative content analysis — as noted recently by other researchers in the field [Kitto et al., 2023]. Regarding the automation of such qualitative analyses, the evidence from our illustrative case study suggests that, while the best-performing models (a multi-language BERT) perform much better than chance, they still cannot accurately reproduce human coding for all kinds of codes. We found that the level of abstraction of the code is an important aspect to consider, as is the provision of enough examples of each code within the dataset (to avoid excessive class imbalance in the supervised learning task). Our exploratory findings thus suggest that gathering data for longer periods of time may be needed for reliable analysis and automation, but also that other analytics techniques may be needed to deal with higher-level codes or constructs of interest to students and researchers.

As to step 4. *Individualized Modeling*, our illustrative case study explored the added value of these models over cohort-based ones (our second research question). We found that the personalized models, which used quantitative variables defined by students themselves, provided more actionable insights (and gave students a voice in the direction of the research and the interventions — which is key in ethical, responsible research). Further, we also found occasions in which the insights from a learner's data countered the across-learner average trends (as also demonstrated by [Fisher et al., 2018] in the field of psychology), thus supporting the importance of such an idiographic approach to provide support that works for each individual. The addition of qualitative open coding data improved the actionability of the insights (in terms of suggesting potential interventions) and contextual richness of the models. Furthermore, the personalized models were found objectively more predictive than across-learner ones. This is not surprising, as personalized models have a much easier task scope (i.e., predict the progress of just one student). Here again, we found that models enriched with qualitative analysis data performed objectively better. Our findings also highlight aspects that future technology designers proposing SCLA4DE systems should pay special attention to. For instance, the number of predictor variables in the ML models grew very quickly (especially, after we added the data stemming from the open coding process) — to the point that there may be more predictors than data points. This highlights the importance of dimensionality reduction and variable selection mechanisms, of which the stepwise regression used in our case study is just one potential alternative. Our results also highlighted inherent limitations of the SCLA4DE approach, like the importance of having enough data volume (i.e., sufficient number of diary entries), in order to obtain reliable personalized models.

While our illustrative case study did not explore step 5. *Personalized Sense-making, Interventions and Re-definition* in depth, or how undergoing it actually affects doctoral students' sense of agency, several insights can be gleaned from our illustrative findings. First, they highlight the limits of the data and models produced, which are correlational and exploratory in nature, and based on only brief mentions of an event or factor in a short narrative entry. These models will not be able to suggest causal relationships and will need further sense-making and contextualization by students themselves (cf. [Shibani et al., 2019]). Similarly, interventions or changes in behavior will also need to be defined by and with students, so that they are fit-for-context (and increase students' agency) [Strosahl et al., 2012]. It is worth noting that the need for a “researcher/expert in the loop” in this and other steps of the SCLA4DE approach is a limitation to its practical application at scale, as it is resource-intensive. This also suggests that further

computational automation or support could be developed for these human-intensive steps, to reduce researcher/expert effort.

Finally, we explored step 6. *Cohort-based modeling* in the first research question of our illustrative case study (what kinds of insights can be gleaned from cohort-based models of SCLA4DE data). The analysis of our authentic (but small) dataset suggests that the simple models built with the quantitative variables that were common across learners led to limited insights that applied across the group of nine doctoral students. In our study, the addition of data coming from qualitative analysis (step #3 above) led to more interesting across-learners models that were, at the same time, more actionable — as they pointed to particular contextual features that tended to relate to better/worse progress (e.g., use of productivity techniques, or feelings of time scarcity). However, several critical limitations curtail the generalizability of these models, which should not be taken as applying to every doctoral candidate out there. While varied in terms of country, discipline, and stage within the doctorate, the small sample size of our illustrative study (which is an endemic problem of doctoral education research that does not use survey methods, e.g., [Di Mitri et al., 2017]) should raise concerns about the generalizability of our findings. Further, the fact that only one coder performed the qualitative analysis limits greatly the generalizability and reliability of the cohort-based models that use qualitative data, which should be taken as merely illustrative and not general truths about doctoral students.

6 Conclusions and a future SCLA4DE research agenda

This paper started with the severe dropout and well-being challenges that doctoral education faces, and its need for personalized support, given doctoral student learning process's uniqueness. Inspired by modern evidence-based psychotherapy approaches to behavior change [Strosahl et al., 2012, Hofmann, 2021], we have proposed SCLA4DE, a novel human-centered learning analytics approach to address that need, and we have described how human actors and computational elements would collaborate to provide such support while taking human stakeholders and their agency as a primary consideration. Our illustrative case study using an authentic doctoral education dataset has shown how three computational elements would support doctoral education address these challenges, balancing explainability and performance as needed.

Overall, the main technical contribution of this paper is the definition of a socio-technical system (more concretely, a human-AI team) tailored to support doctoral education, by structuring the interactions between human actors and computational components in specific ways to provide such support. Our approach also defines specific kinds of computational elements, their expected properties and goals (e.g., using high-explainability linear models to aid data interpretation by students, or high-performance foundation models for mimicking researcher labeling of unstructured data). In this sense, it is (to the best of our knowledge) the first practical implementation of a human-AI approach in doctoral education, and the first one that focuses on human agency after design-time (i.e., during its deployment and operation). Particular elements of the approach are also rare within the field of LA, such as the use of student-defined variables and unstructured data to uncover individual context influences (which aim to enhance the student agency).

We showed that a hybrid human-AI approach is both technically feasible and value-adding, illustrating its practical usefulness to stakeholders like PhD candidates — in the context of doctoral education and its dropout and well-being challenges. We have defined properties of the involved computational elements (e.g., in terms of model explainability

or necessary data volume), and requirements for the human stakeholders involved in the process (e.g., active data collection by students over longer periods of time, or the coding of initial unstructured data by researchers).

However, this paper has barely scratched the surface of many of the research-worthy questions around the SCLA4DE approach. These would define a tentative future research agenda including:

- The use of a wider variety of data gathering methods within a SCLA4DE approach, including multimodal LA techniques [Ochoa, 2017]. This also includes the use of different and more advanced analytics techniques, from ENA (cf. [Prieto et al., 2021] for an example) to reinforcement learning.
- Investigating the potential role of additional relevant stakeholders (like doctoral supervisors, coaches, therapists, etc.) which could be added in extended versions of the SCLA4DE approach. Doctoral supervisors could be involved, for instance, to provide additional data that help deal with diaries' issues of confirmability and social desirability bias [Mittelmeier et al., 2021].
- Related to the previous point, the SCLA4DE process, in its current incarnation, is quite labor-intensive for humans (from the coding of diaries by researchers to the supporting of students in problem/data definition and sense-making. While some of these aspects could be aided by researcher collaboration and later automation of the coding process (as we have explored using BERT models in our illustrative case study), the development of further computational elements that support (or even automate) the different human-led tasks in SCLA4DE process, could be attempted (once the SCLA4DE process is considered stable). These new elements could range from behavioral intervention recommender systems to sense-making assistants that help students with the interpretation of the personalized models.
- Investigating more deeply the ethical issues implied by a SCLA4DE approach, including challenges balancing data privacy (e.g., student-owned data) with the ability for researchers to derive across-learner insights (through anonymization, which may not be trivial for unstructured diary data).
- Practical and motivational issues arising from SCLA4DE's requirement of active (repeated measures) data gathering by doctoral students (which may lead to e.g., participation fatigue or worsening stress due to time scarcity) still need to be addressed. Providing valuable insights and data visualizations to students as early as possible, the use of novel human-computer interaction techniques to make such interactions less effortful, or using more passive (but still individually relevant) data sources as a complement to self-reports could play a role in addressing this issue. The use of coaching and motivational techniques by experts and other human actors could also contribute to tackle these issues.
- Given the variety of causes that may lead to a lack of progress, low well-being and eventual student dropout (in part, due to the variety of learning tasks at different stages of the doctorate [Di Mitri et al., 2017]), it would be important to explore the limits of this approach. Finding out which problems are most effectively dealt with using SCLA4DE (and which are not) should also be part of our future inquiry.
- The use of an SCLA4DE approach to design HCLA systems does not invalidate the rest of design-time methods that the HCLA community has proposed (see Section

2.1). Delving into what are the most beneficial design processes for developing SCLA4DE systems specifically (e.g., using participatory methods as suggested by [Holstein and Aleven, 2021] or value-sensitive design techniques as proposed by [Chen and Zhu, 2019]) would be another avenue worthy of inquiry.

- The generation or expansion of empirically-grounded doctoral education theory (and the role of student agency) using SCLA4DE-generated data, once we accumulate enough personalized data. Using a self-regulated learning [Zimmerman, 1990] or social-emotional learning [Conley, 2015] as theoretical starting points may seem especially fitting. The use of methodologies like comparative analyses [Smelser, 2013] or even causal analyses [Brodersen et al., 2015] may be routes for this theorizing, as randomized experiments will not always be feasible (or ethical).

The modularity of the SCLA4DE approach will enable different levels of automation in the future, be it with the goal of looking for complementarity between human and AI capabilities [Wilder et al., 2020] or proceeding to full automation (cf. [Molenaar, 2021]). Either way, the heavy involvement of learners themselves not only in the design but also the deployment and everyday operation of SCLA4DE systems aims at ensuring trustworthy, responsible computation, and provides a very concrete example of how to augment human intelligence with machine intelligence.

Acknowledgements

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 669074 and the Erasmus Plus programme, grant agreement 2019-1-NO01-KA203-060280. This project has received funding from Estonian Research Council’s Personal Research Grant (PRG) project PRG1634. The Universidad de Valladolid co-authors acknowledge grant PID2020-112584RB-C32 funded by MCIN/AEI/10.13039/501100011033.

References

- [Ahn et al., 2019] Ahn, J., Campos, F., Hays, M., and Digiacomio, D. (2019). Designing in Context: Reaching Beyond Usability in Learning Analytics Dashboard Design. *Journal of Learning Analytics*, 6(2):70–85.
- [Ali and Gregg Kohun, 2007] Ali, A. and Gregg Kohun, F. (2007). Dealing with Social Isolation to Minimize Doctoral Attrition – A Four Stage Framework. *International Journal of Doctoral Studies*, 2:033–049.
- [Amabile and Kramer, 2011] Amabile, T. and Kramer, S. (2011). *The Progress Principle: Using Small Wins to Ignite Joy, Engagement, and Creativity at Work*. Harvard Business Press.
- [Avrahami et al., 2020] Avrahami, D., Williams, K., Lee, M. L., Tokunaga, N., Tjahjadi, Y., and Marlow, J. (2020). Celebrating Everyday Success: Improving Engagement and Motivation using a System for Recording Daily Highlights. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13, Honolulu HI USA. ACM.
- [Bair and Haworth, 2004] Bair, C. R. and Haworth, J. G. (2004). Doctoral student attrition and persistence: A meta-synthesis of research. In *Higher Education: Handbook of Theory and Research*, pages 481–534. Springer.
- [Blodgett and Madaio, 2021] Blodgett, S. L. and Madaio, M. (2021). Risks of AI foundation models in education. *arXiv:2110.10024 [cs.CY]*.

- [Bommasani et al., 2021] Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., Donahue, C., Doumbouya, M., Durmus, E., Ermon, S., Etchemendy, J., Ethayarajh, K., Fei-Fei, L., Finn, C., Gale, T., Gillespie, L., Goel, K., Goodman, N., Grossman, S., Guha, N., Hashimoto, T., Henderson, P., Hewitt, J., Ho, D. E., Hong, J., Hsu, K., Huang, J., Icard, T., Jain, S., Jurafsky, D., Kalluri, P., Karamcheti, S., Keeling, G., Khani, F., Khattab, O., Koh, P. W., Krass, M., Krishna, R., Kuditipudi, R., Kumar, A., Ladhak, F., Lee, M., Lee, T., Leskovec, J., Levent, I., Li, X. L., Li, X., Ma, T., Malik, A., Manning, C. D., Mirchandani, S., Mitchell, E., Munyikwa, Z., Nair, S., Narayan, A., Narayanan, D., Newman, B., Nie, A., Niebles, J. C., Nilforoshan, H., Nyarko, J., Ogut, G., Orr, L., Papadimitriou, I., Park, J. S., Piech, C., Portelance, E., Potts, C., Raghunathan, A., Reich, R., Ren, H., Rong, F., Roohani, Y., Ruiz, C., Ryan, J., Ré, C., Sadigh, D., Sagawa, S., Santhanam, K., Shih, A., Srinivasan, K., Tamkin, A., Taori, R., Thomas, A. W., Tramer, F., Wang, R. E., Wang, W., Wu, B., Wu, J., Wu, Y., Xie, S. M., Yasunaga, M., You, J., Zaharia, M., Zhang, M., Zhang, T., Zhang, X., Zhang, Y., Zheng, L., Zhou, K., and Liang, P. (2021). On the Opportunities and Risks of Foundation Models. *arXiv:2108.07258 [cs.LG]*.
- [Brodersen et al., 2015] Brodersen, K. H., Gallusser, F., Koehler, J., Remy, N., and Scott, S. L. (2015). Inferring causal impact using Bayesian structural time-series models. *The Annals of Applied Statistics*, 9(1):247–274.
- [Buckingham Shum, 2023] Buckingham Shum, S. (2023). What could Learning Analytics learn from Human-Computer Interaction? In Bartimote-Aufflick, K., Gasevic, D., and Howard, S., editors, *Theory Informing and Arising from Learning Analytics*. Springer. Forthcoming.
- [Buckingham Shum et al., 2019] Buckingham Shum, S., Ferguson, R., and Martinez-Maldonado, R. (2019). Human-centred learning analytics. *Journal of Learning Analytics*, 6(2):1–9.
- [Cai et al., 2019] Cai, Z., Siebert-Evenstone, A., Eagan, B., Shaffer, D. W., Hu, X., and Graesser, A. C. (2019). nCoder+: A Semantic Tool for Improving Recall of nCoder Coding. In *International Conference on Quantitative Ethnography*, pages 41–54. Springer.
- [Chen and Zhu, 2019] Chen, B. and Zhu, H. (2019). Towards Value-Sensitive Learning Analytics Design. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, pages 343–352, Tempe AZ USA. ACM.
- [Choe et al., 2014] Choe, E. K., Lee, N. B., Lee, B., Pratt, W., and Kientz, J. A. (2014). Understanding quantified-selfers’ practices in collecting and exploring personal data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1143–1152, Toronto Ontario Canada. ACM.
- [Conley, 2015] Conley, C. S. (2015). SEL in higher education. In *Handbook of Social and Emotional Learning: Research and Practice*, pages 197–212. The Guilford Press.
- [De Clercq et al., 2021] De Clercq, M., Frenay, M., Azzi, A., Klein, O., and Galand, B. (2021). All You Need is Self-Determination: Investigation of PhD Students’ Motivation Profiles and Their Impact on the Doctoral Completion Process. *International Journal of Doctoral Studies*, 16:189–209.
- [De Silva et al., 2022] De Silva, L. M. H., Chounta, I.-A., Rodríguez-Triana, M. J., Roa, E. R., Gramberg, A., and Valk, A. (2022). Toward an Institutional Analytics Agenda for Addressing Student Dropout in Higher Education: An Academic Stakeholders’ Perspective. *Journal of Learning Analytics*, 9(2):179–201.
- [Devlin et al., 2019] Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv:1810.04805 [cs]*.
- [Devos et al., 2017] Devos, C., Boudrenghien, G., Van der Linden, N., Azzi, A., Frenay, M., Galand, B., and Klein, O. (2017). Doctoral students’ experiences leading to completion or attrition: A matter of sense, progress and distress. *European journal of psychology of education*, 32(1):61–77.

- [Di Mitri et al., 2017] Di Mitri, D., Scheffel, M., Drachsler, H., Börner, D., Ternier, S., and Specht, M. (2017). Learning pulse: A machine learning approach for predicting performance in self-regulated learning using multimodal data. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, pages 188–197. ACM.
- [Dimitriadis et al., 2021] Dimitriadis, Y., Martínez-Maldonado, R., and Wiley, K. (2021). Human-Centered Design Principles for Actionable Learning Analytics. In Tsiatsos, T., Dimitriadis, S., Mikropoulos, A., and Dagdilelis, V., editors, *Research on E-Learning and ICT in Education*, pages 277–296. Springer International Publishing, Cham.
- [Dollinger et al., 2019] Dollinger, M., Liu, D., Arthars, N., and Lodge, J. (2019). Working Together in Learning Analytics Towards the Co-Creation of Value. *Journal of Learning Analytics*, 6(2):10–26.
- [Ellis, 1999] Ellis, M. V. (1999). Repeated Measures Designs. *The Counseling Psychologist*, 27(4):552–578.
- [Epskamp et al., 2018] Epskamp, S., Waldorp, L. J., Möttus, R., and Borsboom, D. (2018). The Gaussian Graphical Model in Cross-Sectional and Time-Series Data. *Multivariate Behavioral Research*, 53(4):453–480.
- [European Commission. Directorate General for Education and Culture. et al., 2015] European Commission. Directorate General for Education and Culture., CHEPS., and NIFU. (2015). *Dropout and Completion in Higher Education in Europe: Main Report*. Publications Office, LU.
- [Evans et al., 2018] Evans, T. M., Bira, L., Gastelum, J. B., Weiss, L. T., and Vanderford, N. L. (2018). Evidence for a mental health crisis in graduate education. *Nature Biotechnology*, 36(3):282–284.
- [Fei et al., 2022] Fei, N., Lu, Z., Gao, Y., Yang, G., Huo, Y., Wen, J., Lu, H., Song, R., Gao, X., Xiang, T., Sun, H., and Wen, J.-R. (2022). Towards artificial general intelligence via a multimodal foundation model. *Nature Communications*, 13(1):3094.
- [Ferguson et al., 2016] Ferguson, R., Brasher, A., Clow, D., Griffiths, D., and Drachsler, H. (2016). Learning analytics: Visions of the future. <https://oro.open.ac.uk/45312/>.
- [Ferguson et al., 2019] Ferguson, R., Clow, D., Griffiths, D., and Brasher, A. (2019). Moving forward with learning analytics: Expert views. *Journal of Learning Analytics*, 6(3):43–59.
- [Fernández and Mairal, 2017] Fernández, E. N. and Mairal, J. B. (2017). Behavioral activation versus cognitive restructuring to reduce automatic negative thoughts in anxiety generating situations. *Psicothema*, 29(2):172–177. Publisher: Colegio Oficial De Psicólogos Del Principado De Asturias.
- [Fisher et al., 2018] Fisher, A. J., Medaglia, J. D., and Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences*, 115(27):E6106–E6115.
- [Fitzpatrick, 2018] Fitzpatrick, G. (2018). A Short History of Human Computer Interaction: A People-Centred Perspective. In *Proceedings of the 2018 ACM SIGUCCS Annual Conference*, page 3, Orlando Florida USA. ACM.
- [Godskesen and Kobayashi, 2016] Godskesen, M. and Kobayashi, S. (2016). Coaching doctoral students – a means to enhance progress and support self-organisation in doctoral education. *Studies in Continuing Education*, 38(2):145–161.
- [Grover, 2007] Grover, V. (2007). Successfully Navigating the Stages of Doctoral Study. *International Journal of Doctoral Studies*, 2:009–021.
- [Hakami and Hernández Leo, 2020] Hakami, E. and Hernández Leo, D. (2020). How are learning analytics considering the societal values of fairness, accountability, transparency and human well-being?: A literature review. In Martínez-Monés, A., Álvarez, A., Caeiro-Rodríguez, M., and Y., D., editors, *LASI-SPAIN 2020: Learning Analytics Summer Institute Spain 2020: Learning Analytics. Time for Adoption?*, pages 121–141. CEUR Workshop Proceedings.

- [Hardre et al., 2019] Hardre, P. L., Liao, L., Dorri, Y., and Beeson Stoesz, M. (2019). Modeling American Graduate Students' Perceptions Predicting Dropout Intentions. *International Journal of Doctoral Studies*, 14:105–132.
- [Hayes and Hofmann, 2021] Hayes, S. C. and Hofmann, S. G. (2021). “Third-wave” cognitive and behavioral therapies and the emergence of a process-based approach to intervention in psychiatry. *World Psychiatry*, 20(3):363–375.
- [Hofmann, 2021] Hofmann, S. G. (2021). The Future of Cognitive Behavioral Therapy. *Cognitive Therapy and Research*, 45(3):383–384.
- [Holstein and Aleven, 2021] Holstein, K. and Aleven, V. (2021). Designing for human-AI complementarity in K-12 education. *arXiv:2104.01266 [cs]*.
- [Holstein et al., 2019] Holstein, K., McLaren, B. M., and Aleven, V. (2019). Co-designing a real-time classroom orchestration tool to support teacher-AI complementarity. *Journal of Learning Analytics*, 6(2):27–52.
- [Hsieh and Shannon, 2005] Hsieh, H.-F. and Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative health research*, 15(9):1277–1288.
- [J. Pifer and L. Baker, 2016] J. Pifer, M. and L. Baker, V. (2016). Stage-Based Challenges and Strategies for Support in Doctoral Education: A Practical Guide for Students, Faculty Members, and Program Administrators. *International Journal of Doctoral Studies*, 11:015–034.
- [Jackman et al., 2021] Jackman, P. C., Jacobs, L., Hawkins, R. M., and Sisson, K. (2021). Mental health and psychological wellbeing in the early stages of doctoral study: A systematic review. *European Journal of Higher Education*, 12(3):293–313.
- [Kitto et al., 2023] Kitto, K., Manly, C. A., Ferguson, R., and Poquet, O. (2023). Towards more replicable content analysis for learning analytics. *Proceedings of Learning Analytics and Knowledge 2023 (LAK23)*.
- [Kloos et al., 2022] Kloos, C. D., Dimitriadis, Y., Hernández-Leo, D., Alario-Hoyos, C., Martínez-Monés, A., Santos, P., Muñoz-Merino, P. J., Asensio-Pérez, J. I., and Safont, L. V. (2022). H2o learn-hybrid and human-oriented learning: Trustworthy and human-centered learning analytics (tahcla) for hybrid education. In *2022 IEEE Global Engineering Education Conference (EDUCON)*, pages 94–101. IEEE.
- [Knight et al., 2017] Knight, S., Wise, A. F., and Chen, B. (2017). Time for change: Why learning analytics needs temporal analysis. *Journal of Learning Analytics*, 4(3):7–17.
- [Leony et al., 2015] Leony, D., Merino, P. J. M., Valiente, J. A. R., Pardo, A., and Kloos, C. D. (2015). Detection and evaluation of emotions in massive open online courses. *J. UCS*, 21(5):638–655.
- [Mackie and Bates, 2019] Mackie, S. A. and Bates, G. W. (2019). Contribution of the doctoral education environment to PhD candidates' mental health problems: A scoping review. *Higher Education Research & Development*, 38(3):565–578.
- [Maher et al., 2020] Maher, M. A., Wofford, A. M., Roksa, J., and Feldon, D. F. (2020). Exploring Early Exits: Doctoral Attrition in the Biomedical Sciences. *Journal of College Student Retention: Research, Theory & Practice*, 22(2):205–226.
- [Martínez-Maldonado et al., 2015] Martínez-Maldonado, R., Pardo, A., Mirriahi, N., Yacef, K., Kay, J., and Clayphan, A. (2015). Latux: An iterative workflow for designing, validating, and deploying learning analytics visualizations. *Journal of Learning Analytics*, 2(3):9–39.
- [Mavrikis et al., 2019] Mavrikis, M., Karkalas, S., Cukurova, M., and Papapesiou, E. (2019). Participatory Design to Lower the Threshold for Intelligent Support Authoring. In Isotani, S., Millán, E., Ögan, A., Hastings, P., McLaren, B., and Luckin, R., editors, *Artificial Intelligence in Education*, volume 11626, pages 185–189. Springer International Publishing, Cham.

- [McAlpine and Amundsen, 2009] McAlpine, L. and Amundsen, C. (2009). Identity and agency: Pleasures and collegiality among the challenges of the doctoral journey. *Studies in Continuing Education*, 31(2):109–125.
- [Milicev et al., 2021] Milicev, J., McCann, M., Simpson, S. A., Biello, S. M., and Gardani, M. (2021). Evaluating Mental Health and Wellbeing of Postgraduate Researchers: Prevalence and Contributing Factors. *Current Psychology*.
- [Mittelmeier et al., 2021] Mittelmeier, J., Rienties, B., Zhang, K. Y., and Jindal-Snape, D. (2021). *Using Diaries in Mixed Methods Designs*, pages 15–28. Routledge, first edition.
- [Molenaar, 2021] Molenaar, I. (2021). Personalisation of learning: Towards hybrid human-AI learning technologies. *OECD Digital Education Outlook 2021 Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots: Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots*, pages 57–77.
- [Ochoa, 2017] Ochoa, X. (2017). Multimodal learning analytics. In *Handbook of Learning Analytics*, pages 129–141. Society for Learning Analytics Research.
- [Pishtari, 2021] Pishtari, G. (2021). *Aligning Learning Design and Analytics in Mobile Learning*. PhD thesis, Tallinn University.
- [Poduthase, 2015] Poduthase, H. (2015). Rigor in Qualitative research: Promoting quality in Social Science Research. *Research Journal of Recent Sciences*, 4:25–28.
- [Prestigiacomio et al., 2020] Prestigiacomio, R., Hadgraft, R., Hunter, J., Locker, L., Knight, S., van den Hoven, E., and Martinez-Maldonado, R. (2020). Learning-centred translucence: An approach to understand how teachers talk about classroom data. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, pages 100–105, Frankfurt Germany. ACM.
- [Prieto et al., 2022] Prieto, L., Odriozola-González, P., Rodríguez-Triana, M. J., Dimitriadis, Y., and Ley, T. (2022). Progress-Oriented Workshops for Doctoral Well-being: Evidence From a Two-Country Design-Based Research. *International Journal of Doctoral Studies*, 17:39–66.
- [Prieto et al., 2021] Prieto, L. P., Rodríguez-Triana, M. J., Ley, T., and Eagan, B. (2021). The Value of Epistemic Network Analysis in Single-Case Learning Analytics: A Case Study in Lifelong Learning. In Ruis, A. R. and Lee, S. B., editors, *Advances in Quantitative Ethnography*, volume 1312, pages 202–217. Springer International Publishing, Cham.
- [Prieto-Alvarez et al., 2018] Prieto-Alvarez, C. G., Martinez-Maldonado, R., and Buckingham Shum, S. (2018). Mapping learner-data journeys: Evolution of a visual co-design tool. In *Proceedings of the 30th Australian Conference on Computer-Human Interaction*, pages 205–214, Melbourne Australia. ACM.
- [Rigler et al., 2017] Rigler, K. L., Bowlin, L. K., Sweat, K., Watts, S., and Throne, R. (2017). Agency, Socialization, and Support: A Critical Review of Doctoral Student Attrition. Technical report.
- [Rodríguez-Triana et al., 2018] Rodríguez-Triana, M. J., Prieto, L. P., Martínez-Monés, A., Asensio-Pérez, J. I., and Dimitriadis, Y. (2018). The teacher in the loop: Customizing multimodal learning analytics for blended learning. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, pages 417–426, Sydney New South Wales Australia. ACM.
- [Ruiz et al., 2016] Ruiz, F. J., Hernández, D. R., Falcón, J. C. S., and Luciano, C. (2016). Effect of a one-session ACT protocol in disrupting repetitive negative thinking: A randomized multiple-baseline design. *International Journal of Psychology and Psychological Therapy*, 16(3):213–233.
- [Satinsky et al., 2021] Satinsky, E. N., Kimura, T., Kiang, M. V., Abebe, R., Cunningham, S., Lee, H., Lin, X., Liu, C. H., Rudan, I., Sen, S., Tomlinson, M., Yaver, M., and Tsai, A. C. (2021). Systematic review and meta-analysis of depression, anxiety, and suicidal ideation among Ph.D. students. *Scientific Reports*, 11(1):14370.

- [Sellés et al., 2015] Sellés, P. M., Tomás, M. T. C., Costa, J. A. G., and Mahía, F. C. (2015). Predictors of weekly alcohol drinking and alcohol-related problems in binge-drinking undergraduates. *Adicciones*, 27(2):119–131. Publisher: Sociedad Científica Española de Estudios sobre el Alcohol, el Alcoholismo y
- [Selwyn, 2019] Selwyn, N. (2019). *Should robots replace teachers?: AI and the future of education*. John Wiley & Sons.
- [Shaffer, 2017] Shaffer, D. W. (2017). *Quantitative Ethnography*. Lulu.com.
- [Shaffer and Ruis, 2017] Shaffer, D. W. and Ruis, A. (2017). Epistemic network analysis: A worked example of theory-based learning analytics. In *Handbook of Learning Analytics*, pages 175–187. Society for Learning Analytics Research.
- [Shaffer and Serlin, 2004] Shaffer, D. W. and Serlin, R. C. (2004). What good are statistics that don't generalize? *Educational researcher*, 33(9):14–25.
- [Shibani et al., 2019] Shibani, A., Knight, S., and Shum, S. B. (2019). Contextualizable Learning Analytics Design: A Generic Model and Writing Analytics Evaluations. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, pages 210–219, Tempe AZ USA. ACM.
- [Shneiderman, 2020] Shneiderman, B. (2020). Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. *International Journal of Human-Computer Interaction*, 36(6):495–504.
- [Siemens, 2012] Siemens, G. (2012). Learning analytics: Envisioning a research discipline and a domain of practice. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, pages 4–8. ACM.
- [Slade and Prinsloo, 2013] Slade, S. and Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*, 57(10):1510–1529.
- [Smelser, 2013] Smelser, N. J. (2013). *Comparative Methods in the Social Sciences*. Quid Pro Books.
- [Strosahl et al., 2012] Strosahl, K., Robinson, P., and Gustavsson, T. (2012). *Brief Interventions for Radical Change: Principles & Practice of Focused Acceptance & Commitment Therapy*. New Harbinger Publications, Oakland, CA.
- [Sverdlik et al., 2018] Sverdlik, A., C. Hall, N., McAlpine, L., and Hubbard, K. (2018). The PhD Experience: A Review of the Factors Influencing Doctoral Students' Completion, Achievement, and Well-Being. *International Journal of Doctoral Studies*, 13:361–388.
- [Sverdlik and Hall, 2020] Sverdlik, A. and Hall, N. C. (2020). Not just a phase: Exploring the role of program stage on well-being and motivation in doctoral students. *Journal of Adult and Continuing Education*, 26(1):97–124.
- [Terrell et al., 2012] Terrell, S. R., Snyder, M. M., Dringus, L. P., and Maddrey, E. (2012). A grounded theory of connectivity and persistence in a limited residency doctoral program. *The Qualitative Report*, 17(31):1–14.
- [Tzimas and Demetriadis, 2021] Tzimas, D. and Demetriadis, S. (2021). Ethical issues in learning analytics: A review of the field. *Educational Technology Research and Development*, 69(2):1101–1133.
- [van Gog et al., 2020] van Gog, T., Hoogerheide, V., and van Harsel, M. (2020). The Role of Mental Effort in Fostering Self-Regulated Learning with Problem-Solving Tasks. *Educational Psychology Review*, 32(4):1055–1072.
- [Viera and Garrett, 2005] Viera, A. J. and Garrett, J. M. (2005). Understanding interobserver agreement: The kappa statistic. *Family Medicine*, 37(5):360–363.
- [Wang and Hannafin, 2005] Wang, F. and Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational Technology Research and Development*, 53(4):5–23.

- [Weiser, 2019] Weiser, S. (2019). Requirements of Trustworthy AI. <https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines/1>.
- [Wild and Schulze Heuling, 2020] Wild, S. and Schulze Heuling, L. (2020). Student dropout and retention: An event history analysis among students in cooperative higher education. *International Journal of Educational Research*, 104:101687.
- [Wilder et al., 2020] Wilder, B., Horvitz, E., and Kamar, E. (2020). Learning to Complement Humans. *arXiv:2005.00582 [cs]*.
- [Wise and Jung, 2019] Wise, A. F. and Jung, Y. (2019). Teaching with Analytics: Towards a Situated Model of Instructional Decision-Making. *Journal of Learning Analytics*, 6(2):53–69.
- [Wollast et al., 2018] Wollast, R., Boudrenghien, G., Van der Linden, N., Galand, B., Roland, N., Devos, C., De Clercq, M., Klein, O., Azzi, A., and Frenay, M. (2018). Who Are the Doctoral Students Who Drop Out? Factors Associated with the Rate of Doctoral Degree Completion in Universities. *International Journal of Higher Education*, 7(4):143–156.
- [Yang et al., 2021] Yang, S. J., Ogata, H., Matsui, T., and Chen, N.-S. (2021). Human-centered artificial intelligence in education: Seeing the invisible through the visible. *Computers and Education: Artificial Intelligence*, 2:100008.
- [Zimmerman, 1990] Zimmerman, B. J. (1990). Self-Regulated Learning and Academic Achievement: An Overview. *Educational Psychologist*, 25(1):3–17.
- [Zimmerman and Forlizzi, 2014] Zimmerman, J. and Forlizzi, J. (2014). Research through design in hci. In *Ways of Knowing in HCI*, pages 167–189. Springer.