Learning and Individual Differences 114 (2024) 102499

Contents lists available at ScienceDirect

Learning and Individual Differences

journal homepage: www.elsevier.com/locate/lindif

Idiographic artificial intelligence to explain students' self-regulation: Toward precision education

Mohammed Saqr [a,*], Rongxin Cheng [b], Sonsoles López-Pernas [a], Emorie D Beck [b]

[a] University of Eastern Finland, School of Computing, Joensuu, Yliopistokatu 2, fi-80100 Joensuu, Finland
[b] University of California, Davis, Psychology Department, USA

ARTICLE INFO

Keywords:
Idiographic
Artificial intelligence
Machine learning
Learning analytics
Self-regulated learning
Person-specific
Personalized learning
Explainable AI
Within-person

ABSTRACT

Existing predictive learning analytics models have exclusively relied on aggregate data which not only have obfuscated individual differences but also made replicability and generalizability difficult. This study takes a radical departure and uses a person-specific approach to predicting and explaining students' self-regulation (SRL). A person-specific approach entails developing a predictive model for each individual using their own data (i.e., idiographic, single-subject or $N = 1$). We also use explainable and interpretable artificial intelligence (AI) models that allow us to identify the variables that explain students' SRL and guide data-informed decisions. Our study has shown that idiographic single-subject models are tenable, informative, and can accurately capture the individualized students' SRL process. Predictions varied vastly across students regarding accuracy and predictors. The traditional average model did not match any student regarding the predictors' order. These findings are a testament that the "average" is rare and often does not represent any individual student. The variability in our study has shown that no single model can accurately and reliably capture all students. To account for the unique learning processes of individual students, idiographic methods could provide a solution.

Educational relevance statement: Individualized artificial intelligence is feasible and reliable and can help understand each person using their own data. Using idiographic models, we can deliver solutions that are precise, accurate and interventions that are more likely to work.

1. Introduction

Learning analytics emerged at the outset of the last decade to harness the opportunities created by the abundance of students' data. The premise was that big data with massive computing power and advanced methods would bring major breakthroughs and help deliver real-life impact (Siemens, 2013). A large number of learning analytics applications have been developed ever since; among those, predictive analytics. Researchers would collect logs of students' online behavior —e.g., views of learning resources, social interactions, and assessment work— to build a predictive model to flag potential underachievers (Macfadyen & Dawson, 2010; Shafiq et al., 2022). Identifying an underachieving student early in a course paves the way for proactive intervention (Saqr et al., 2017). Several studies have reported successful prediction of underachievers within convenience samples (i.e., samples with conveniently accessible data) (Ifenthaler & Yau, 2020). However, replicating or applying these findings in other contexts (i.e., portability) has been disappointing (Gašević et al., 2016). Such difficulty in replication was attributed to contextual variabilities, learning design, and more importantly to individual differences (Saqr et al., 2022). This is because existing predictive models rely on aggregate data collected from *different* students that obfuscate individual differences and personal variabilities (Bobrowicz et al., 2024; Saqr et al., 2024). In doing so, the aggregate models perform well in general but not in any particular case. These shortcomings call for a paradigm shift that ***prioritizes the person*** where the learning process takes place to create individualized person-specific models (Beltz et al., 2016; Saqr & Lopez-Pernas, 2021). It stands to reason that if the ***person*** is our concern, then, the methods we should use are person-specific (Molenaar & Campbell, 2009), which remains a gap that our study aims to fill.

This study takes a radical departure from the existing status quo and uses a person-specific approach to predicting and explaining students' self-regulation. Self-regulation significantly enhances student performance and interventions promoting self-regulation have shown a positive impact on academic achievement across various settings, including classrooms, online environments, and workplaces (J. Broadbent & Poon,

* Corresponding author.
E-mail address: mohammed.saqr@uef.fi (M. Saqr).

https://doi.org/10.1016/j.lindif.2024.102499
Received 6 February 2024; Received in revised form 22 June 2024; Accepted 1 July 2024
Available online 6 July 2024
1041-6080/© 2024 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

2015; Heikkinen et al., 2022). A person-specific approach entails developing a predictive algorithm for each individual student using their own data (i.e., idiographic models) (Beck & Jackson, 2022). By idiographic models, we mean single-subject ($N = 1$) machine learning models trained exclusively on each individual's data where the results of the analysis are specific to the modeled individual (Soyster et al., 2021). Building predictive algorithms at the resolution of the single student paves the way for precision education which puts the power of person-specific artificial intelligence in the hands of learners (Cook et al., 2018). We use explainable AI models to reveal the factors that explain student's self-regulation, and thereby could guide data-informed decisions (Khosravi et al., 2022). Further, we assess the similarities and differences between individual student models (idiographic models) and traditional machine learning models (models created from all aggregate data). The research questions of this study are:

RQ1: (a) To what extent can idiographic machine learning models predict students' self-regulation of effort, metacognition, motivation, and enjoyment? (b) How much does the prediction vary across different students and models?

RQ2: What are the most important predictors of learners' self-regulation and how do they differ among students?

RQ3: How do individual machine learning models compare to traditional models? In other words, to what extent does each idiographic model (single subject) compare to the traditional model (all data) in terms of explanatory variables?

The remainder of this study is structured as follows. The next section is the Background of the study and discusses the challenges of prediction in education and offers an overview of idiographic methods, how they differ from existing approaches as well as group-to-individual generalizability which is —at least partially— implicated in existing challenges in prediction. The Background section also establishes the theoretical grounding of the aims of the current study and presents a review of previous research. The subsequent section describes the methodology followed in the study, including the participants, the data collection, and the data analysis. Next is the Results section, followed by a Discussion section, the Limitations of the study, and a Conclusions section with some closing remarks.

2. Background

2.1. Challenges in predictive learning analytics

Studies examining the portability of algorithms across contexts found that —even within the same program— courses have different predictors that are not shared by any two courses (Finnegan et al., 2008; Gašević et al., 2016). Studies using data from the same program, with the same pedagogical and curricular design found that some predictors were relatively consistent across course iterations (e.g., session count), while others yielded different results between iterations of the same course offering (Saqr et al., 2022). One year, the predictor was positively associated with students' performance and, another, such association was non-existent.

The failure to replicate or transfer results across contexts is not limited to education. In fact, several large-scale replications have raised concerns about generalizability in social science research in what is referred to as a “replicability crisis” (Hagger et al., 2016; Klein et al., 2018; Open Science Collaboration, 2015; Plesser, 2018). Researchers have cited the “usual suspects” and pitfalls in research rigor, methodological issues, contextual differences, and publication bias among others (Hagger et al., 2016; Hernández-García et al., 2024; Open Science Collaboration, 2015). Most importantly, individual differences, heterogeneity, and lack of group-to-individual generalizability (i.e., how results from group-based research generalize to individuals) have been recognized as major contributing factors to the elusive generalizability (Bryan et al., 2021; Jovanović et al., 2021). Addressing these issues may help improve research results, individualize learning, and offer precise educational solutions (Hamaker, 2012; Saqr & López-Pernas, 2024). In our study, our focus centers on group-to-individual generalizability and we aim to provide solutions that entail the creation of individualized person-specific models.

2.2. Group-to-individual generalizability

Research is commonly performed using aggregated data from a group of individuals (often referred to as nomothetic). The tacit assumption is that, if we use a representative sample, inferences made from the aggregated group-based data can capture the population and therefore, can be used to generate laws and norms that apply to everyone (Hamaker, 2012; Valsiner et al., 2009). Statistically speaking, such inferences assume that capturing inter-individual variance across many individuals “yield the same results as an analysis of intra-individual variation” (Molenaar & Campbell, 2009, p. 112). For this assumption to hold, the examined phenomenon must be ergodic, that is, equivalent across individuals (i.e., invariant and homogeneous across the population) and stable over time (stationary without temporal progression) (Molenaar & Campbell, 2009). By equivalence, we mean that the distribution of the studied variables is similar for each person. Then, it does not matter *which* individuals we measure, they all have a similar data distribution generated by the same generating mechanism. By the same token, the stationarity criterion entails that it does not matter *when* we measure an individual, the distribution of the variable (i.e. mean and variability) is stable over time.

Both of the aforementioned assumptions —stationarity and homogeneity— are empirically unattainable (Fisher et al., 2018; Molenaar & Campbell, 2009). We know that psychological phenomena —like learning— vary across the population and continuously change over time. A vast body of empirical evidence has refuted such assumptions with a multitude of examples across the years (Molenaar, 2004; Richters, 2021; Valsiner et al., 2009). For instance, Fisher et al. (2018) found significant differences between intra-individual and inter-individual variance in six different samples with six different experiments. The authors concluded that “the temptation to use aggregate estimates to draw inferences at the basic unit of social and psychological organization—***the person***—is far less accurate or valid than it may appear in the literature” (p. E6113). The authors also added that the lack of group-to-individual generalizability poses a credible threat to research at large (Fisher et al., 2018). Thereupon, inferences made with group-based insights hardly —if at all— “apply to each and every individual in the population, or even to a majority of the individuals in a population” (Hamaker, 2012, p. 43). Assuming that what is true for the population in aggregate is true for the individuals is often referred to as an ecological fallacy (Fisher et al., 2018). It is therefore necessary to model the *individual* if our aim is to devise or generate person-specific insights.

In contrast to group-level research based on aggregate data, idiographic research uses a person-specific approach to capture individual human processes based on single-subject data (Beck & Jackson, 2021; Molenaar, 2004). Idiographic methods achieve this goal by collecting several measurement points from the same individual that are enough to capture his/her intra-individual processes. The analysis is then performed for each individual separately. Given that this data is person-specific, the analysis reflects precisely and accurately the distinct person processes based on their particularities.

This is not to imply that group-level population studies are useless. In fact, both group-level and idiographic methods serve different purposes. On the one hand, group-level or traditional statistics allow the inference of the general trends, e.g., what works for a considerable proportion of the population. Yet, we cannot know exactly whom —among the population— the results apply to. On the other hand, idiographic methods allow us to understand the individual phenomena or what works for the very individual we are studying. Idiographic models cannot — and strictly so— claim generalizability beyond the studied individual (Howard & Hoffman, 2018).

The illustration in Fig. 1A offers a simplified representation of inter-individual data from different people (i.e., between-person group-based or nomothetic) and Fig. 1B shows intra-individual (within-person, idiographic or single subject) variance collected from the same person over time (Fig. 1B). Please note that all idiographic models ($N = 1$) rely on within-person variance and therefore, we use both terms interchangeably in our study. Yet, within-person variance can also refer to studies with $N > 1$ (e.g., Saqr, 2023b).

2.3. Previous research on within-person processes

Despite the centrality of the "individual" or the "person" and the importance of the individual mechanisms of learning, there has been a "neglect" of within-person research in education (Murayama et al., 2017). Existing research —although relatively scarce— has focused on studying the within-person processes to optimize between-person models but not the true person-specific idiographic (N = 1) models. For example, using multilevel models (MLM) Trautwein et al. (2009) found a positive relationship between doing homework and school performance. Similarly, Martin et al. (2015) used MLM to study the momentary within-person interplay between engagement and motivation emphasizing the importance of temporal variation across time. Similar other studies have investigated the within-person variation of engagement (Bakker et al., 2015), emotional experiences (Webster & Hadwin, 2015), memory variations, and motivations, among others (Collie et al., 2020; L. E. Malmberg et al., 2013; Martin et al., 2020). Other methods, too, were used, like multilevel structural equation modeling to study the interplay between effort regulation and task values (Dietrich et al., 2017). While the aforementioned methods have brought valuable insights into the within-person variability, they do not capture the person-specific idiographic variance but they are rather "informed by and shrunk toward group-level averages" (Beck & Jackson, 2021). In other words, they are group-based research optimized to better capture the variations within large samples.

Recently, within-person research has emerged in education. Examples include studies in which intensive longitudinal data has been used to capture the within-person variations in SRL (J. Malmberg et al., 2022; Saqr & López-Pernas, 2024). Saqr and Lopez-Pernas (2021) used dynamic networks to capture the daily variation in a single student over a whole month. Recently, Saqr (2023a) used within-person variations of engagement to predict students' performance and reported that augmenting traditional MLM with within-person models vastly improved predictive performance. Another study by the same author used psychological networks to compare average within-person networks with traditional group-based networks (Saqr, 2023b). The author reported that group-based engagement —while sharing similarities with within-person engagement patterns— shows stark differences across indicators. To that end, it is clear that within-person research is lacking, especially when it comes to predicting and explaining individual students' behavior. More importantly, research using $N = 1$ idiographic models is non-existent.

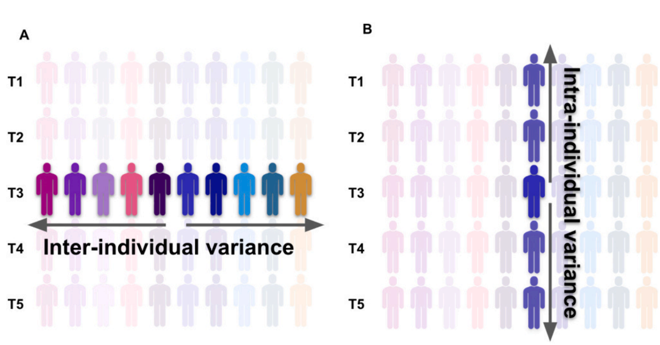

Fig. 1. A) Inter-individual variance measures across individuals (nomothetic) at the T3 time point. **B)** Intra-individual variance (within-person) measures from the same individual across different time points (idiographic).

2.4. Theoretical underpinning

This study is grounded within the framework of self-regulated learning (SRL). Self-regulated learning may be defined as "an active, constructive process whereby learners set goals for their learning and then attempt to monitor, regulate, and control their cognition, motivation, and behavior, guided and constrained by their goals and the contextual features in the environment" (Pintrich, 2000). The self-regulation process as the name implies revolves around the quintessential role of "self" whereby students regulate, control and adapt their learning. Further, self-regulation emphasizes self-directedness, self-control, autonomy, and the agency of the learner to make decisions about their goals, to choose the means of achieving such learning goals, and later to self-evaluate and adjust (Winne, 1996). Students' self-awareness of their learning enhances their learning and control (Panadero, 2017).

Despite the emphasis on the individual in SRL and how the individual changes (controls, adjusts, or adapts), most of the existing research —if not all— has modeled aggregate group-level processes. In this study, we take a different approach that implements individualized models to predict and explain the person-specific processes. Namely, we examine the possibility of predicting students' effort regulation. Effort was chosen given its centrality as a determinant of task execution, students' learning, and its tight relationship with learning achievement (Biwer et al., 2023; Jaclyn Broadbent et al., 2023; Trautwein et al., 2009). Besides, we explore the possibility of predicting other elements of self-regulation and determinants of achievements, namely, motivation, enjoyment, and meta-cognition (Heikkinen et al., 2022).

3. Method

3.1. Participants

The participants of this study were students attending the last year of their secondary school in an international school in Finland. Students studied mathematics, science, languages, and social sciences among others. A total of 41 students were invited to the study during an in-class presentation; a total of 21 agreed to contribute for the duration of the study (45 days) by responding to a survey twice a day: once at school during the regular classrooms with permission from their teachers, and once at home (see next subsection for details). Weekends were excluded as they did not attend school. Three students dropped out and we excluded one student who had <30 responses. The remaining 17 participants responded to the survey a total of 821 times. The number of times each individual completed the survey ranged from 32 to 63 across the 45 days. Whereas the sample size may seemingly be small, it is considered adequate in within-person design in which the number of repeated measurements for every student is what matters (Maas & Hox, 2005; Martin et al., 2015; Raudenbush & Bryk, 2002).

3.2. Measurement

The data collection instrument was designed to capture ecologically valid SRL data, i.e., SRL as it occurs in real life (school during lessons and home during studying). Several SRL instruments and protocols exist (e.g., questionnaires, log data, and interviews) with different temporal granularities (Jaclyn Broadbent et al., 2023; Panadero, 2017; Rovers et al., 2019). These existing instruments typically capture the stable aptitude of SRL using cross-sectional surveys that have items that read like "I always", or "I regularly" or use the present tense to reflect an enduring habit like "I learn from my mistakes". These instruments fall short of capturing the momentary or daily changes in SRL and are considered suboptimal if applied to daily data (Flake & Fried, 2020; Hall et al., 2021).

The data were collected using ecological momentary assessment (EMA) which is an intensive data collection technique wherein students respond to a questionnaire several times a day. The instrument was constructed in a way that each of the questions covers a construct (or single item measure) (e.g., motivation, goal-setting, or enjoyment) (L. E. Malmberg et al., 2016, 2013; Martin et al., 2015). This is a standard EMA research practice where, for instance, resilience (Gilmore et al., 2019), well-being (Cheung & Lucas, 2014), burnout (Dolan et al., 2015), and self-efficacy (Hoeppner et al., 2011) are measures through single item questionnaires. Other instruments capture elements of engagement using single-item constructs (e.g., Manwaring et al., 2017; Martin et al., 2015; Xie et al., 2019). Several studies have shown that these methods are as "reliable as their multi-item counterparts" (Allen et al., 2022, p. 4). This is particularly true when "a construct is unambiguous or narrow in scope" (Allen et al., 2022, p. 1). More importantly, short surveys are more likely to encourage compliance and do not overburden students the way long surveys may do over multiple applications per day. Whereas there are some existing short surveys, they are validated for aptitude cross-sectional between-person data that are therefore inappropriate for our context (Flake & Fried, 2020; Hall et al., 2021).

Based on the aforementioned considerations, we built *the Concise SRL Survey* where every question captures a single construct or in other words, a collection of single-item questions that each captures a single construct (available from Saqr & López-Pernas, 2024). The *Concise SRL Survey* was built after an extensive review of nine existing SRL surveys (Jaclyn Broadbent et al., 2023; Jansen et al., 2018; Pichardo et al., 2014; Pintrich & De Groot, 1991; Rovers et al., 2019). A twelve-item questionnaire was selected to cover the main phases of SRL planning, monitoring, effort, regulation, organizing, help-seeking, environment, time management, applying feedback, and self-evaluation, as well as emotions (anxiety and enjoyment) and motivation. Yet, some items were highly correlated and therefore, were combined into single items: planning and monitoring, effort and effort regulation, and feedback and self-assessment. The final items included: *Effort* (Effort + Regulation), *Metacognition* (Feedback + Evaluating), *Motivation*, *Enjoyment*, *Planning* (Planning + Monitoring), *Environment*, *Help*, *Organizing*, and *Anxiety* (Table S1; please refer to Saqr and López-Pernas (2024) for the full instrument).

To ensure the content of *Concise SRL Survey* is validated, we performed content validity following the COSMIN methodology (a method for evaluating and ensuring the quality of research instruments) with a group of domain experts (five researchers) who assessed the contents of the questionnaire regarding relevance (to the construct of interest, target population, context, recall period, and options), the coverage of the key concepts, and the comprehensiveness (regarding the items coverage, instructions, and wording) (Terwee et al., 2018). Furthermore, the questionnaire was administered to a class of 14 students —who are similar in age and in studies to our sample— who gave their opinions and input about the ease of response, relevance, understanding, and ambiguity of items, the possibility of distress or judgment to establish the face validity of the questionnaire (Allen et al., 2022).

The questions in the *Concise SRL Survey* were presented on a scale of 1–100 to capture daily variations as recommended in EMA studies (Shiffman et al., 2008; Wright & Zimmermann, 2019). The *survey* was distributed using an EMA-specialized app (Avicenna, https://avicennaresearch.com) that allows notifications to be sent at preset times (in our case 10:15 for school and 17:30 for home with minor variations). The scheduled times were chosen based on a conversation with the students and the teachers of the school. Furthermore, students were given a demonstration of the app and their questions were answered. Each student signed a detailed informed consent, and their parents and teachers were informed about the procedure.

Besides the questionnaire items, we created time features to account for the relationship between students' responses in SRL variables and response time. The time features include two dummy codes for the time of day (morning and evening) and five for the weekdays (Monday through Friday). Next, we created cyclical trend variables by computing the cumulative time (in hours) from the first beep (i.e., response) to create linear, quadratic, and cubic time trends as well as one- and two-period sine and cosine functions across each 24-h period. These features were included to account for response time variations—which were not constrained by design— but will not be discussed in the present article due to their low value in explaining students' SRL. Interested readers may refer to the supplemental materials Fig. S1–S3 to compare the feature importance of psychological factors with that of time features.

3.3. Outcome variables

We use self-regulation to guide the data collection and analysis. In particular, our study explores the possibility of predicting key determinants of learners' performance based on their self-regulation using idiographic models. The reliance of our study on intensive daily measurements made it impractical to obtain students' grades twice daily as an outcome measure (note that an outcome needs to match the frequency of the survey). Therefore, we opted for outcomes that are both strongly associated with learning outcomes and reflect self-regulation and investment in learning tasks (Heikkinen et al., 2022; Panadero, 2017; Schneider & Preckel, 2017). Given that predictions at the idiographic level are still new, it was decided to test different SRL behaviors to compare their predictability. As such the outcomes were: effort regulation (henceforth *effort* for brevity), *metacognition*, *motivation* and *emotions*. In particular, effort was chosen because it is an important measure of the energy that students invest in learning or studying. The choice was also motivated by the centrality of effort regulation as a determinant of students' performance (Cole et al., 2008; Schneider & Preckel, 2017). Further, effort is an observable and malleable disposition that can be used as a target of support (Biwer et al., 2023). Targeting effort for support plays a "substantial role in increasing students' achievement" (Biwer et al., 2023; Stewart, 2008; Trautwein et al., 2006). Additionally, we also explore the possibility of predicting other key elements of self-regulation, namely, *motivation*, *enjoyment*, and *metacognition*. We chose motivation because of its importance in driving students' self-regulation and its importance in planning, achievement of goals, and learning outcomes (Pintrich, 2000; Schneider & Preckel, 2017; Urdan & Kaplan, 2020). Similarly, we also attempt to predict *metacognition* as a higher-order thinking skill that is tightly linked to cognitive engagement and applying deep-learning strategies. Lastly, we selected *enjoyment* as representative of emotion.

3.4. Descriptive statistics

Table 1 presents the descriptive statistics (mean and standard deviation) of each construct per student. For each student, the mean and standard deviation of each outcome were computed across all their responses collected throughout the study for each outcome. We see students who are fairly constant, with small standard deviations across all outcomes (e.g., Participant 2); students who are highly variable, with high standard deviations across all outcomes (e.g., Participant 4), and students who are stable in some outcomes and fluctuating in others (e.g., Participant 17).

3.5. Data analysis

The data analysis included applying machine learning techniques to investigate: (a) the predictability of four SRL variables (*effort*, *metacognition*, *motivation*, *enjoyment*) at the individual level (Fig. 2 left) (RQ1a), (b) the variations in predictability among individuals (RQ1b), (c) the most influential features at the individual level and group (RQ2), and (d) the extent of individual differences in the importance of these features and how they compare to the traditional model (RQ3).

To answer RQ1, we predicted the key SRL variables (*effort*,

Table 1
Descriptive statistics of each participant: Mean (Standard Deviation).

Ppant.	N	Effort	Metacognition	Motivation	Enjoyment	Planning	Help	Environment	Organizing	Anxiety
1	62	54.9 (11.4)	63.0 (5.5)	47.4 (22.7)	47.6 (18.8)	62.9 (7.8)	85.6 (9.0)	55.6 (13.4)	83.3 (10.0)	74.5 (11.9)
2	59	87.2 (5.2)	89.4 (4.6)	88.2 (4.7)	88.1 (4.9)	85.6 (5.3)	87.4 (5.1)	88.7 (5.2)	88.0 (5.5)	88.7 (5.9)
3	61	41.5 (22.6)	22.0 (21.6)	50.9 (23.2)	39.5 (23.8)	47.3 (24.6)	52.5 (24.5)	49.5 (26.0)	47.2 (26.7)	25.6 (22.0)
4	59	68.2 (33.9)	64.1 (12.6)	45.7 (31.7)	75.6 (28.4)	88.3 (23.7)	92.2 (23.9)	89.7 (23.1)	84.9 (18.5)	32.1 (33.5)
5	57	34.1 (25.5)	19.7 (15.1)	27.9 (26.8)	30.4 (25.5)	32.2 (27.0)	58.2 (38.7)	79.4 (31.1)	25.5 (23.0)	26.2 (34.9)
6	57	53.7 (18.9)	59.4 (9.1)	54.0 (16.3)	53.4 (15.1)	66.1 (13.0)	92.3 (7.1)	87.6 (11.0)	72.5 (9.4)	77.0 (7.5)
7	54	45.1 (13.9)	46.8 (6.9)	27.9 (14.4)	39.0 (14.4)	62.7 (10.3)	100.0 (0.0)	74.5 (9.3)	71.7 (9.3)	71.1 (17.9)
8	50	49.5 (19.5)	55.7 (16.8)	45.4 (25.5)	52.2 (21.3)	51.2 (19.0)	63.4 (17.3)	66.6 (15.5)	56.3 (18.1)	41.4 (16.0)
9	46	66.5 (7.5)	73.4 (6.1)	72.0 (9.0)	74.6 (9.1)	74.9 (6.5)	94.1 (3.0)	92.8 (5.1)	73.5 (6.9)	81.3 (5.1)
10	44	60.6 (17.6)	47.9 (15.7)	33.5 (22.4)	43.5 (19.4)	59.8 (17.7)	25.9 (25.1)	67.3 (20.3)	57.6 (19.1)	28.0 (24.2)
11	43	39.8 (21.3)	54.5 (15.8)	29.6 (20.9)	27.3 (18.9)	42.9 (19.9)	67.7 (31.2)	38.9 (34.4)	45.1 (17.2)	56.5 (18.5)
12	43	66.5 (15.9)	54.4 (13.3)	70.9 (14.3)	63.7 (16.9)	71.4 (11.2)	54.2 (16.2)	75.8 (12.4)	72.9 (10.5)	65.8 (24.8)
13	42	60.1 (14.3)	60.8 (11.9)	57.1 (17.5)	58.9 (14.5)	64.0 (14.0)	66.9 (13.2)	66.7 (14.5)	68.2 (11.4)	61.1 (14.1)
14	40	63.1 (17.1)	55.5 (15.3)	61.3 (20.8)	55.7 (19.7)	65.8 (17.0)	57.1 (18.1)	73.2 (23.0)	64.6 (20.6)	40.4 (36.5)
15	35	82.4 (15.0)	89.9 (9.4)	68.1 (33.4)	82.3 (25.4)	88.0 (10.8)	87.5 (13.9)	85.8 (15.5)	89.6 (10.5)	24.0 (27.9)
16	32	42.6 (17.8)	43.6 (9.4)	38.1 (18.2)	49.8 (13.5)	50.2 (11.2)	61.4 (12.0)	52.8 (16.5)	49.9 (15.6)	68.2 (10.0)
17	33	75.3 (33.1)	99.3 (2.7)	65.6 (37.5)	67.6 (32.6)	91.7 (18.1)	99.7 (1.7)	82.0 (21.1)	50.5 (29.1)	98.8 (4.8)

Note. Ppant. = Participant. N = Total number of responses.

Fig. 2. An illustration of the analytical plan: On the left, P1 to P7 are participants, for each and every one of them, we operationalize three machine learning algorithms: Random Forest (RF), BISCWIT (Bis) and Elastic Net Regression (ENR). On the right we use the whole sample, for which we use the three aforementioned machine learning algorithms in a traditional general way, by averaging the data throughout the whole study for each student.

metacognition, *motivation*, and *enjoyment*) using the items of the EMA questionnaire: *effort*, *metacognition*, *motivation*, *enjoyment*, *planning*, *environment*, *help*, *organizing*, and *anxiety*. Each outcome variable predicted was omitted from its list of predictors. For instance, to predict *effort* we used all predictors except for *effort*, that is, *metacognition*, *motivation*, *enjoyment*, *planning*, *environment*, *help*, *organizing*, and *anxiety*. Missing data were imputed using the R package *Amelia* (Honaker et al., 2011). Due to the possibility of autocorrelation, correlations between variables were tested for each student and variables with high correlations were removed.

Guided by prior research that performed prediction using data captured through EMA and related techniques (Barrigón et al., 2017; Beck & Jackson, 2022; Hart et al., 2022; Kaiser et al., 2021; Soyster et al., 2021), we used three prevalent machine learning regression models: (a) random forest models (Kim et al., 2019), (b) elastic-net regression (ENR; Friedman et al., 2010), and (c) the best-items scale that is cross-validated, correlation-weighted, informative, and transparent (BISCWIT; Elleman et al., 2020). RF is an ensemble machine learning method that combines the predictions of several decision trees and therefore it is able to capture complex nonlinear relationships and handle high-dimensional data. ENR is a regularization technique that is effective in providing feature selection in the presence of high-dimensional data with potential multicollinearity (such as SRL dimensions) (Soyster et al., 2021). BISCWIT is a correlation-based machine learning technique that offers a simpler alternative to complex machine learning models while offering more parsimonious models with comparable —or sometimes even higher— performance especially for small samples and effects in the presence of high measurement error (Elleman et al., 2020; Kaiser et al., 2021). We used the feature selection methods that are most appropriate for each model (Barredo Arrieta et al., 2020; Soyster et al., 2021). In the RF model, feature importance was estimated using a permutation-based test where importance was quantified by the reduction in model fit before and after random shuffling of feature values. Larger decreases in model fit would suggest greater importance. In the ENR model, feature importance was determined by the absolute regression coefficients. In the BISCWIT model, feature importance was determined by the absolute correlation coefficient between the feature and the outcome variable.

For the RF model, we used the R package *parsnip* (Kuhn & Vaughan, 2023). We set the mode to regression, the engine to "ranger", the number of trees to 1000, and specified 20 candidate parameter sets for tuning the model. For the ENR model, we also used *parsnip*; we set the mode to regression, the engine to "glmnet", and specified 10 values for penalty and 10 values for mixture to make the regular grid. For the BISCWIT model, we used the implementation provided by the R package *psych* (Revelle, 2023). We set the best number of items from 3 to 21 (the largest possible value smaller than the number of predictors, 23) in increments of 3 to increase the speed of computation. We fixed the model parameters to be the same for each individual model and the general model to maximize the comparability of feature selection across individuals. Otherwise, we would not be able to know if the observed differences in the selected features across participants are due to individual differences in self-directed learning or model parameter specification.

Comparing the results generated by the three models provides us with a less biased understanding of the important features predicting each outcome and further allows us to choose the best-performing model. Given that we have several predictors, we chose methods with variable selection procedures and methods for reducing overfitting. In each of the methods, we used k-fold cross-validation and set k to 5. The out-of-sample prediction was evaluated by adjusted R-squared (adjusted RSQ) and root-mean-square-error (RMSE); neither performance metric has well-established cut-offs.

The interpretability and explainability of machine learning models are critical in understanding the learning process, allowing nuanced feedback and insight into how the process works (Barredo Arrieta et al., 2020; Khosravi et al., 2022). Therefore, we used methods to estimate the importance of each feature in each model (RQ2). We fit machine learning models to each participant to obtain idiographic models. An idiographic model is an $N = 1$ model that captures the within-person variance of a single individual using their own data to predict their own behavior (Beck & Jackson, 2022; Lavelle-Hill et al., 2023). In addition, we fit machine learning models to all available data to obtain a between-person general model for comparison using the same technique, algorithms and outcomes for the idiographic models (henceforth referred to as the general model) (RQ3). In the result section, we only reported findings about the machine learning models selected by RMSE. The corresponding findings from models selected by RSQ can be found in the supplementary material Fig. S4 to S8.

4. Results

4.1. *To what extent can idiographic machine learning models predict key elements of the learning process? (RQ1A)*

First, we examined how well idiographic models predicted SRL key behaviors (i.e., *effort*, *metacognition*, *motivation*, and *enjoyment*) using the remaining SRL variables for each person. Expectedly, we found that each of the three machine learning models had different average levels of predictive performance, as demonstrated in Fig. 3 where we show the descriptive statistics and distribution of RMSE and RSQ for each outcome and machine learning algorithm per person. When using RSQ, out-of-sample prediction performance was similar for all three models —albeit slightly higher for ENR ($M_{RSQ} = 0.25$–0.45 for random forest; 0.26–0.49 for ENR; 0.31–0.49 for BISCWIT). In general, the mean RSQ scores of models predicting *effort* were close to 0.50, being the highest among the four tested outcomes. In other words, *effort* was the most predictable outcome in all the models. The average RSQ for other outcomes ranged from 0.3 to 0.4, being lowest for *metacognition* which was the least predictable behavior. Taken together, the average RSQ for all models shows that all outcomes were fairly predictable —on average— and higher most of the time than prior studies (e.g., 0.27; Soyster et al., 2021). Using RMSE to evaluate out-of-sample prediction, the ENR models outperformed random-forest and BISCWIT models ($M_{RMSE} = 0.98$–1.1 for ENR; 11.43–18.59 for random forest; 13.71–22.46 for BISCWIT).

4.2. *Individual variations in model predictability differ across machine learning methods (RQ1B)*

RSQ had a wide range across each predicted SRL behavior for each student and model. For instance, when predicting *effort* using random forest, RSQ ranged from 0.07 to 0.95 across students, similar results were also found for ENR (range 0.02–0.9) and BISCWIT (range 0.13–0.9). The mean value of RSQ for *motivation* was 0.4 (SD = 0.28) and ranged from 0.04 to 0.94 which shows that the motivation of some of the students was almost perfectly predictable (close to 1) while that of the others was almost impossible to predict (close to 0). *Metacognition* was the overall least predictable for students with an average of 0.26 (SD = 0.19) and ranged from 0 to 0.64. Using the standard deviation to assess the variability of model performance, the ENR models detected less individual variation in RMSE than the other two models ($SD_{RMSE} =$

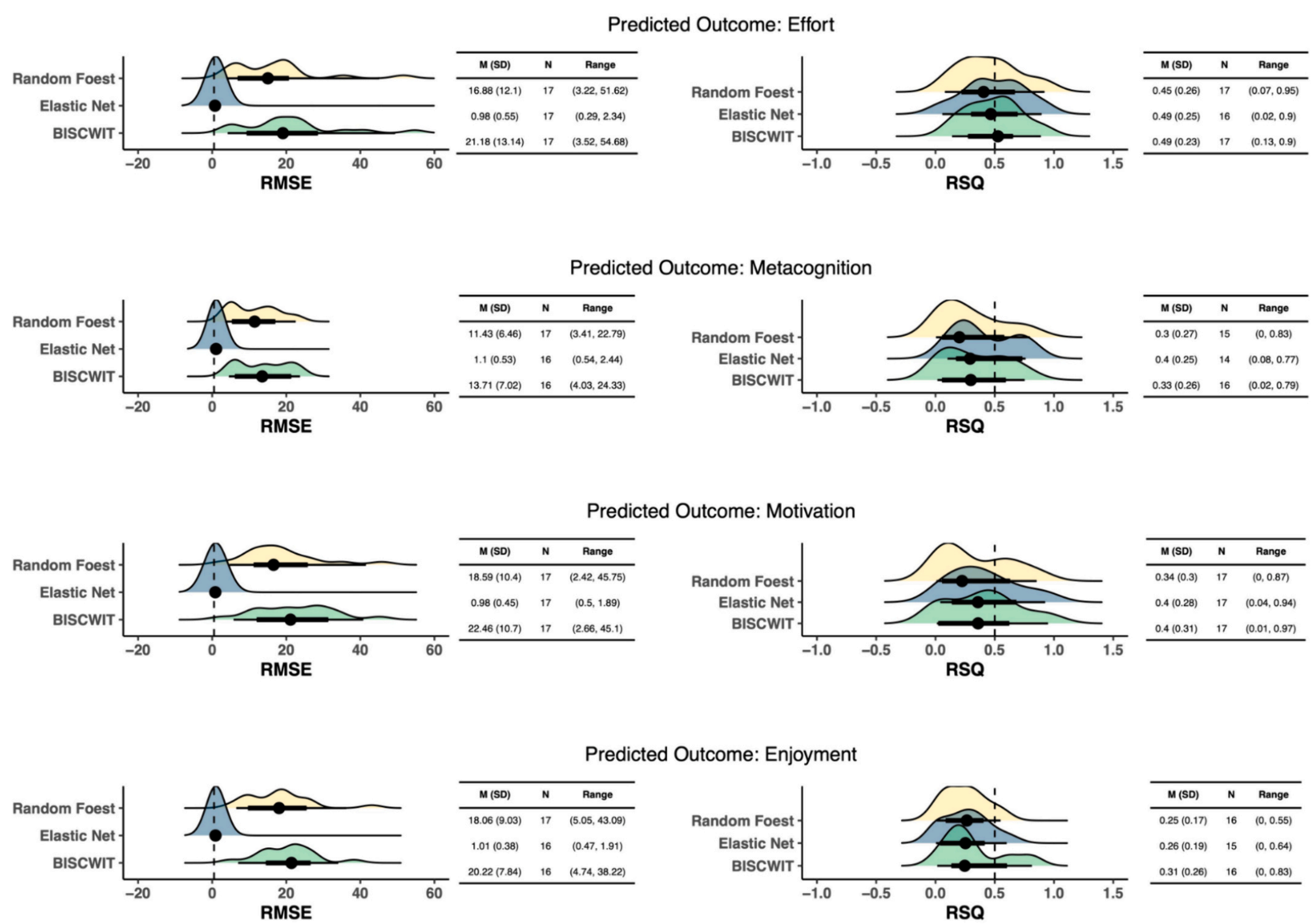

Fig. 3. Plots of RMSE and RSQ as an indication of predictability across students and outcomes. Note: Final model fit index could not be computed in some cases, either because the outcome variable has a constant value in the test data or because none of the features were selected in some data folds.

0.38–0.55 for ENR; SD_{RMSE} = 6.46–12.10 for random forest; SD_{RMSE} = 7.02–13.14 for BISCWIT), as shown in Fig. 3. In comparison, when using RSQ, all three models detected a similar amount of individual variation in model performance (SD_{RSQ} = 0.17–0.30 for random forest; SD_{RSQ} = 0.19–0.28 for ENR; SD_{RSQ} = 0.23–0.31 for BISCWIT).

On the level of individual students, some of them were highly predictable across all SRL behaviors and algorithms while others —a minority— were hard to predict with any model. For instance, using the ENR model, Participant 2 was highly predictable across all outcome variables and algorithms, the RSQ for *Motivation* was 0.91, 0.9 for *Effort*, 0.74 for *Metacognition*, and 0.63 for *Enjoyment*. On the contrary, Participant 5 was fairly unpredictable, whereby the RSQ for *Effort* was 0.02, for *Enjoyment* was 0.003, for *Motivation* was 0.29, and for *Metacognition* was 0.26. Given the superior performance of ENR with RSQ and RMSE, we will use it as the main model going forward to report our findings.

4.3. Individual variation in feature importance (RQ2)

Given that ENR models outperformed the other two algorithms, the presented results will be based on ENR. We present the results of the prediction of *effort* with more space and emphasis as it is the important variable that indicates students' investment in learning and regulation.

Similar to the variability in predictability (RQ1), yet more profoundly, the important explanatory predictors varied for each student, for each outcome, and for each algorithm (RQ2) as illustrated in Fig. 4 which presents the top five predictors for each student. For Participant 10, *planning* was the single most important predictor of *effort* while other predictors (*organizing*, *environment*, *metacognition*, and *anxiety*) contributed little to predicting *effort*. In contrast, for Participant 4, *motivation* was the most important predictor of *effort* and other predictors (*anxiety*, *help*, *planning*, and *organizing*) made important contributions to *effort* prediction. In total, *planning* was the most important predictor of *effort* in 6 students (35.3 %), *motivation* was the most important in 4 students (23.5 %), *anxiety*, *seeking help* and *organizing* were the most important in 2 students (11.8 %), and *metacognition* was for 1 student (5.9 %). Among the 5 most important predictors, *planning* was the most frequent predictor and appeared among the top 5 predictors in 16 out of the 17 students (94.1 %). *Anxiety* was the second most frequent predictor appearing on the top 5 predictors in 14 students 82.3 %, *environment*, and *metacognition* appeared 10 times (58.8 %), *enjoyment*, *help*, and *motivation* were in 9 students (52.8 %), and *organizing* in 8 students (47 %). Most strikingly, the same order of predictors was not shared by any two students for the same outcome and algorithm, pointing out the vast diversity among students.

Furthermore, Fig. 5 presents the percentage of individuals for whom the presented features were in their top five important features that predicted *effort*, *metacognition*, *motivation* and *enjoyment*. Across algorithms, the most commonly shared feature was *planning* in predicting *effort*; *enjoyment* in predicting *metacognition* and *motivation;* and *motivation* in predicting *enjoyment*. These aggregated results from the individual models were relatively close to the general model —the aggregate traditional model which combines data from all students— for the top feature (see Fig. 6). Yet predictors differed significantly for the rest of the features (from the second predictors onwards). For all three algorithms, about only a third of the features were shared by <50 % of all students.

Additionally, we aggregated the feature importance scores of all students in Fig. 6 to compile the *average* importance. Comparing the median value of each density distribution, Fig. 6 suggests that on the aggregate level, *planning* was the most important predictor of *effort*, *enjoyment* was the most important predictor of *metacognition* and *motivation*, and *motivation* was the most important predictor of *enjoyment*.

Fig. 4. The top five features that predicted *Effort* for each participant using the ENR algorithm for each student. Please note that no two students share the same order.

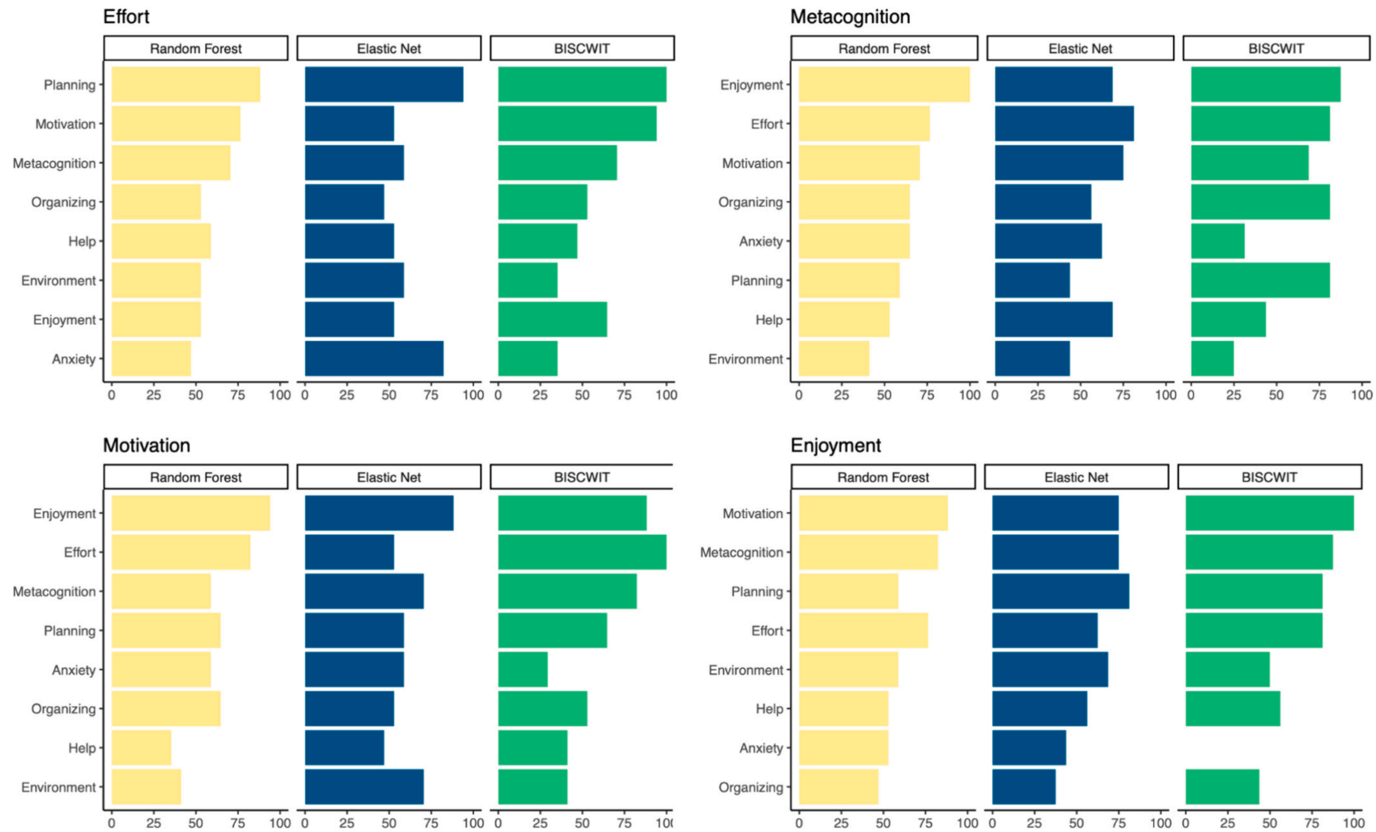

Fig. 5. Percentage of individuals for whom the presented features were in the top five important features in their individual models.

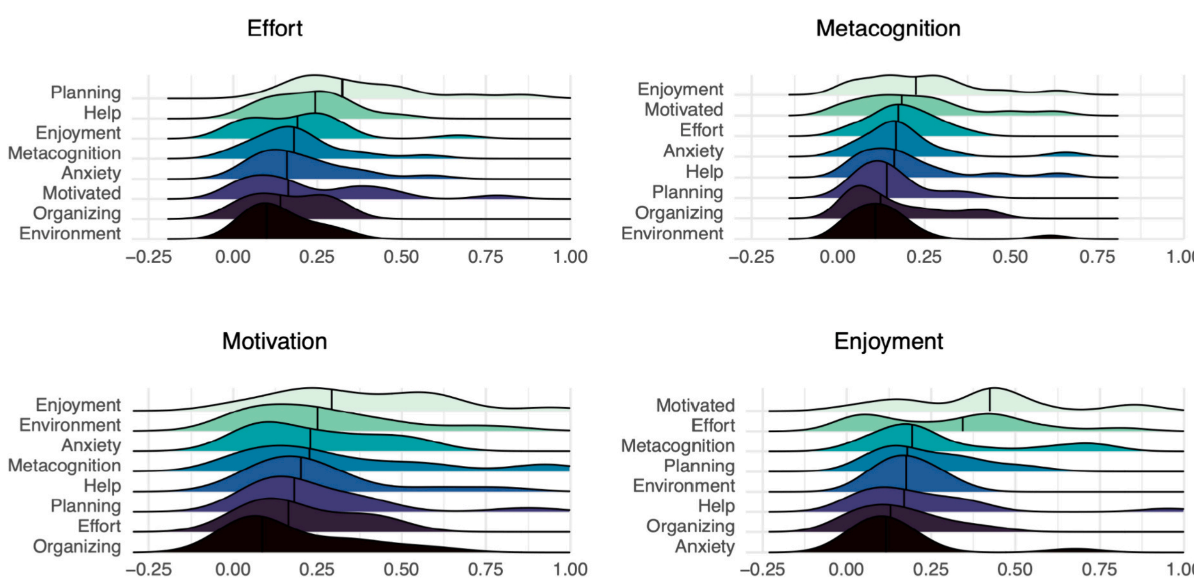

Fig. 6. The figure shows the aggregate absolute value of importance scores (x-axis) of each feature across all students (results from ENR individual models selected by RMSE). The direction of the effect is not depicted in the figure. Time features were included in the machine learning models but were NOT depicted in this visualization.

Again, not a single student had the same order of features as in Fig. 6.

4.4. Results from the general machine learning model (RQ3)

To compare the performance of the idiographic models with the general machine learning models, we fit a general model to all available data of all students. Similar to individual models, ENR outperformed the other two algorithms according to RMSE. All three algorithms performed similarly well when comparing RSQ values. The general machine learning model performed better than any individual models with better RSQ which ranged from 0.5 to 0.73, see the detailed performance of each outcome and algorithm in Fig. 7. Fig. 8 shows that *planning* was the most important predictor of *effort*, *anxiety* was the most important predictor of *metacognition*, effort was the most important predictor of

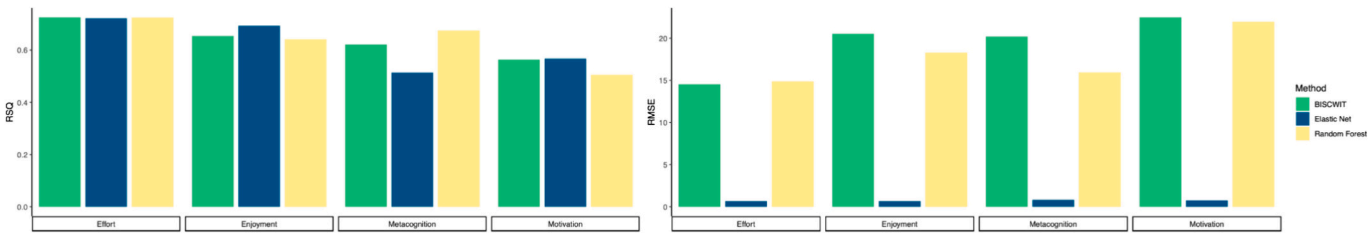

Fig. 7. Model fit indices of the general models.

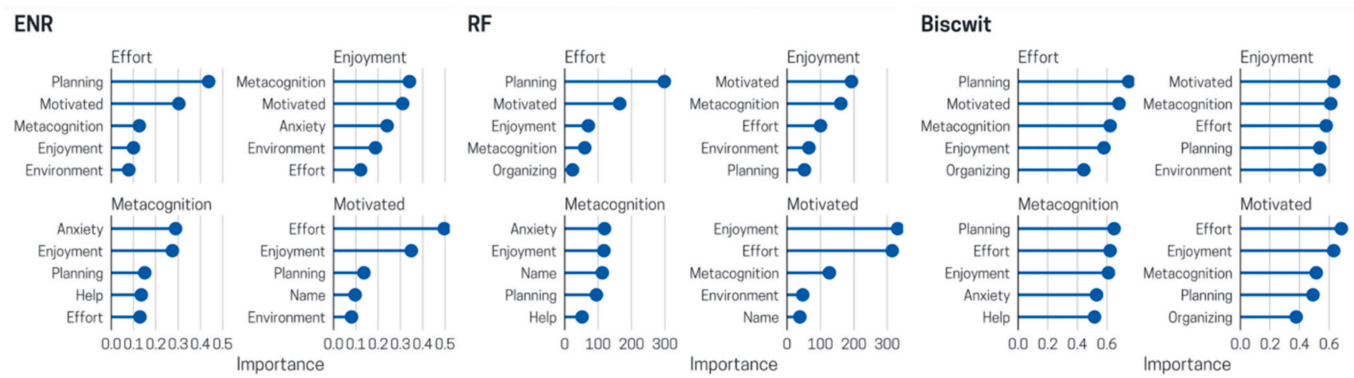

Fig. 8. Top five predictors of the general models predicting each focal outcome.

motivation, and *metacognition* was the most important predictor of *enjoyment*. Interestingly, the top five predictors differed from the individual models, as well as from the aggregated model in Fig. 6. Most importantly, not a single student shared the exact five predictors as the general model.

5. Discussion

For several decades, educational researchers have harnessed the power of prediction to test theoretical models, optimize students' learning, and forecast learning outcomes (Cornog & Stoddard, 1925; Kelley, 1914). Yet, prediction has been elusive, hard to transfer or generalize across contexts —let alone replicate (Conijn et al., 2017; Saqr et al., 2022). Our study took a radically different approach based on the latest insights that recommended modeling the individual where the learning process takes place (Bryan et al., 2021; Saqr et al., 2024). Instead of using others' data to devise an aggregate average that is supposed to work for some —and hopefully many; we used data collected from the same individual, created a unique individualized predictive single-subject or idiographic model for every individual, and evaluated them.

5.1. Idiographic models

First, our study establishes an important finding: idiographic models are tenable, informative, and can deliver individualized predictions and possibly personalized support for most students. Notwithstanding the variability, all the tested outcomes were reasonably predicted in a considerable number of students. These findings speak to the potential of idiographic machine learning models for delivering personalized explainable insights to students based on their own data that may possibly work better than relying on others' data (Beck & Jackson, 2022; Hamaker, 2012; Soyster et al., 2021). This performance is notable given that the daily behavioral variations —predicted by the idiographic models— are fluctuant and harder to predict when compared to the “broad patterns that slowly change” (Inzlicht et al., 2021, p. 331) that are predicted by traditional models (Soyster et al., 2021). The difference is analogous to weather and climate: daily weather patterns are harder to predict compared to the climate seasonal patterns (Inzlicht et al., 2021).

There is certainly a tradeoff between the specificity of insights (the idiographic approach), and the scale and breadth (the group-level analysis) (Howard & Hoffman, 2018). Idiographic models are person-specific and therefore, are not warranted to be generalizable. Of course, the idiographic models are costly. Instead of creating a single machine learning for the entire sample, we created an individual algorithm for each and every individual. This entails frequent data collection, analysis and reporting for each individual.

Through the development of unobtrusive and passive data collection —e.g., from digital sources and sensors—, data can be collected on a local device, analyzed, and presented to the student. In that context, analytics is a bottom-up approach where students collect their own data, have it analyzed on their own devices, and become the sense-makers themselves or elect to consult educators for help (López-Pernas & Saqr, 2021). Guided by the positive effects in other fields, for instance, how activity trackers improved health for users who followed their physical health with fit bands, individualized tracking of learning may facilitate students' learning, autonomy, and self-directedness (Brickwood et al., 2019; Saqr & Lopez-Pernas, 2021). Furthermore, a bottom-up approach could offer more privacy as the data will not leave the student's device (López-Pernas & Saqr, 2021).

5.2. Variability among students

Whereas students were individually predicted, their predictability varied from each other, and from the general model (i.e., the traditional model that has all data combined). Some students were perfectly predictable and some —a minority— were impossible to predict. Even in the case of predicting *effort regulation* —which was the most predictable behavior— still, two students were unpredictable with all algorithms. This could be explained by individual characteristics, survey reporting problems, unmeasured factors or algorithm performance (Greene et al., 2022). Indeed, the fact that some individuals defy predictive algorithms and are impossible to predict is becoming increasingly reported in the literature (Greene et al., 2022). Future research should possibly investigate those unpredictable students and find better ways to offer them support based on their unique characteristics.

Surprisingly, not a single student shared the same order of the top predictors for any outcome with another student. This finding underscores the stark differences between students and the peculiarities of each student's approach. Even more surprising was our finding that not a single student shared the same order of top predictors with the general model or the aggregate model (created by averaging the top individual predictors across all students in Fig. 6). These findings are telling loud and clear that students are essentially different, and an "average" is surprisingly rare and often does not match any student, let alone the majority of students as always claimed. Therefore, if personalization or understanding the individual processes is our goal, then, it stands to reason that we use person-specific methods.

This is not to discredit traditional statistics or dismiss findings from methods based on group-based research. Traditional methods remain a fair approximation and representative of any phenomena that would apply to many students at least partially. Recall that planning was the most important variable in predicting effort regulation in the general model (traditional statistics), it was also the most important predictor in around a third of the students and appeared among the top 5 in all students except one. In that way, traditional methods can answer the question of what works, but will fall short when it comes to which specific individual the results would work for or apply to. Shall we seek to personalize or understand individual students, idiographic methods must be the answer (Howard & Hoffman, 2018).

The variability demonstrated in our study for each person, each outcome and each algorithm contradicts the homogeneity premise (Beltz et al., 2016). Some students were driven by motivation, some by planning, some by anxiety, and others were not predictable in any way and did not conform to the postulated theoretical models (Heikkinen et al., 2022). Such inconsistency could be a major factor behind the inconsistent generalizability of SRL research findings (Bryan et al., 2021; Moeller, 2021). In such cases, the components of the blend of students determines the directions of the results: if the sample has a large proportion of students who are motivation driven, the results will confirm the theory about motivation.

To the best of our knowledge, this is the first study to investigate idiographic machine learning in education and therefore, a comparison with another study is not possible. The closest studies to our work come from recent studies in psychology. Beck and Jackson (2022) studied the utility of idiographic machine learning for the prediction of procrastination, loneliness, and studying among university students; the authors were able to accurately predict students' behavior with "a striking degree of prediction accuracy across participants" (p. 1767). Similar to our study, the authors reported great variations in "every aspect of the models —in accuracy, in feature sets, and in the importance of specific features" (Beck & Jackson, 2022, p. 1779). Another study by Soyster et al. (2021) investigated drinking behavior among college students using idiographic machine learning models. Similarly, the authors reported high predictive accuracy as well as great variability in predictors —that resulted in "a unique number and combination" (p. 303)— and accuracies for each student. It is worth noting that both of the aforementioned studies predicted a binary outcome, which is far easier than a continuous outcome.

5.3. Implications for theory and practice

Oftentimes, SRL theoretical models hypothesize learning processes occurring within the individual or the self (e.g., self-reflection, self-control, self-monitoring etc.). Yet, modeling of such processes is performed by aggregating across a group of individuals (Curran & Bauer, 2011; Inzlicht et al., 2021; Saqr & López-Pernas, 2024). In order to capture the "self" processes (i.e., the intra-individual mechanisms) and reconcile the "mismatch" between the implausible homogeneity assumptions —that individuals are the same— and the actual realities of heterogeneity, within-person methods are therefore necessary (Richters, 2021; Saqr et al., 2024). As Winne (2017) argues, a model based on group-based data "poorly forecasts what any individual learner can expect" (p. 6).

In this study, we demonstrated a method for modeling the person-specific SRL processes that can provide a reliable answer to modeling the hypothesized individualized mechanisms. Idiographic models offered the precise determinants of the learning process for each person without being tainted with others. In doing so, we provided a viable alternative to the well-tested models that rely on aggregate between-person which has so far fallen short. These results can be used to offer personalized support, help students understand and improve their learning. More importantly, the results pave the way for a bottom up learning analytics system whereby a student can implement on their own (López-Pernas & Saqr, 2021).

6. Conclusions

Our study has shown that idiographic single-subject machine learning models are tenable, informative, and can accurately capture the individualized students' mechanisms. The person-specific methods demonstrated in our study can offer precise individualized insights about learners, support, and possible precise intervention. Since person-specific methods require only a single student, they can be applied to courses of any size. In doing so, our study can be considered a step-forward in delivering precision education in line with the growing calls in other fields, e.g., precision medicine.

Predictions varied vastly across students regarding accuracy and predictors and the general average model did not match any student regarding the predictors' order. What is more, we report here that we found a minority of students who were highly predictable across all algorithms and others who were unpredictable regarding any outcome and with all the algorithms we tried.

6.1. Limitations

Our study has limitations that can be attributed to the data collection method, analysis, and context. First, self-reports can be subjective: some students may not respond accurately by under- or overestimating their responses. Self-reports are also prone to recall bias. Still, EMA is known as being less prone to recall bias because it is applied in real time to capture experiences as they happen (and therefore is ecologically valid). Furthermore, we used the survey two times a day. It is unclear that this is the best strategy and frequency of data collection given that there is no research about the optimal frequency. It is possible that some of our models may improve with frequent data collection. Nevertheless, there is always a balance between frequency and burdening students. The models we have chosen were selected based on previous research and suitability for the purpose of analysis. Yet, it is possible that some other techniques may result in improved accuracy in one or more of the predictions. The sample, while not small on the idiographic level (i.e., our data captures reasonable information about each student), is not a large sample that can represent the population. In other words, our results are limited to each and every student in our study (and that it is indeed our goal to create single-subject models) and does not generalize —or at least should not be deemed— applicable to any other student.

Most importantly, our study can be viewed as a proof of concept: a test of the possibilities, challenges and limitations. Our main goal in this article was to demonstrate the process. The findings, being person-specific, may not apply widely, but the process of the explainable models applies to any scenario where enough data exists for understanding and optimizing the person-specific behavior.

We focused on effort as the main predicted outcome as a determining factor for learning and learning outcomes. Yet, an arguably better way would have been to use some kind of assessment of learning on daily basis. Yet, this is far from practical and largely exhausting to test students several times a day. In this study we tested several outcomes as demonstration of the feasibility as an ancillary to our experiment.

Funding and acknowledgement

This study is co-funded by the Research Council of Finland (Academy of Finland) for the project "Towards Precision Education: Idiographic Learning Analytics (TOPEILA)" Decision Number 350560, awarded to the first author. The authors extend their gratitude to the school director, coordinator, and students for their invaluable contributions and support during this study.

CRediT authorship contribution statement

Mohammed Saqr: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Rongxin Cheng:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Sonsoles López-Pernas:** Writing – review & editing, Writing – original draft, Validation, Project administration, Methodology, Formal analysis, Data curation, Conceptualization. **Emorie D Beck:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Formal analysis, Conceptualization.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.lindif.2024.102499.

References

Allen, M. S., Iliescu, D., & Greiff, S. (2022). Single item measures in psychological science. *European Journal of Psychological Assessment: Official Organ of the European Association of Psychological Assessment, 38*(1), 1–5. https://doi.org/10.1027/1015-5759/a000699

Bakker, A. B., Sanz Vergel, A. I., & Kuntze, J. (2015). Student engagement and performance: A weekly diary study on the role of openness. *Motivation and Emotion, 39*(1), 49–62. https://doi.org/10.1007/s11031-014-9422-5

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., … Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *An International Journal on Information Fusion, 58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Barrigón, M. L., Berrouiguet, S., Carballo, J. J., Bonal-Giménez, C., Fernández-Navarro, P., Pfang, B., … MEmind study group. (2017). User profiles of an electronic mental health tool for ecological momentary assessment: MEmind. *International Journal of Methods in Psychiatric Research, 26*(1). https://doi.org/10.1002/mpr.1554

Beck, E. D., & Jackson, J. J. (2021). Within-person variability. In *The handbook of personality dynamics and processes* (pp. 75–100). Elsevier. https://doi.org/10.1016/b978-0-12-813995-0.00004-2.

Beck, E. D., & Jackson, J. J. (2022). Personalized prediction of behaviors and experiences: An idiographic person–situation test. *Psychological Science, 33*(10), 1767–1782. https://doi.org/10.1177/09567976221093307

Beltz, A. M., Wright, A. G. C., Sprague, B. N., & Molenaar, P. C. M. (2016). Bridging the nomothetic and idiographic approaches to the analysis of clinical data. *Assessment, 23*(4), 447–458. https://doi.org/10.1177/1073191116648209

Biwer, F., Wiradhany, W., Oude Egbrink, M. G. A., & de Bruin, A. B. H. (2023). Understanding effort regulation: Comparing "Pomodoro" breaks and self-regulated breaks. *The British Journal of Educational Psychology, 93*(Suppl. 2), 353–367. https://doi.org/10.1111/bjep.12593

Bobrowicz, K., López-Pernas, S., Teuber, Z., Saqr, M., & Greiff, S. (2024). Prospects in the field of learning and individual differences: Examining the past to forecast the future using bibliometrics. *Learning and Individual Differences, 109*, Article 102399. https://doi.org/10.1016/j.lindif.2023.102399

Brickwood, K.-J., Watson, G., O'Brien, J., & Williams, A. D. (2019). Consumer-based wearable activity trackers increase physical activity participation: Systematic review and meta-analysis. *JMIR mHealth and uHealth, 7*(4), Article e11819. https://doi.org/10.2196/11819

Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *The Internet and Higher Education, 27*, 1–13. https://doi.org/10.1016/j.iheduc.2015.04.007

Broadbent, J., Panadero, E., Lodge, J. M., & Fuller-Tyszkiewicz, M. (2023). The self-regulation for learning online (SRL-O) questionnaire. *Metacognition and Learning, 18*(1), 135–163. https://doi.org/10.1007/s11409-022-09319-6

Bryan, C. J., Tipton, E., & Yeager, D. S. (2021). Behavioural science is unlikely to change the world without a heterogeneity revolution. *Nature Human Behaviour, 5*(8), 980–989. https://doi.org/10.1038/s41562-021-01143-3

Cheung, F., & Lucas, R. E. (2014). Assessing the validity of single-item life satisfaction measures: Results from three large samples. *Quality of Life Research: an International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation, 23*(10), 2809–2818. https://doi.org/10.1007/s11136-014-0726-4

Cole, J. S., Bergin, D. A., & Whittaker, T. A. (2008). Predicting student achievement for low stakes tests with effort and task value. *Contemporary Educational Psychology, 33*(4), 609–624. https://doi.org/10.1016/j.cedpsych.2007.10.002

Collie, R. J., Malmberg, L.-E., Martin, A. J., Sammons, P., & Morin, A. J. S. (2020). A multilevel person-centered examination of teachers' workplace demands and resources: Links with work-related well-being. *Frontiers in Psychology, 11*, 626. https://doi.org/10.3389/fpsyg.2020.00626

Conijn, R., Snijders, C., Kleingeld, A., & Matzat, U. (2017). Predicting student performance from LMS data: A comparison of 17 blended courses using moodle LMS. *IEEE Transactions on Learning Technologies, 10*(1), 17–29. https://doi.org/10.1109/TLT.2016.2616312

Cook, C. R., Kilgus, S. P., & Burns, M. K. (2018). Advancing the science and practice of precision education to enhance student outcomes. *Journal of School Psychology, 66*(March 2018), 4–10. https://doi.org/10.1016/j.jsp.2017.11.004

Cornog, J., & Stoddard, G. D. (1925). Predicting performance in chemistry. *Journal of Chemical Education, 2*(8), 701. https://doi.org/10.1021/ed002p701

Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology, 62*, 583–619. https://doi.org/10.1146/annurev.psych.093008.100356

Dietrich, J., Viljaranta, J., Moeller, J., & Kracke, B. (2017). Situational expectancies and task values: Associations with students' effort. *Learning and Instruction, 47*, 53–64. https://doi.org/10.1016/j.learninstruc.2016.10.009

Dolan, E. D., Mohr, D., Lempa, M., Joos, S., Fihn, S. D., Nelson, K. M., & Helfrich, C. D. (2015). Using a single item to measure burnout in primary care staff: A psychometric evaluation. *Journal of General Internal Medicine, 30*(5), 582–587. https://doi.org/10.1007/s11606-014-3112-6

Elleman, L. G., McDougald, S. K., Condon, D. M., & Revelle, W. (2020). That takes the BISCUIT. *European Journal of Psychological Assessment: Official Organ of the European Association of Psychological Assessment, 36*(6), 948–958. https://doi.org/10.1027/1015-5759/a000590

Finnegan, C., Morris, L. V., & Lee, K. (2008). Differences by course discipline on student behavior, persistence, and achievement in online courses of undergraduate general education. *Journal of College Student Retention: Research, Theory & Practice, 10*(1), 39–54. https://doi.org/10.2190/CS.10.1.d

Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences of the United States of America, 115*(27), E6106–E6115. https://doi.org/10.1073/pnas.1711978115

Flake, J. K., & Fried, E. I. (2020). Measurement Schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science, 3*(4), 456–465. https://doi.org/10.1177/2515245920952393

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software, 33*(1). https://doi.org/10.18637/jss.v033.i01

Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *Internet and Higher Education, 28*, 68–84. https://doi.org/10.1016/j.iheduc.2015.10.002

Gilmore, G., Welsh, S., & Loton, D. (2019). An Australian case for relational resilience. Building academic pathways in first year, preservice teacher education. *Pedagogy, Culture and Society, 27*(3), 441–461. https://doi.org/10.1080/14681366.2018.1522365

Greene, A. S., Shen, X., Noble, S., Horien, C., Hahn, C. A., Arora, J., … Constable, R. T. (2022). Brain–phenotype models fail for individuals who defy sample stereotypes. *Nature, 609*(7925), 109–118. https://doi.org/10.1038/s41586-022-05118-w

Hagger, M. S., Chatzisarantis, N. L. D., Alberts, H., Anggono, C. O., Batailler, C., Birt, A. R., … Zwienenberg, M. (2016). A multilab preregistered replication of the Ego-depletion effect. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science, 11*(4), 546–573. https://doi.org/10.1177/1745691616652873

Hall, M., Scherner, P. V., Kreidel, Y., & Rubel, J. A. (2021). A systematic review of momentary assessment designs for mood and anxiety symptoms. *Frontiers in Psychology, 12*, Article 642044. https://doi.org/10.3389/fpsyg.2021.642044

Hamaker, E. L. (2012). Why researchers should think "within-person": A paradigmatic rationale. In M. R. Mehl (Ed.)*, 676. Handbook of research methods for studying daily life* (pp. 43–61). The Guilford Press. xxvii https://psycnet.apa.org/fulltext/2012-05165-003.pdf.

Hart, A., Reis, D., Prestele, E., & Jacobson, N. C. (2022). Using smartphone sensor paradata and personalized machine learning models to infer participants' well-being: Ecological momentary assessment. *Journal of Medical Internet Research, 24*(4), Article e34015. https://doi.org/10.2196/34015

Heikkinen, S., Saqr, M., Malmberg, J., & Tedre, M. (2022). Supporting self-regulated learning with learning analytics interventions – A systematic literature review.

Education and Information Technologies. https://doi.org/10.1007/s10639-022-11281-4

Hernández-García, Á., Cuenca-Enrique, C., Del-Río-Carazo, L., & Iglesias-Pradas, S. (2024). Exploring the relationship between LMS interactions and academic performance: A learning cycle approach. *Computers in Human Behavior, 155* (108183), Article 108183. https://doi.org/10.1016/j.chb.2024.108183

Hoeppner, B. B., Kelly, J. F., Urbanoski, K. A., & Slaymaker, V. (2011). Comparative utility of a single-item versus multiple-item measure of self-efficacy in predicting relapse among young adults. *Journal of Substance Abuse Treatment, 41*(3), 305–312. https://doi.org/10.1016/j.jsat.2011.04.005

Honaker, J., King, G., & Blackwell, M. (2011). Amelia II: A program for missing data. *Journal of Statistical Software, 45*, 1–47. https://doi.org/10.18637/jss.v045.i07

Howard, M. C., & Hoffman, M. E. (2018). Variable-centered, person-centered, and person-specific approaches: Where theory meets the method. *Organizational Research Methods, 21*(4), 846–876. https://doi.org/10.1177/1094428117744021

Ifenthaler, D., & Yau, J. Y. K. (2020). Utilising learning analytics to support study success in higher education: A systematic review. *Educational Technology Research and Development: ETR & D, 0123456789*. https://doi.org/10.1007/s11423-020-09788-z

Inzlicht, M., Werner, K. M., Briskin, J. L., & Roberts, B. W. (2021). Integrating models of self-regulation. *Annual Review of Psychology, 72*, 319–345. https://doi.org/10.1146/annurev-psych-061020-105721

Jansen, R. S., van Leeuwen, A., Janssen, J., & Kester, L. (2018). Validation of the revised self-regulated online learning questionnaire. *Lifelong Technology-Enhanced Learning, 116–121*. https://doi.org/10.1007/978-3-319-98572-5_9

Jovanović, J., Saqr, M., Joksimović, S., & Gašević, D. (2021). Students matter the most in learning analytics: The effects of internal and instructional conditions in predicting academic success. *Computers & Education, 172*(April), Article 104251. https://doi.org/10.1016/j.compedu.2021.104251

Kaiser, T., Butter, B., Arzt, S., Pannicke, B., Reichenberger, J., Ginzinger, S., & Blechert, J. (2021). Time-lagged prediction of food craving with qualitative distinct predictor types: An application of BISCWIT. *Frontiers in Digital Health, 3*. https://doi.org/10.3389/fdgth.2021.694233

Kelley, T. L. (1914). *Educational Guidance: An Experimental Study in the Analysis and Prediction of Ability of High School Pupils, by Truman Lee Kelley* (No. 71). Teachers college, Columbia university.

Khosravi, H., Shum, S. B., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., … Gašević, D. (2022). Explainable Artificial Intelligence in education. *Computers and Education: Artificial Intelligence, 3*, Article 100074. https://doi.org/10.1016/j.caeai.2022.100074

Kim, H., Lee, S., Lee, S., Hong, S., Kang, H., & Kim, N. (2019). Depression prediction by using ecological momentary assessment, Actiwatch data, and machine learning: Observational study on older adults living alone. *JMIR mHealth and uHealth, 7*(10), Article e14149. https://doi.org/10.2196/14149

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Jr., Alper, S., … Bahník, Š. (2018). Many labs 2: Investigating variation in replicability across samples and settings. *Advances in Methods and Practices in Psychological Science, 1*(4), 443–490.

Kuhn, M., & Vaughan, D. (2023). Parsnip: A common API to modeling and analysis functions. https://CRAN.R-project.org/package=parsnip.

Lavelle-Hill, R., Smith, G., & Murayama, K. (2023). Machine learning meets traditional statistical methods in psychology: Challenges and future directions. https://files.osf.io/v1/resources/6xt82/providers/osfstorage/653fb1b9282745178fb86616.

López-Pernas, S., & Saqr, M. (2021). Idiographic learning analytics: A within-person ethical perspective. In *Companion proceedings 11th international conference on Learning Analytics & Knowledge (LAK21)* (pp. 310–315).

Maas, C. J. M., & Hox, J. J. (2005). Sufficient Sample Sizes for Multilevel Modeling. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 1*(3), 86–92. https://doi.org/10.1027/1614-2241.1.3.86

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & Education, 54*(2), 588–599. https://doi.org/10.1016/j.compedu.2009.09.008

Malmberg, J., Saqr, M., Järvenoja, H., & Järvelä, S. (2022). How the monitoring events of individual students are associated with phases of regulation. *Journal of Learning Analytics, 9*(1), 77–92. https://doi.org/10.18608/jla.2022.7429

Malmberg, L. E., Lim, W. H. T., Tolvanen, A., & Nurmi, J. E. (2016). Within-students variability in learning experiences, and teachers' perceptions of students' task-focus. *Frontline Learning Research, 4*(5), 62–82. https://doi.org/10.14786/flr.v4i5.227

Malmberg, L. E., Walls, T. A., Martin, A. J., Little, T. D., & Lim, W. H. T. (2013). Primary school students' learning experiences of, and self-beliefs about competence, effort, and difficulty: Random effects models. *Learning and Individual Differences, 28*, 54–65. https://doi.org/10.1016/j.lindif.2013.09.007

Manwaring, K. C., Larsen, R., Graham, C. R., Henrie, C. R., & Halverson, L. R. (2017). Investigating student engagement in blended learning settings using experience sampling and structural equation modeling. *The Internet and Higher Education, 35*, 21–33. https://doi.org/10.1016/j.iheduc.2017.06.002

Martin, A. J., Mansour, M., & Malmberg, L.-E. (2020). What factors influence students' real-time motivation and engagement? An experience sampling study of high school students using mobile technology. *Educational Psychology, 40*(9), 1113–1135. https://doi.org/10.1080/01443410.2018.1545997

Martin, A. J., Papworth, B., Ginns, P., Malmberg, L. E., Collie, R. J., & Calvo, R. A. (2015). Real-time motivation and engagement during a month at school: Every moment of every day for every student matters. *Learning and Individual Differences, 38*, 26–35. https://doi.org/10.1016/j.lindif.2015.01.014

Moeller, J. (2021). Averting the next credibility crisis in psychological science: Within-person methods for personalized diagnostics and intervention. *Journal for Person-Oriented Research, 7*(2), 53–77. https://doi.org/10.17505/JPOR.2021.23795

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdisciplinary Research and Perspectives, 2*(4), 201–218. https://doi.org/10.1207/s15366359mea0204_1

Molenaar, P. C. M., & Campbell, C. G. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychological Science, 18*(2), 112–117. https://doi.org/10.1111/j.1467-8721.2009.01619.x

Murayama, K., Goetz, T., Malmberg, L.-E., Pekrun, R., Tanaka, A., & Martin, A. J. A. J. (2017). Within-person analysis in educational psychology: Importance and illustrations, 71–87. BJEP Monograph Series II, 12 https://motivationsciencelab.com/wp-content/uploads/2022/02/Murayama_Goetz_et_al_2017_BPSMonograph.pdf.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251). https://doi.org/10.1126/science.aac4716. aac4716–aac4716.

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology, 8*, 422. https://doi.org/10.3389/fpsyg.2017.00422

Pichardo, C., Justicia, F., de la Fuente, J., Martínez-Vicente, J. M., & Berbén, A. B. G. (2014). Factor structure of the Self-Regulation Questionnaire (SRQ) at Spanish universities. *The Spanish Journal of Psychology, 17*, E62. https://doi.org/10.1017/sjp.2014.63

Pintrich, P. R. (2000). Chapter 14 - The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451–502). Academic Press. https://doi.org/10.1016/B978-012109890-2/50043-3.

Pintrich, P. R., & De Groot, E. V. (1991). Motivated strategies for learning questionnaire. *Journal of Educational Psychology*. https://doi.org/10.1037/t09161-000

Plesser, H. E. (2018). Reproducibility vs. replicability: A brief history of a confused terminology. *Frontiers in Neuroinformatics, 11*(January), 1–4. https://doi.org/10.3389/fninf.2017.00076

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. SAGE.

Revelle, W. (2023). psych: Procedures for psychological, psychometric, and personality research. *Northwestern University*. https://CRAN.R-project.org/package=psych.

Richters, J. E. (2021). Incredible utility: The lost causes and causal debris of psychological science. *Basic and Applied Social Psychology, 43*(6), 366–405. https://doi.org/10.1080/01973533.2021.1979003

Rovers, S. F. E., Clarebout, G., Savelberg, H. H. C. M., de Bruin, A. B. H., & van Merriënboer, J. J. G. (2019). Granularity matters: Comparing different ways of measuring self-regulated learning. *Metacognition and Learning, 14*(1), 1–19. https://doi.org/10.1007/s11409-019-09188-6

Saqr, M. (2023a). Modelling within-person idiographic variance could help explain and individualize learning. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.13309

Saqr, M. (2023b). Group-level analysis of engagement poorly reflects individual students' processes: Why we need idiographic learning analytics. *Computers in Human Behavior, 107991*. https://doi.org/10.1016/j.chb.2023.107991

Saqr, M., Fors, U., & Tedre, M. (2017). How learning analytics can early predict under-achieving students in a blended medical education course. *Medical Teacher, 39*(7), 757–767. https://doi.org/10.1080/0142159X.2017.1309376

Saqr, M., Jovanovic, J., Viberg, O., & Gašević, D. (2022). Is there order in the mess? A single paper meta-analysis approach to identification of predictors of success in learning analytics. *Studies in Higher Education, 1–22*. https://doi.org/10.1080/03075079.2022.2061450

Saqr, M., & Lopez-Pernas, S. (2021). Idiographic learning analytics: A single student (N=1) approach using psychological networks. In *Companion proceedings 11th international conference on Learning Analytics & Knowledge (LAK21)* (pp. 397–404).

Saqr, M., & López-Pernas, S. (2024). Mapping the self in self-regulation using complex dynamic systems approach. *British Journal of Educational Technology: Journal of the Council for Educational Technology.*. https://doi.org/10.1111/bjet.13452

Saqr, M., Vogelsmeier, L. V. D. E., & López-Pernas, S. (2024). Capturing where the learning process takes place: A person-specific and person-centered primer. *Learning and Individual Differences, 113*, Article 102492. https://doi.org/10.1016/j.lindif.2024.102492

Schneider, M., & Preckel, F. (2017). Variables associated with achievement in higher education: A systematic review of meta-analyses. *Psychological Bulletin, 143*(6), 565–600. https://doi.org/10.1037/bul0000098

Shafiq, D. A., Marjani, M., Habeeb, R. A. A., & Asirvatham, D. (2022). Student retention using educational data mining and predictive analytics: A systematic literature review. *IEEE Access: Practical Innovations, Open Solutions, 10*, 72480–72503. https://doi.org/10.1109/access.2022.3188767

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology, 4*, 1–32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *The American Behavioral Scientist, 57*(10), 1380–1400. https://doi.org/10.1177/0002764213498851

Soyster, P. D., Ashlock, L., & Fisher, A. J. (2021). Pooled and person-specific machine learning models for predicting future alcohol consumption, craving, and wanting to drink: A demonstration of parallel utility. *Psychology of Addictive Behaviors: Journal of the Society of Psychologists in Addictive Behaviors, 36*(3), 296–306. https://doi.org/10.1037/adb0000666

Stewart, E. B. (2008). School structural characteristics, student effort, peer associations, and parental involvement: The influence of school- and individual-level factors on academic achievement. *Education and Urban Society, 40*(2), 179–204. https://doi.org/10.1177/0013124507304167

Terwee, C. B., Prinsen, C. A. C., Chiarotto, A., Westerman, M. J., Patrick, D. L., Alonso, J., … Mokkink, L. B. (2018). COSMIN methodology for evaluating the content validity of patient-reported outcome measures: A Delphi study. *Quality of Life Research: an International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation, 27* (5), 1159–1170. https://doi.org/10.1007/s11136-018-1829-0

Trautwein, U., Lüdtke, O., Schnyder, I., & Niggli, A. (2006). Predicting homework effort: Support for a domain-specific, multilevel homework model. *Journal of Educational Psychology, 98*(2), 438–456. https://doi.org/10.1037/0022-0663.98.2.438

Trautwein, U., Schnyder, I., Niggli, A., Neumann, M., & Lüdtke, O. (2009). Chameleon effects in homework research: The homework–achievement association depends on the measures used and the level of analysis chosen. *Contemporary Educational Psychology, 34*(1), 77–88. https://doi.org/10.1016/j.cedpsych.2008.09.001

Urdan, T., & Kaplan, A. (2020). The origins, evolution, and future directions of achievement goal theory. *Contemporary Educational Psychology, 61*, Article 101862. https://doi.org/10.1016/j.cedpsych.2020.101862

Valsiner, J., Molenaar, P. C. M., Lyra, M. C. D. P., & Chaudhary, N. (2009). Dynamic process methodology in the social and developmental sciences. *Dynamic Process Methodology in the Social and Developmental Sciences, 2018*, 1–668. https://doi.org/10.1007/978-0-387-95922-1. December.

Webster, E. A., & Hadwin, A. F. (2015). Emotions and emotion regulation in undergraduate studying: Examining students' reports from a self-regulated learning perspective. *Educational Psychology, 35*(7), 794–818. https://doi.org/10.1080/01443410.2014.895292

Winne, P. (1996). A metacognitive view of individual differences in self-regulated learning. *Learning and Individual Differences, 8*(4), 327–353. https://doi.org/10.1016/S1041-6080(96)90022-9

Winne, P. (2017). Leveraging big data to help each learner and accelerate learning science. *Teachers College Record, 119*(3), 1–24. https://www.tcrecord.org/books/pdf.asp?ContentID=21769.

Wright, A. G. C., & Zimmermann, J. (2019). Applied ambulatory assessment: Integrating idiographic and nomothetic principles of measurement. *Psychological Assessment, 31* (12), 1467–1480. https://doi.org/10.1037/pas0000685

Xie, K., Heddy, B. C., & Vongkulluksn, V. W. (2019). Examining engagement in context using experience-sampling method with mobile technology. *Contemporary Educational Psychology, 59*, Article 101788. https://doi.org/10.1016/j.cedpsych.2019.101788