The engage taxonomy: SDT-based measurable engagement indicators for MOOCs and their evaluation

Alexandra I. Cristea[1] · Ahmed Alamri[5] · Mohammed Alshehri[6] ·
Filipe Dwan Pereira[2] · Armando M. Toda[3] · Elaine Harada T. de Oliveira[4] ·
Craig Stewart[1]

Abstract

Massive Online Open Course (MOOC) platforms are considered a distinctive way to deliver a modern educational experience, open to a worldwide public. However, student engagement in MOOCs is a less explored area, although it is known that MOOCs suffer from one of the highest dropout rates within learning environments in general, and in e-learning in particular. A special challenge in this area is finding early, measurable indicators of engagement. This paper tackles this issue with a unique blend of data analytics and NLP and machine learning techniques together with a solid foundation in psychological theories. Importantly, we show for the first time how *Self-Determination Theory (SDT) can be mapped onto concrete features* extracted from tracking student behaviour on MOOCs. We map the dimensions of *Autonomy*, *Relatedness* and *Competence*, leading to methods to characterise engaged and disengaged MOOC student behaviours, and exploring what triggers and promotes MOOC students' interest and engagement. The paper further contributes by building the *Engage Taxonomy*, the *first taxonomy of MOOC engagement tracking parameters, mapped over 4 engagement theories: SDT, Drive, ET, Process of Engagement*. Moreover, we define and analyse students' engagement tracking, with a larger than usual body of content (6 MOOC courses from two different universities with 26 runs spanning between 2013 and 2018) and students (initially around *218.235*). Importantly, the paper also serves as the *first large-scale evaluation of the SDT theory itself*, providing a blueprint for large-scale theory evaluation. It also provides for the first-time *metrics for measurable engagement* in MOOCs, including *specific measures for Autonomy, Relatedness and Competence;* it evaluates these based on existing (and expanded) measures of success in MOOCs: *Completion rate, Correct Answer ratio and Reply ratio*. In addition, to further illustrate the use of the proposed SDT metrics, this study

✉ Alexandra I. Cristea
alexandra.i.cristea@durham.ac.uk

Extended author information available on the last page of the article

is the *first to use SDT constructs extracted from the first week, to predict active and non-active students in the following week*.

Keywords Student engagement · Learning analytics · SDT · MOOCs

1 Introduction

There is no doubt that the global digital revolution and growing availability of broadband have paved the way for new forms of education (Phillips 2005). These include, but are not limited to, online learning, digital educational content production and delivery, and mobile learning. The recent advent of Massive Online Open Courses (MOOCs,[1] relatively short online courses) which target large student numbers and international audiences (De Freitas et al. 2015)—has raised the interest of students, educators and researchers alike. However, regardless of the low access barriers of MOOCS, compared to traditional higher education, concerns have arisen regarding the tremendously low completion rate—only 12% by a meta-analysis of 221 courses in this medium (Jordan 2015). Recently, the dynamics of engagement and motivation in MOOC systems has been especially targeted (Ferguson et al. 2015). However, to the best of our knowledge, *there are no works to date, which systematically employ motivational theories, mapping online student behaviour onto them, to analyse the drives and triggers promoting student engagement*. Moreover, in the past, engagement theories have been created often based on theoretical findings from psychology, or small-scale experiments (Moreno-Murcia et al. 2013; Shen et al. 2009; Langdon et al. 2014). We advocate that, in addition, it is vital to provide numerical, tested engagement measures, for direct application in MOOCs and numerical comparisons. We *consider the advent of 'big data' as a chance to evaluate these theories at scale*.

To address these gaps, we use raw multimodal, multi-dimensional data from comprehensive tracking of student behaviour, as well as aggregate features—such as those extracted from natural language processing (NLP)—to cluster students, then analysing the engagement parameters of these clusters, based on solid motivational theories—starting with one of the most well-known, cited and used one, especially in the education domain—the *Self-Determination Theory (SDT)*(Zhou 2016; Duncan et al. 2020; Deci and Ryan 2013). Many previous works focus only on some particular aspects of the student behaviour (Cristea 2018, Shi and Cristea 2018a). Instead, here, we triangulate multimodal tracking data at various granularity levels—*temporal mode* (time-stamp, day, week, course), *action mode* (where we compute frequencies of student actions for a given time interval), *natural language mode* (where we analyse the language exchange content, including its sentiment mode, etc.)—resulting in 17 indicators. To obtain significant, generalisable results, we perform this engagement analysis on a large longitudinal dataset (6 MOOC courses with 26 runs,[2] spanning 2013–2018, delivering to 218,235 students).

---

[1] MOOC refers to a web-based class, designed to support a large number of participants. It can deliver learning content online to anyone wanting to take a course, with no limit on attendance (Dutta 2021).

[2] A run in FutureLearn means a course being in effect, usually for 4–9 weeks; runs of a course usually are repeated on a yearly basis.

Thus, the main research question addressed in this paper is: "*Can engagement theories help in identifying student success on MOOCs?*". The follow-up question addressed is: *"How are engagement theories applicable in MOOCs?"*.

The main contributions of our work are thus:

- the *Engage Taxonomy*: a new and, to the best of our knowledge, first of its kind taxonomy mapping multimodal student behaviour and data, based on a temporal mode, action mode and language mode in MOOCs, over several motivational theories, allowing for measurable engagement;
- Providing *Engagement Measures* for MOOCs (specifically measurable SDT axes, such as measurable *Autonomy*, *Competence* and *Relatedness*);
- A methodology of how to use the Engage Taxonomy, specifically, the SDT mapping, for identifying and semantically labelling student clusters, potentially leading to improved course design and personalised, adaptive intervention methods;
- A large-scale evaluation of the *SDT* theory for online learning and MOOCs, based on success measures;
- An analysis of a large dataset;
- Applying machine learning techniques, by using SDT constructs as an input, to predict active and non-active students in the following week (week-by-week prediction).
- Importantly, this novel work could bring a new way of thinking in terms of merging the prior top-down, theory-driven learning system building, to the current bottom-up, data-intensive approaches (such as MOOCs).

The remainder of the paper is structured as follows. Section 2 presents related works, including the most influential works from the engagement literature, and then literature targeted specifically at engagement and motivation of students in MOOCs. Section 3 describes the *Engage Taxonomy*, i.e. the model that we designed, by mapping the concepts from motivational theories to raw and aggregated tracked student data, together with *measures for engagement*. Section 4 presents the data, methods and tools that were used for conducting the study. Section 5 compares engagement clusters and shows the results of machine learning prediction. Section 6 presents the discussion concerning our results, including limitations. Finally, Sect. 7 presents our conclusions and future work.

2 Related work

As studies on mapping motivational theories of engagement onto MOOC features with the goal of evaluating these top-down theories in a bottom-up fashion do not exist in the literature (to the best of our knowledge) we are exploring here first the notion of 'engagement', and then popular engagement theories (to substantiate also our choice of SDT). Finally, we analyse the current state of the art in engagement and motivation-related studies in MOOCs, as our closest 'competitors', individually, as well as in a summary table (Table 1).

Table 1 Summary of comparison of our Engage Taxonomy with related work and state of the art

Study	Theory-driven	Data-driven	Motivation and engagement analysis	Prediction (learning outcomes and success)	Sentiment analysis
(Sunar et al. 2016)		X	X	X	
(Nam et al. 2017)		X	X	X	
(Chiang 2019)		X	X	X	
(Pardo et al. 2016)	X		X		
(Barak et al. 2016)		X	X		
(de Barba et al. 2016)	X	X	X	X	
(Wang et al. 2015)		X	X	X	
(Brinton et al. 2014)		X	X		
(Moreno-Marcos 2018, Moreno-Marcos et al. 2018, Kloos 2018)		X	X		X
Adamopoulos (2013)		X		X	X
Wen et al. (2014)		X	X	X	X
(Arroyo et al. 2007)		X	X		
(Cocea 2007)			X	X	
(Jackson et al. 2009)		X	X		
(Adnan 2021)		X		X	
(Khodeir 2021)		X			X
(Mubarak 2020)		X		X	
The Engage Taxonomy	**X**	**X**	**X**	**X**	**X**

2.1 Engagement

Engagement is a complex concept, and there are several definitions for it (Alarcon and Edwards 2011); indeed, some authors claim that there is no single definition suitable for all contexts (Witchel 2013). One possible definition of the engagement of students in their learning (see, e.g. Deng et al. 2020),which we use in this paper, is *as the behavioural, cognitive, emotional and social connections that MOOC participants make with the course content, the instructor and/or other learners*. Engagement, regardless of its definition, has been shown to be a significant attribute towards students' learning success (Kuh 2003). Challengingly, engagement is notoriously difficult to be achieved, especially in learning environments (Willms 2003; Shernoff et al. 2017). In MOOCs, sustaining engagement is even more difficult, as reflected by the higher rates of dropout (Jordan 2015; Shi and Cristea 2018b). Thus, to study the concept of engagement in MOOCs in a systematic, theoretically viable way, we have chosen to apply engagement theories and frameworks that are most popular (most used

and highly cited) in computer science studies, which are also applicable to e-learning (O'Brien and Toms, 2008, Kearsley and Shneiderman 1998).

2.2 Theories of engagement

A classic work on *Engagement Theory (ET)* (Kearsley and Shneiderman 1998) stated that students are engaged, when they are intrinsically motivated to learn, and the activities they perform involve active cognitive processes, such as problem-solving. The *ET* framework promotes three main principles: *Relate*, which emphasises social interactions, mainly collaborating with other peers; *Create*, i.e. making learning a purposeful activity, related to the students' own pace; and *Donate*, where the student makes useful contributions while learning, applying the knowledge onto something practical.

Another highly influential conceptual framework for user engagement with technology, *Process of Engagement* (O'Brien and Toms 2008), proposed four phases tied to the engaged state: *Point of engagement*, where the user gets acquainted with the application; the *Period of engagement*, where the user is using the tool; the *Disengagement*, where they present reasons on why the user would stop using the tool; and the *Reengagement*, which is an iterating process that returns to the first phase, Point of engagement. At the Point of engagement, the users start to use an application based on its aesthetics, novelty, extrinsic motivation to accomplish a task, interest, immersive experience with the product and the Autonomy provided to use the application. The users continue using the system during the Period of engagement, where the users' experience with the system is one of the engaging factors, as are realism, customisable interfaces, fun, time perception, connection with other people and feedback. They also conceptualised some attributes that are related to the disengagement factor, such as the inability to interact, or the lack of challenges and frustration within the system.

Another fundamental book of high influence and extensive implementation, by Deci & Ryan, proposed the *Self-Determination Theory* (SDT) (Deci and Ryan 2013). This theory suggested that human motivation is sustained by three main constructs: *Autonomy*, which is related to the user control over their actions; *Competence*, related to the skills obtained and used to perform a certain task; and *Relatedness*, concerned with the users' interactions with others, as they perform the given task. The theory helps investigate why humans engage in certain activities and the purpose of those activities (Hofer and Busch 2011). The use of SDT has become commonplace in the educational domain, as the theory supports the idea that the students' intrinsic motivation is a primary determinant in their engagement (Zhou 2016).

A similar theory, *Drive*, described also in a recent, authoritative book, by Pink (2011), brings to the fore similar constructs to SDT, like *Autonomy* and *Mastery* (the latter related to Competence); however, instead of Relatedness, the author proposes the construct of *Purpose*, which is tied to the personal desire of the user to do something meaningful for themselves or the community. However, SDT has far more and deeper studies and instruments that can be used, e.g., for intrinsic motivation.

One of the latest proposals by Marczewski (2015) merges the above two theories into a new intrinsic motivation theory, where all four concepts are supported (Autonomy,

Mastery, Purpose and Relatedness). However, this theory has yet to gain much support (in terms of citations).

2.3 Engagement and motivation in AIED and ITS studies

Engagement and motivation have been also studied in the areas of Artificial Intelligence in Education (AIED) and Intelligent Tutoring Systems (ITS), where the connection to theories is more or less explicit. One of the earliest studies, by Arroyo et al. (2007), showed that students were more likely to get re-engaged with the system by providing monitoring interventions to students in-between problems. Also, the negative feedback messages may have motivated some students to be more attentive, to avoid receiving such feedback in their future.

Another early study was conducted by Cocea (2007), to detect the students' engagement level in an e-Learning system. Students were labelled (engaged/disengaged) based on their performance. For instance, a student who took longer or less than the necessary time to perform a given task was considered as a disengaged student. The study showed that the average time spent reading was the best indicator for engagement.

Jackson et al. (2009) divided students into two groups, based on their expectations before they engaged with a system. The first group included students who were "sure" that using the eLearning system would help them improve their knowledge, and the second group had the students who were unsure whether a system could assist them. The results of the post-survey indicated that the students in the first group thought that the system was significantly more enjoyable, more motivating. In contrast, the students in the second group did not like the engagement and found it less motivating.

A more recent work of Walker and Ogan (2016) proposed a vision of the future, in which students would be able to build social relationships with their eLearning systems that are context-sensitive, evolve, and are carefully implemented, in order to enhance the learning environment through the theory-driven approach.

In the last few years, the movement has been more towards Educational Data Mining. As such, there have been studies that have proposed using several machine learning techniques at the same time, to build their prediction models. One very recent study Adnan (2021) used seven different machine learning techniques, including RF, SVM, K-NN, ANN, ExtraTrees, AdaBoost and Gradient Boosting to predict four categories (Withdrawn, Fail, Pass and Distinction). Additionally, another very recent study (Khodeir 2021) deployed the state of art language modelling Transformers (BERT) for the natural language processing task, by using the student comments in a MOOC as input, to predict students who require urgent intervention. (Mubarak 2020) targeted struggling learners who need early intervention, to keep the engagement, by designing a prioritising at-risk student temporal model to predict whether the student will dropout next week.

Compared to all of these theoretical and practical works, which are somewhat related, our work maps, for the first time, to the best of our knowledge, MOOC behaviour onto these classic, highly respected engagement and motivation theories from the field of psychology and learning psychology. Thus, concrete behaviour

extracted by exploiting tracking technologies and techniques, some of which have only recently been made available, can be directly connected to concrete engagement constructs, potentially forming measurable indicators of engagement. To establish the latter, we cluster students according to their tracked behaviour, to find promising predictors, and analysing these clusters in terms of engaged and disengaged behaviour.

2.4 Engagement and motivation in MOOCs

Finally, (see Table 1), we analyse the state-of-the-art in engagement and motivation-related studies in MOOCs. In a recent work conducted by Sunar et al. (2016), the authors investigated how social interactions impact on course completion in MOOCs. According to the authors, dropout rates could be reduced by increasing the users' engagement with social interactions in the systems. The authors presented descriptive statistics and a literature review on the prediction of user behaviours in MOOCs. However, the authors did not base their investigation on any existing motivational theories. Their survey showed that many works (8 out of 15 of their reviewed studies) concerned with predicting students' behaviour, focused on course attrition, by analysing clickstream data and student activity within the system. They argued that social interaction also needed to be analysed—thus supporting the social interaction analysis we do in our work. However, none of the surveyed works analysed the engagement of the students within the course, as a *continuous* variable, nor focused on extracting *measurable* engagement indicators.

A recent work by Nam et al. (2017) predicted disengaging behaviour using data-driven methods on a small scale (25 students) Intelligent Tutoring System (ITS). The authors developed a model to predict when the students were not engaged with the activity of vocabulary training. According to the authors, different attributes could be used to predict engagement and context-sensitive information improved the prediction accuracy. They performed a feature analysis, to select the best prediction features, but they only used supervised learning (with manually labelled data). Although they claimed to create temporal engagement patterns, in fact, whilst some of their analysed features were of a temporal nature, this did not translate into a continuous value for the engagement of students. The continuous data provides more informative and more statistical power to distinguish between groups (Creveling et al. 2006; Jiang et al. 2013).In our work, we use distinct values of the feature on a continuous scale, as interaction with the system is continuously monitored. More importantly, they also did not systematically employ engagement theories, but instead labelled the data manually as engaged/disengaged.

Another very recent work of Chen (2019) proposed the prediction of learning outcomes through learning behaviour, including engagement, in short online courses. In this work, the authors adopted time spent and completion rate to model engagement. They also stated that by using social learning network features, they could increase the prediction accuracy over time. Although relevant, the attributes used to model engagement were not based on any engagement or motivational theory.

In the study conducted by Pardo et al. (2016), the authors focused on creating an approach that combined data about self-regulated learning skills and online activities

in a blended learning course. In this study, engagement was measured through a self-reported questionnaire at the end of the course, and not in real-time, using variables from the system.

Another study, conducted by Barak et al. (2016), explored the motivation to learn in MOOCs. They analysed existing theories and created a model of motivational components, to inspect the influence of language and social engagement in a MOOC. They found a positive relationship between the social interactions e.g. the number of messages posted in online forums and the number of members in online communities. The authors developed a model to measure students' motivation based on pre- and post-course questionnaires; only 325 out of 13,405 students participated in the questionnaire (which is not that surprising, considering the response rates in MOOCs are generally lower for questionnaires Mihalec-Adkins et al. 2016). The authors defined success as students completing essay questions, ten weekly quizzes and a final project. However, they did not focus on predicting success or tying the motivation to success measures. According to their findings, the larger the number of posts, the higher the students' motivation. Finally, the authors also provided a classification of MOOC users, based on their interactions. Another study by Lan and Hew (2020) used self-reported questionnaires and interviews to measure learner engagement and motivation for both completers and non-completers. They also received a low response rate from participants (82 out of 693 students agreed to be interviewed). The study showed that completers are more willing to participate in the social network and receive a certificate.

Moreover, research by Stone (2021) used self-reported questionnaires with only 68 learners to examine "why students are motivated to enrol". The scope of this research focuses only on the motivation that led students to enrol in MOOC courses based on the questionnaire's outcome. The author did not analyse students' interaction data to explore the link between student activities and their motivation. In the same line of research, de Barba et al. (2016) focused on using student motivation as a predictor for learning performance. The authors created a theory-rooted predictive model, to estimate students' intrinsic motivation and participation, and found that the number of attempts in answering quizzes was a good predictor of their final grades. However, this study included only students who were active during the last three weeks of the course and responded to the questionnaire on the last week. Meanwhile, the majority of participants in MOOCs will drop out much earlier. As a result, analysing such learners at an early stage is critical, in order to give early assistance and maintain engagement.

Wang et al. (2015) analysed comments within a MOOC and aimed at creating a predictive model that relates the learning gains with social interactions. The authors defined a taxonomy for the comments and concluded that students that present active (who actively practiced the learning material via quizzes) and constructive discussions (who produced content, e.g. explanations and examples, based on the course material) had significant learning gains. According to the authors, constructive behaviours produced more learning gains than active behaviour. However, they only explored the engagement related to social interactions, to the detriment of other aspects, including motivational theories.

The study conducted by Brinton et al. (2014) focused on providing statistical evidence and a Generative Model (a model used to learn the distribution of data using

unsupervised learning) based on user interaction within a MOOC. The authors performed a large-scale study (73 courses) and found that teachers' participation within forums was positively associated with the level of interaction (e.g. increased discussion threads), but did not affect the decline rate of participation in the course. By conducting this analysis, the authors focused on improving social learning in MOOCs. Although promising results were presented, the authors did not evaluate or predict student performance.

Concerning sentiment analysis, this new field has gained traction in recent MOOC studies (Moreno-Marcos 2018). Adamopoulos (2013) used sentiment analysis to predict student completion of a MOOC. According to the authors, positive sentiments towards the course instructor, assignments and course material were positively related to course retention and completion. However, the authors did not relate the sentiment analysis with student motivation, but only to positive attitudes (comments) in the course. Additionally, their course retention variable was self-reported and not collected from the system log.

Another similar work, conducted by Wen et al. (2014), focused on a sentiment analysis in a MOOC, aiming to understand the relation between students' comments and course success. The authors developed a model to identify motivated students, based on their comments in a course forum. According to the authors, students who had a positive behaviour towards the course (according to their motivation model) had lower rates of dropping out from the course. Our work is pioneering work, in that there are no works that create an engagement taxonomy, with measurable engagement parameters, and allowing for an engagement theory (here, *SDT*) to be evaluated on a large scale. A comparison of the state-of-the-art in engagement and motivation-related studies in MOOCs and our current work depicted in this paper can be observed in Table 1. As can be seen, our proposed method, the Engage Taxonomy (see definition in Sect. 3), brings together all these strands: theory-driven and data-driven approaches, as well as the important motivation and engagement analysis, prediction in terms of learning outcomes and success, as well as emotional engagement (via sentiment analysis). This is done in a focused attempt to express motivational engagement theories via data-driven approaches, and then evaluate them, as further explained. Next, we introduce our Engage Taxonomy.

3 Engage taxonomy: mapping of MOOC indicators onto engagement theories

3.1 Extracting raw and computing aggregated MOOC indicators

As we have observed, most of the motivational theories discussed in Sect. 2 have relatively similar concerns about what triggers student engagement and motivation. The challenge is, then, to map their respective engagement concepts onto MOOC behaviour, available as tracked data, to extract concrete measures that address them. Thus, our next step is to have a collection of potentially relevant data that can be tracked from MOOCs in general. The point being that the more available the data is across different MOOC types and platforms, the more likely it is that this process

is generalisable. Most MOOCs track access to the material, answers to any quizzes, and store chats of their students. Based on indicators used by prior MOOC research for clustering (Oyelade et al. 2010) (Rana and Garg 2016), prediction (Alamri 2019, Alshehri 2021, Chen 2019), data analytics (Moreno-Marcos et al. 2020; Shorfuzzaman et al. 2019), we gather the following:

- *Number of accesses steps per week*: how many steps are accessed by a given student per week.
- *Number of correct answers per week*: questions within quizzes answered correctly by a given student.
- *Number of wrong answers per week*: as students can have multiple attempts to a question, before they get it right, this counter controls how many of those wrong attempts were made by a given student.
- *Number of attempts per week*: number of wrong answers per week plus the number of correct answers per week. This is a measure of the total activity of a student in terms of quizzes per week, showing their engagement along this axis.
- *Number of comments per week*: as students can comment on any 'step', they can produce varying numbers of comments each week. This is the most straightforward way to measure their social contribution and engagement.
- *Number of replies posted per week*: to incorporate part of the social interactive engagement element, we track how many replies a student places to others, thus where they go beyond uttering their opinions in public, but instead, consider the opinions of others. Please note this only reflects on the quantity of replies, and not capturing only prosocial comments.
- *Number of replies received per week*: this is a social construct, but also a measure of the popularity, and hence influence of a given student on their peers—receiving comments shows how engaged other students are with this particular one.
- *Number of likes received per week*: this is a clearly positive social construct; it is a measure of the influence, popularity, as well as of the engagement of other students with a given student.
- *Number of positive comments per week*: this is an aggregate measure, derived based on sentiment analysis, to measure the (positive) nature of engagement of a student with peers.
- *Number of negative comments per week*: this is a similar measure as the one above, measuring negative engagement.
- *Number of (positive–negative-neutral) replies posted per week*: these three self-developed aggregated constructs, based on sentiment analysis, are a measure of the nature of the social engagement of the current student with their peers.
- *Number of (positive–negative-neutral) replies received per week*: these three self-developed aggregated constructs are a measure of the impact of a student on others, as well as of the nature of this impact.

3.2 Mapping indicators to engagement theories

Next, the mapping between the engagement and motivation concepts and the potential indicators within MOOCs was done independently by three experts (see details on

the procedure in Sect. 4.4). Table 2 presents the mapping performed by our experts between the engagement and motivation concepts and the potential indicators within MOOCs that can be tracked, resulting in the *Engage Taxonomy*. As can be seen from the table, due to the generic nature of the MOOC indicators selected, our Engage Taxonomy is available to the research community for further exploitation of other motivational theories. In fact, Table 2 shows mapping onto *four* popular motivational theories, *SDT, Drive, Engagement Theory and Process of Engagement*. Thus, this paper can be used to showcase how to tackle this exploitation, and continue using both SDT and the other motivational theories.

Analysing the expert mapping, for example, experts found that MOOC learners' independent, voluntary activities, such as *#Accessed Steps*, could be related to their '*Autonomy' (AUT)*. Any activity showcasing users' skills, such as *#Correct Answers*, could be related to the '*Competence' (COM)* construct in the *SDT* theory, and the '*Mastery' (MAS)* construct in the *Drive theory*, respectively. User's skills in MOOCs could be tracked by storing numbers—such as the number *(#)* of completed courses or *steps*, as well as number of quizzes answered. Similarly, the users' social interactions address the '*Relate' (REL)* concept in the *Engagement Theory*, the '*Relatedness' (REL)* concept in SDT and other social attributes within the *'Period of Engagement' (PER)*; this may be represented in MOOCs by indicators, such as the number of *# Main Comments posted* and interactions within those comments, or replies and likes. Please note that similar constructs, such as AUT in SDT and Drive, have a similar mapping. *'Create' (CRE)* in the Engagement Theory is somewhat related to autonomy, but it is more concerned with the process of creating a students' own path, or, even more interestingly, creating new information via comments. On the other hand, none of the above indicators were considered appropriate for the '*Point of Engagement' (POI)*, which would need to be extracted from first-access data. All constructs regarding comments and replies were considered to be a good mapping for *'Relatedness' (REL)* in both SDT and Engagement Theory. However, *'Competence' (COM)* in SDT was only showcased in Comments by the *#Likes Received* and in Replies by positive replies posted or received—although the *#Replies Posted* also was considered a measure of competence, possibly as the ones feeling ready to answer other learners' questions would show a degree of competence.

However, this proposed mapping (according to our experts), whilst potentially of use to the research community, evaluated in various ways in the rest of the paper and showed to be performant, is not claiming to be optimal, but can be further improved upon. Moreover, the process of obtaining and testing it are, we believe, just as useful as the final product.

Next, we explain why and how we select SDT, for the evaluation of the expert mapping.

3.3 SDT as illustrator of the engage taxonomy

We only work with *Self-Determination Theory (SDT)* (Deci and Ryan 2013), as it is arguably the most well-known engagement theory, and is well-supported by socio-psychological literature (Gerber and Anaki 2021). *SDT* is a macro-theory linking

Table 2 *Engage Taxonomy*: motivational theories mapped onto students' activities (indicators) in MOOCs

Theories:	SDT			Drive			Engagement theory			Process of Engagement			
MOOC features (indicators): Activities per week	AUT	COM	REL	AUT	MAS	PUR	CRE	REL	DON	POI	PER	DIS	REE
Steps													
# Accessed Steps	x			x			x				x		
Quiz													
# Correct Answers		x			x				x		x		
# Wrong Answers		x			x				x		x	x	
# Attempts to answering questions	x			x							x		
Comments													
# Main Comments posted	x		x	x		x	x	x	x		x		x
# Like Received		x	x		x			x					
# Positive comments posted	x		x	x		x	x	x			x		x
# Negative comments posted	x		x	x		x	x	x			x	x	
# Neutral comments posted	x		x	x		x	x	x			x		x
Replies													
# Replies Posted	x	x	x	x		x	x	x			x		x
# Replies Received			x					x					
# Positive Replies posted		x	x			x		x			x		x
# Positive replies received		x	x		x	x		x			x		x
# Neutral replies posted			x			x		x			x		x
# Negative replies received			x			x		x				x	
# Negative replies posted			x			x		x			x	x	
# Neutral replies received			x			x		x					

personality, human motivation, and optimal functioning. It stems from research on two main types of motivation —intrinsic and extrinsic— that are further thought to shape human behaviour (Deci and Ryan 2013). SDT posits that humans become self-determined when their needs for Competence, Relatedness and Autonomy are fulfilled. Self-determined individuals believe they are in control of their lives, take responsibilities for their behaviours, are self-motivated and determine their actions based on internal values and goals. SDT has led to various sub-theories, such as organismic integration theory and causality orientation; and questionnaires (Hagger and Hamilton 2021), such as the Aspiration Index, Basic Psychological Needs Scale, Christian Religious Internationalisation Scale (Courtney 2018), to name but a few. In education (Standage et al. 2005), students are more likely to learn and succeed when they are intrinsically motivated by their need for Competence, than when extrinsically motivated. Studies within SDT provide strong psychological evidence to support a more interactive, multidimensional picture of human nature in various sociocultural contexts (Chirkov 2009). SDT has been used in musical education (Evans 2015), physical education (Vasconcellos et al. 2020), science education (Lavigne et al. 2007), medical education, amongst others. It is worth mentioning that SDT is further connected with the *self-regulated learning* (Littlejohn et al. 2016) method—which is the predominant approach in MOOCs, and thus we consider SDT an excellent choice of a first analysis of motivational theory application and testing for MOOCs.

Whilst SDT is, as mentioned, well-known and frequently applied, nevertheless, SDT, as well as other motivational theories, have not yet been evaluated on large-scale data. This opportunity is given to us in the context of online learning and MOOCs. Finding out to what extent SDT is really applicable is thus a useful endeavour. Hence, SDT represents a good starting point to experiment with mapping of MOOC features onto motivational theories.

In this paper we propose and use early *measurable indicators of engagement (*the *Engage Taxonomy* for SDT*)* from the first week activities (see Sects. 3.2 and 3.3), as mapped by experts (see Sect. 4.4), and apply these onto the concrete data from our MOOCs to all student clusters. These early behavioural clusters are analysed in terms of their semantics derived from *SDT*, on the axes of Autonomy, Competence and Relatedness. The idea being that, if we find semantically relevant clusters, build based on SDT variables, and then further find that they are correlated to the success parameters, this confirms that the motivational theory-rooted methodology can be applied to characterise students' success.

After performing the above SDT mapping, we need to establish how to measure its success. This is further explained in Sect. 3.5. Next, we tackle how to compute the engagement measures.

3.4 Engagement measures: computing SDT aggregate constructs

As mentioned, we proceed in our further analysis with the SDT model and mapping. Thus, we analyse the following SDT-related constructs, proposing hereby also numerical ways of computing them:

Autonomy (per week): we created this aggregate construct, based on (a generalised version of the data in) Table 2, computed as a normalised value, $Aut \in [0,1]$, as follows:

$$Aut(s,w) = \frac{\sum_{c \in C_{Aut}}^{w} \left(we_c * \frac{c_w^s}{max_{ss \in S(w)} c_w^{ss}} \right)}{\sum_{c \in C_{Aut}}^{w}} \tag{1}$$

Here C_{Aut} are all constructs (extracted from tracking data) usable for establishing the Autonomy of students; where c_w^s is the value of construct $c \in C_{Aut}$ for student $s \in S(w)$ in week w, normalised by dividing it by the maximum of all values of c in that week w, for all students $ss \in S(w)$; we_c is the weight of construct c in the computation of the Autonomy, and should be a value between $[0,1]$; this weight allows to have different constructs influence the result in a different way; currently, we used $we_c = 1$, although further experimentation could render more exact results. Finally, to ensure $Aut(s,w) \in [0,1]$, we normalise the result by dividing it by the number of constructs in C_{Aut}. E.g., if in week *w1* there are 6 steps in total, 3 quizzes, and students have posted together 5 comments (out of which, 2 are positive, 2 neutral and 1 negative), and there have been in total 2 replies; a student *s1* has accessed 3 out of the max 6 steps, answered 1 of the 3 quizzes, and, for simplicity, 0 out of 5 comments in week *w1*, as $\sum_{c \in C_{Aut}}^{w} 1 = 7$ as we have 7 features on Autonomy for SDT in Table 2. Then Aut(s1,w1) = (3/6 + 1/3 + 0/5 + 0/2 + 0/1 + 0/2 + 0/1)/7 = 0.119. A student s2 performing the maximum number of activities related to autonomy constructs would have a total of *Aut(s2,w1)* *= (6/6 + 3/3 + 5/5 + 2/2 + 1/1 + 2/2 + 1/1)/7 = 1*, representing the maximum value of the autonomy computable in that week.

Competence (per week): is also our created aggregate construct, on Table 2, defined similarly to the Autonomy construct above (Eq. 1), but summing only over C_{Com}, the Competence-related constructs (formed of tracked data). For example, if a student *s3* has 1 correct answer out of a maximum of 3, 2 wrong answers out of 3, 1 like received out of 2, 2 replies posted out of 6, 1 positive reply posted out of 3 and 2 positive replies received out of a maximum 2, in week *w1*, as $\sum_{c \in C_{Com}}^{w} 1 = 6$, as we have 6 features on competence for SDT in Table 2, *Com(s3,w1) = (1/3 + 2/3 + 1/2 + 2/6 + 1/3 + 2/2)/6 = 0.53*.

$$Com(s,w) = \frac{\sum_{c \in C_{Com}}^{w} \left(we_c * \frac{c_w^s}{max_{ss \in S(w)} c_w^{ss}} \right)}{\sum_{c \in C_{Com}}^{w}} \tag{2}$$

To further illustrate the usefulness of the weights, it is possible that, instead of using the same weight overall, we would consider in a further iteration of this research (not further explored here beyond this section) that, for the Competence construct, quiz results are much more important than comments and replies. Thus, with the rest of the data as above, instead of $we_c = 1$ for all, we could have $we_{\#LikeReceived} = we_{\#RepliesPosted} = we_{\#PositiveRepliesposted} = we_{\#Positiverepliesreceived} = 0.5$; $we_{\#CorrectAnswers} = we_{\#WrongAnswers} = 1$. Thus, if student *s4* would have the maximum number of correct

answers, and no other competence-related accomplishments in week *w1*, whereas student *s5* would have the maximum number of comments liked during the same week, but no other accomplishments, in this case we would have $Com(s4, w1) < Com(s5, w1)$, as:

$$Com(s4, w1) = \frac{\left(\frac{3}{3}\right)*1+\left(\frac{0}{3}\right)*1+\left(\frac{0}{2}\right)*.5+\left(\frac{0}{6}\right)*.5+\left(\frac{0}{3}\right)*.5+\left(\frac{0}{2}\right)\%.5}{6} = 0.167$$

$$< Com(s5, w1) = \left(\left(\frac{0}{3}\right)*1+\left(\frac{0}{3}\right)*1+\left(\frac{2}{2}\right)*.5+\left(\frac{0}{6}\right)*.5+\left(\frac{0}{3}\right)*.5+\left(\frac{0}{2}\right)\%.5\right)/6 = 0.083$$

Relatedness (per week): is our final aggregate construct, again based on Table 2, and defined similarly to the Autonomy and Competence constructs (Eqs. 1 and 2), summing here over C_{Rel}, the Relatedness-related constructs (formed of tracked data):

$$\boldsymbol{Rel(s, w)} = \frac{\sum_{c \in C_{Rel}}^{w} \left(\boldsymbol{we}_c * \frac{c_w^s}{max_{ss \in S(w)} c_w^{ss}}\right)}{\sum_{c \in C_{Rel}}^{w}} \tag{3}$$

From the generic way we have created these formulas, one could see that they are immediately applicable onto the other theoretical engagement and motivation approaches analysed in Table 1, such as *Drive, Engagement Theory*, and *Process of Engagement*. Please note that we do not claim this model to be optimal; it is however a simple one, thus an Occam-razor based approach.

The overall connection between the definitions above, the mapping in Sect. 3.2 (Table 2) and the indicators in Sect. 3.1 are further shown in Fig. 1. For instance, raw
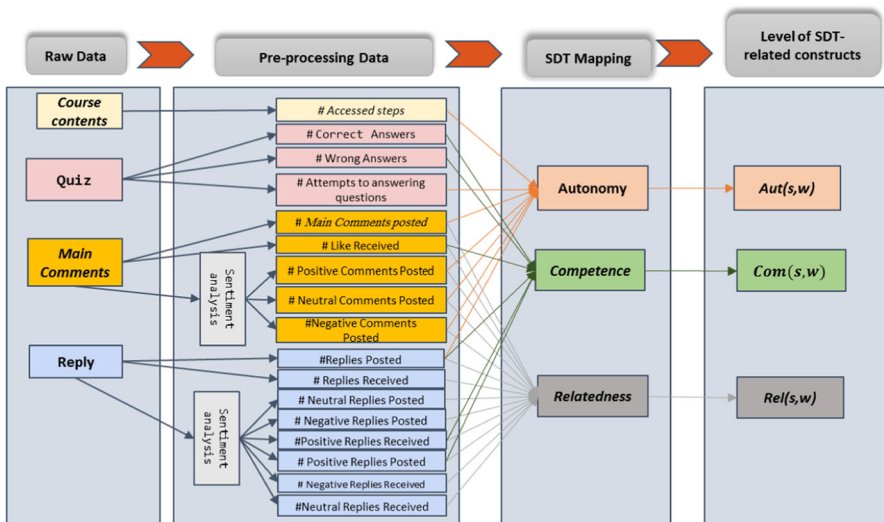

Fig. 1 SDT Theory, mapped to students' activities in MOOCs

data, such as course contents, maps, via pre-processed data *#Accessed_steps* (number of steps accessed by the current student), to the construct on Autonomy in the SDT mapping. Similarly mapped to Autonomy are *#Attempts to answering questions, #Main Comments posted*, etc.). Based on this mapping, the *Aut(s,w)* measurable, SDT-related construct, is defined. Thus, Fig. 1 illustrates how, from raw student tracking data, we can obtain measurable motivational theory constructs. It shows the sequence of operations for extracting these features and turning them into the input data used for clustering and machine learning.

3.5 Computing SDT success

How to measure success in MOOCs is a debatable issue (Davis et al. 2013). Whilst completion is the most frequently used parameter for success (Mohamed and Salleh 2021; Loizzo et al. 2017), it is not unanimously agreed upon as the best way of measuring the perceived, or even the actual success of MOOC students. Students may have different objectives when they embark on a MOOC-journey. Students may wish to learn a new topic in its entirety, but also, they may wish to use the obtained knowledge for different aims. Thus, besides completion, it is useful to look at other measures of success. We do this here in two ways: one, by using a new, rigorous methodological approach based on engagement and motivation theory—and mapping SDT constructs onto track data (raw or derived) of the students, from 12 features, including lesser used features, such as sentiment-related ones (as described in Sects. 3.1, 3.2). The second way is in measuring the success in terms of the other ways a student can interact with a MOOC, beside reading (and completing) pages: by answering quizzes and posting comments. Thus, we measure the student's success additionally via the Correct Answer ratio and the Reply ratio from week 2 to the last week (Fig. 2 and see also Sect. 4.8).

In terms of completion itself, various studies proposed different formulas to estimate the completion (Sunar 2017). We use here the 80% threshold to define Active students in the following week as in (Alamri 2021) (see Sect. 4.9). Please note that this binary way of computing completion has been applied in several prior studies and is accepted in the literature (Alamri 2021, Alamri 2019, Alamri 2020).

We use the measures of student success, as defined in Sect. 4.7, to evaluate the clustering (as explained in Sect. 4.6) and ML prediction based on the proposed SDT theory and the Engage Taxonomy, as illustrated in Sect. 4.9.

Figure 2 shows the input values (SDT elements from Students' activities in the first week) and output results from both the Success measures from clusters analysis and Active students prediction from Machine learning.
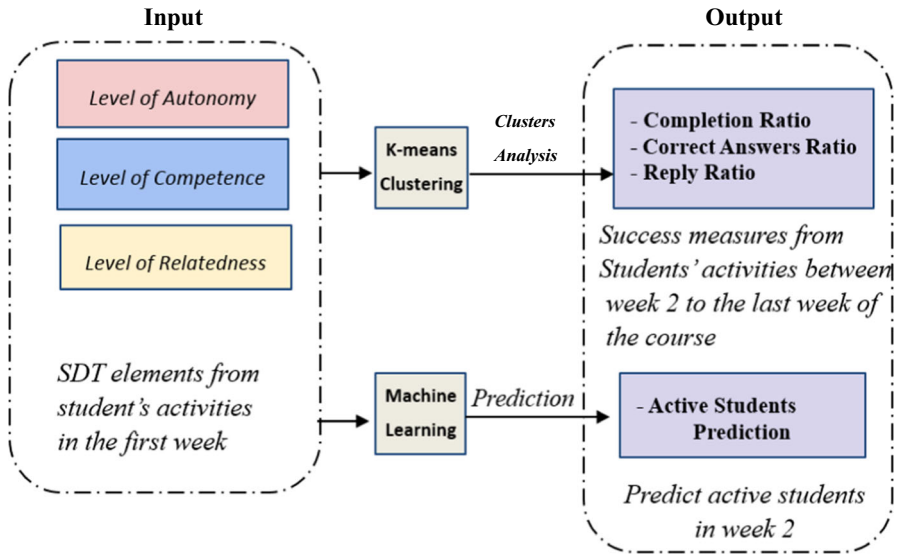
Fig. 2 SDT constructs versus success measures

4 Methodology

4.1 Courses description

FutureLearn[3] is one of the youngest MOOC platforms (since 2012), and the European counterpart to USA's Coursera,[4] EdX,[5] etc., which now supports 327 courses, created by 83 partners, and reached 3 million students in 2018.[6] In general, FutureLearn's philosophy is grounded in social learning theories,[7] where the students' interactions with other students may impact positively on the individual success of each one.

As it is however a newer platform, there are fewer studies performed on it. We fill this gap, by selecting courses delivered through it. In this study, we analyse a large dataset of 26 runs of 6 multidisciplinary courses, which fall under 3 main categories: Literature, Business and Psychology. The courses have been delivered through FutureLearn by two universities in the United Kingdom (< names removed here for anonymity >) between 2013 and 2018. The studied courses have a length of 4 to 10 weeks. The structure of these courses is based on a weekly learning unit. Every learning week includes so-called *'steps'*, which cover images, videos, articles and quizzes. Having joined a given course, students can access these steps, and optionally mark them as completed, or solve quizzes. These steps also allow *comments*, *replies* and *likes* on

---

[3] https://futurelearn.com/

[4] https://www.coursera.org/

[5] www.edx.org/

[6] https://about.futurelearn.com/press-releases/futurelearn-has-3-million-learners

[7] https://www.futurelearn.com/courses/learning-network-age/0/steps/24637

these comments, from different users enrolled within the course. Moreover, quizzes can be frequently attempted, until the Correct Answer is obtained.

4.2 Multimodal and multi-dimensional tracking data

From the very first interaction with the website, every student's activity has been logged, along with a timestamp, into our dataset, with their unique student IDs. Our longitudinal dataset, consisting of the repetition of the 6 courses delivered over consecutive years has enabled a deep exploration of student behaviour. In this in-depth analysis, we use multi-modal data, from many perspectives. Firstly, we use *time-related data*, as well as *numerical data* and *textual data*. Second, we use data of *different granularity*: from a time-related perspective, we analyse data at the level of a timestamp, a day, a week, several weeks, or the whole course.

4.3 Data preparation and pre-processing

We started with 6 courses with 27 runs with 2167 steps in total (an average of around 80.3 steps per run), and 136 quizzes (on average, about 5.2 quizzes per run). The courses are: Course 1, 'Open Innovation in Business (OI)'; Course 2, 'Leading and Managing People-Centred Change (LMPCC)'; Course 3, 'Babies in Mind (BIM)'; Course 4, ' Shakespeare (SHK)'; Course 5, 'Supply Chains (SUP)'; Course 6, 'The Mind is Flat (THM)' as shown in Table 3.

Originally, 218,235 students enrolled in these courses. The first step of the data preparation refined the raw data, by removing all students who had enrolled on one of the courses, but never viewed (accessed) any of the materials (steps). These students clearly never engaged, and thus are irrelevant for our online course behaviour analysis with track data; moreover, these students have been analysed, based on earlier parameters, i.e., registration date, in our previous work [ref-removed]. As a result, we were left with analysing 107,771 students, which is still a large number, to extract behavioural patterns from (tracking data as above). The final dataset included 2,824,891 *accesses* (i.e., 33.44 steps of the course), 2,538,406 *completions* (30.04% steps of the course), 2,125,961 *attempts answers*, 1,425,297 *correct answers*, 700,664

Table 3 FutureLearn courses' summary

Course	Enrolled	Accessed	Run
Open innovation in business (OI)	*6071*	*2792*	*4*
Leading and managing people-centred change (LMPCC)	*10,417*	*6599*	*3*
Babies in mind (BIM)	*48,771*	*26,175*	*6*
Shakespeare (SHK)	*63,625*	*29,432*	*4*
Supply chains (SUP)	*5808*	*2912*	*2*
The mind is flat (THM)	*83,543*	*39,894*	*7*
Total	***218,235***	***107,771***	***26***

wrong answers, 196,143 *comments*, of which 154,432 were replies to previous comments, 294,942 *likes*, 154,088 *positive comments* (representing 78.5% from the total number of comments), 24,755 *negative comments* (12.6% from total of comments), and 17,300 *neutral comments* (8.8% from total of comments), 78,362 *positive replies* (67.88% from total of replies), 17,555 *negative replies* (15.20% from total of replies), 19,515 *neutral replies* (16.90% from total of comments). A normality test indicated that these variables are not normally distributed (Shapiro–Wilk test; $p < 0.05$). Thus, this is further evidence that the students' behaviour is not homogenous, and, if engagement indicators are found, these are expected to have different values for different student sub-populations. As a result, we cluster students based on the tracking parameters (raw or derived), before we proceed to the engagement analysis.

4.4 Expert mapping

To map the features extracted (indicators) from MOOCs onto the set of chosen popular Engagement Theories and, importantly for this study, the SDT theory, we used expert labelling. We put forward some requirements to ensure the annotation quality. Therefore, we selected annotators who held at least a PhD degree and were experts in the domain of learning analytics (LA). The mapping between the engagement and motivation concepts and the potential indicators within MOOCs was done independently by the experts (two professors and one PhD holder). In terms of our experts being LA experts, this was considered necessary, as LA experts understand the need for labelling data for any kind of neural network-based automatic machine-learning. Additionally, they also had the educational expertise to understand what motivates students (as per target of this study). In terms of their knowledge of the purpose of the study, the mapping was done independently of, and at a stage prior to the evaluation study. To further increase the quality of the categorical labelling, in the case where two experts disagreed on mapping a specific behaviour onto the theory's constructs, the mapping from the third expert was considered to determine the decision. Moreover, the interrater Fleiss' Kappa agreement test has been used to assess the inter-rater agreement between experts' mapping. The test resulted in $k = 0.72$, which is interpreted as a substantial agreement (Fleiss et al. 1981). The Engage Taxonomy constructed thus is described in detail in Sect. 3. Please note that, beside the efforts taken as described here to ensure the quality of the process, further validation of the experts' mapping is indirectly provided by measuring the success of students based on their SDT values (as further explained in Sects. 4.7, 4.8 and 4.9, and evaluated in Sects. 5.2 and 0). Construct validity is estimated at *face validity* (Nevo 1985), as in Sect. 3.2. Additionally, the stronger, established and effective 'gold standard' measurement of *criterion validity* (Amirkhan 1994) is provided by calculating the correlation between the results of the mapping and the results of the criterion measurement (here, the success measurement, as introduced in Sect. 4.7 and with results in Sects. 5.2). The usage of the resulting 3 SDT constructs in practice for a prediction of active students in the following week provides further 'in-practice' proof of the usefulness of the SDT mapping (see Fig. 1 and Sects. 4.9 and 5.3). However, whilst we took great care with all steps in our and innovative process of evaluating motivational theories via data-driven approaches, and

have had promising results (see Sect. 5), we do not claim each step is optimal; indeed, this process illustrated here is provided to the research community to further improve upon and explore, as also discussed in Sect. 6.

4.5 Sentiment analysis

Sentiment analysis has become valuable to a wide range of problems, to extract opinions and make decisions across different disciplines and fields, including sociology, marketing and advertising, psychology, economics, political science and others (Bakharia 2016). There are relatively few studies that employ sentiment analysis in MOOCs. Here, we use the outcomes of sentiment analysis, to generate some of the potential indicators for the behavioural clustering of students, such as numbers of positive/negative/neutral comments or replies (see Sect. 3.1). To achieve this, a Natural Language Processing (NLP) tool called TextBlob[8] has been employed, in order to classify students' comments into three categories: positive, neutral and negative. TextBlob is an NLP-oriented Python library trained on a movie reviews corpus. TextBlob offers a simple API to measures polarity and subjectivity of a textual dataset for certain tasks, such as sentiment analysis, classification, part-of-speech tagging, extraction and more complex text processing tasks (Brinton et al. 2014). The tool has been widely used on similar datasets extracted from several social media platforms and proved to be an effective tool for sentiment analysis (Vyas and Han 2019; Dutta 2021; Lohar et al. 2021; Gryllos et al. 2017). This would help understand student expectation and overall satisfaction with the course contents and outcomes. We use sentiment analysis to derive some of the aggregate MOOC indicators, as a basis for the Engage Taxonomy (Sect. 3.2). Appendix H shows an example of using TextBlob to classify students' comments in our datasets.

4.6 Clustering students

As we wish to explore commonalities of students, clustering (Alamri 2019, Rana and Garg 2016) seems like the most appropriate technique to employ first. In terms of student clustering, we analyse data at the level of individual students, groups of students, all students in a run of a course, all students for all runs of a course, and even students who perform several runs of a course.

To decide which clustering technique to use, we have applied several techniques, such as Spectral Clustering, Density-based spatial clustering, Gaussian Mixture Models and K-means clustering. We found that the K-means has the best result in term of Silhouette Coefficient.

The K-Means clustering technique is an unsupervised machine learning algorithm and one of the most popular clustering techniques in data analytics (Xu and Wunsch 2005). Previous research used this technique for related tasks—e.g., (Moreno-Marcos 2018) used sentiment analysis and k-means clustering to analyse discussion patterns on FutureLearn. K-means produces a pre-specified number k of clusters. To find the

---

[8] https://textblob.readthedocs.io/en/dev/

optimal k, we used the means silhouette coefficient, i.e., running clustering on a range of values of k (2 ~ 10, in our case), calculated using intra- and inter-cluster distance. Thus, we used it to partition students based on their behaviour.

We use *raw data* and *aggregate data*, i.e. data composed from different raw data sources. We use data for *early prediction* as well as for *general descriptive analysis*. We use data generated with various techniques: e.g., generated by 'simple' tracking of students, by applying motivational theories, by applying sentiment analysis on student information exchange, by applying statistical indexes on the data, etc. Considering the multiple sources and complexity of the data processed, and to limit it somewhat for the current paper, we have decided to perform a first aggregation step based on the weekly learning unit, which is used as a synchronisation point in instructor-led FutureLearn courses.

This approach further allows for early prediction (see Sect. 4.9)—starting by analysing clustering in *week 1*. Additionally, we ensure that tracking data covers all aspects of the motivational theories involved (see Sect. 3)—especially the SDT theory, which is studied here, as being the most widely used one (see Sect. 3.3).

4.7 Student success measure definitions

Although a considerable amount of literature has been published on student success in a MOOC, there is no formal definition of student success. The concept of MOOC success is multidimensional and the researchers in the domain have been using a variety of definitions such as Course completion, Pass/Fail, certificate earners and final exam grade (Gardner and Brooks 2018).

To measure student success in MOOCs, in the clusters we identify as explained in Sect. 4.6, we use an extended set of parameters (besides the 'basic' Completion ratio), as proposed by (Shi et al. 2020) (as explained at a generic level in Sect. 3.5) d.

- *Completion ratio*: this is the most often used success measure in learning in general, and in online learning in particular: did the student complete the course? Here, instead of obtaining a binary value based on various criteria, such as an (often) arbitrary proportion of the course completed (Alshehri et al. 2018), we instead use the actual (normalised) proportion as a target, we normalised the *Completion rate* for each student, by dividing the number of completed steps by total course steps available, which had the effect of scaling all scores between 0 and 1.
- Correct Answer ratio: often, completion is not sufficient for estimating the success of a student. (Shi et al. 2020) have thus proposed to use other measures on a different 'axis', that of quizzes, and to explore how many answers have been correctly answered by a student, from all answers delivered by that student during the same period.
- *Reply ratio*: similarly, the social activity of a student may be considered another type of measure of success, which is here represented by the number of replies a student receives for their comments.

4.8 Analysing student clusters

To analyse the student subpopulations in the clusters obtained, we performed statistical analysis of the input parameters: the *SDT constructs* (Autonomy, Competence and Relatedness), defined via the Engage Taxonomy, see Sect. 3. In addition, the *success measures* (Completion ratio, Correct Answer ratio and Reply ratio, see Sect. 5.2) have been analysed in each cluster. For this purpose, we computed the mean and standard deviation of these parameters. Additionally, the highest two clusters in terms of the mean of the success measures (Completion Ratio, Correct Answers Ratio and Reply Ratio) are further compared pairwise via a Mann–Whitney U test, for each of the success measures. Moreover, the Pearson correlation coefficient test (PCC) (Benesty et al. 2009) has been employed to assess the relation between *SDT constructs* and *success measures.*

4.9 Machine learning prediction

To illustrate the power of the extracted SDT constructs, we evaluate if they can be used directly as early predictors of student activity. For this, we define *Active* students as those that did *access* 80% of the course material (80% = 0.8 of the number of *#steps*; see also Sect. 3.5, (Alamri 2021), and the rest as *Non-Active* students (Eq. 4):

$$NA(s, w) = \begin{cases} 1, \, if \, TAS(s, w) < TS(w) * 0.8 \\ \qquad 0, \, rest \end{cases} \tag{4}$$

$$TAS(s, \boldsymbol{w}) = \sum_{j=0..\boldsymbol{TS(w)}} AS(s, \boldsymbol{w}, \boldsymbol{j})$$

$$AS(s, \boldsymbol{w}, \boldsymbol{j}) = \begin{cases} 1, \, if \, students \, accessed \, step \, j \, in \, week \, w \\ \qquad 0, \, rest \end{cases}$$

where *s*: student, *TAS(s,w)*: total steps accessed by student *s* in week *w*; *TS(w)*: total course steps available in week *w*; where *TAS(s,w) < = TS(w)*, as the maximum number of steps a student could access in week w are all available ones.

As we aim at early prediction, we use machine learning to predict *Non-Active* Students (*NA*) in week 2, by using SDT constructs (defined via the Engage Taxonomy, see Sect. 3) as input, which were extracted from week 1. Thus our prediction problems is:

Given a student *s*, and their SDT constructs from the current week (week $w = 1$) predict if the same student *s* is non-active *(NA(s,w + 1))* in the following week (i.e., week $w + 1 = 2$).

For a comprehensive analysis, we employed several competing ML ensembles methods, as follows: Random Forest (RF) (Breiman 2001), Gradient Boosting Machine (Gradient Boosting) (Friedman 2001), Adaptive Boosting (AdaBoost) (Freund and Schapire 1997), XGBoost (Chen and Guestrin 2016), (ExtraTrees) (Geurts et al. 2006), Logistic Regression (LR) (Rawlings et al. 1998) and (K-nearest Neighbour) (Anchalia and Roy 2014). Ensembles refer to learning algorithms that fit a model via combining
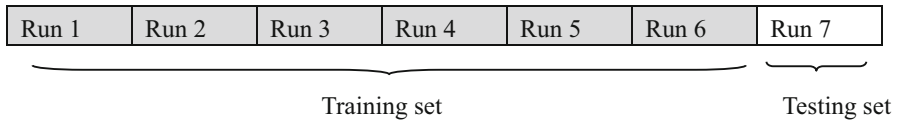
| Run 1 | Run 2 | Run 3 | Run 4 | Run 5 | Run 6 | Run 7 |

Training set   Testing set

Fig. 3 The Mind Is Flat course

several simpler models to strengthen the accuracy (Zhang and Ma 2012). In cases of binary classification (like ours), Gradient Boosting uses a single regression tree to fit the negative gradient of the binomial deviance loss function (Friedman 2001). XGBoost, a library for Gradient Boosting, contains a scalable tree boosting algorithm, which is widely used for structured or tabular data, to solve complex classification tasks (Chen and Guestrin 2016). AdaBoost is another method, performing iterations using a base algorithm. At each interaction, AdaBoost uses higher weights for samples misclassified, so that this algorithm focuses more on difficult cases (Mazini et al. 2019; Kumar et al. 2011). Random Forest is a method that uses several decision trees, constructed via bootstrapping resampling and then applying majority-voting, or averaging, to perform the estimation (Mishina et al. 2015).

The current study used a balanced accuracy score (BA) to evaluate the performance of the models; this metric is widely used to calculate accuracy for imbalanced datasets, by preventing the majority of negative samples from biasing the result (Brodersen et al. 2010). Please note that, although we applied and compared various classifiers, our aim here was not to optimise the prediction of Active students in week 2, but to showcase how the SDT theory, and our mapping of indicators onto SDT constructs, can be used directly as a predictor.

As we have used a massive dataset for different courses, we have prepared the training and testing sets based on the last Run of the course. For example, in the Mind Is Flat course, we trained our ML models by using students' data extracted from early runs (Runs 1 to 6) for students who registered between 2013 and 2016. However, for testing the models, we used a new data set extracted from the last run (Run 7) that contains students' activities in 2017-see Fig. 3 which is similar to some extent to transformer models (Vaswani et al. 2017).

In addition, we combined all datasets (Runs) together for each course to predict Active and Non-Active Students in week 2 by using the tenfold cross-validation, a widely used technique to evaluate a predictive models (An et al. 2007).

5 Results

5.1 Student clusters

The indicators defined in Sect. 3 are aggregated, to obtain 6 datasets corresponding to the 6 courses. Further aggregation would not be applicable, as the structure of the courses varied in length, number of steps, quizzes, resources, etc., available, whereas within each course, these were (relatively) constant, thus progress would have been expected to be equal for each student—all other parameters being equal. As it is

Table 4 Number of students and percentages in each cluster for each course

Course	Cluster 1	Cluster 2	Cluster 3	Total
Babies in mind	2055 (8%)	10,155 (39%)	13,965 (53%)	26,175
Shakespeare	1454 (5%)	12,247 (42%)	15,731 (53%)	29,432
Supply chains	96 (3%)	631 (22%)	2185 (75%)	2912
The mind is flat	1410 (3%)	12,522 (31%)	25,962 (65%)	39,894
Open innovation in business (OI)	164 (6%)	786 (28%)	1842(66%)	2792
LMPCC	314 (5%)	2255 (34%)	3997 (61%)	6566

difficult to compare data from different runs and different sources, we normalised the indicators, by dividing each value by the highest value in the column (activity) within each course, which had the effect of scaling all scores between 0 and 1 (see Sect. 3). Please note that other methods of normalisation could have been used, such as min–max or z-norming (Patro and Sahu 2015; Stocks et al. 2011), all serving the goal to achieve a type of distributional equivalence between runs.

We first clustered the students in the 6 courses for the 26 runs and obtained the main clusters for students, based on the SDT variables from students' activities in the first week. The silhouette coefficient analysis showed that $k = 3$ for K-Means is the most appropriate, when clustering the behavioural indicators. Hence, we obtain 3 clusters (Table 4).

Table 4 shows an overview of students' distribution in each cluster. As can be seen from the table, cluster 3 contains the majority of the students: more than the other two clusters for all six courses. Approximately two-thirds of the students have been clustered in cluster 3 in Supply Chains and 66% of the students have been grouped in cluster 3 in Open Innovation in Business (OI). On the other hand, cluster 1 comprises the minority of the students (3%-8% of the students) and cluster 2 is mid-ranged (28% and 42%).

5.2 Student cluster analysis

To semantically analyse the clusters, Table 5 illustrates the three elements of SDT extracted from **week 1** activities only, versus the success measures of student activities from **week 2 to the last week** (with highest values in bold). This cluster analysis thus allows us to estimate if the SDT values of week 1 would be a good predictor for success in the rest of the course.

The most striking result to emerge from the table is that there is a clear positive correlation between the SDT constructs (Autonomy, Competence and Relatedness) and the success measures (Completion Ratio, Correct Answers Ratio and Reply Ratio). This has further been proven statistically, by using the Pearson correlation coefficient test. The results revealed a positive correlation ($r > 0$), as shown in Appendix A1). The table shows that the Relatedness construct is the most correlated construct, strongly correlated with the Reply-Ratio measure in the six courses, whereas Autonomy and

Table 5 Mean and standard deviation for the 3 SDT construct-based clusters aggregated over the 6 courses, versus the success measures (highlighted in green)

Course			Cluster 1 Mean/SD	Cluster 2 Mean/SD	Cluster 3 Mean/SD
Babies in Mind	**SDT elements** from Students' activities in the first week	Autonomy	**0.37 / 0.08**	0.19 / 0.04	0.04 / 0.04
		Competence	**0.2 / 0.09**	0.09 / 0.03	0.0 / 0.01
		Relatedness	**0.16 / 0.08**	0.02 / 0.03	0.0 / 0.0
	Success measures from Students' activities between week 2 to the last week	Completion Ratio	**0.55 / 0.4**	0.40 / 0.4	0.02 / 0.09
		Correct Answers Ratio	**0.39 / 0.36**	0.27 / 0.34	0.01 / 0.05
		Reply Ratio	**0.02 / 0.06**	0.0 / 0.01	0.0 / 0.0
Shakespeare	**SDT elements** from Students' activities in the first week	Autonomy	**0.39 / 0.07**	0.22 / 0.03	0.04 / 0.04
		Competence	**0.35 / 0.09**	0.20 / 0.03	0.0 / 0.01
		Relatedness	**0.19 / 0.09**	0.01 / 0.03	0.0 / 0.01
	Success measures from Students' activities between week 2 to the last week	Completion Ratio	**0.62 / 0.38**	0.44 / 0.39	0.02 / 0.09
		Correct Answers Ratio	**0.45 / 0.31**	0.31 / 0.31	0.01 / 0.05
		Reply Ratio	**0.02 / 0.07**	0.0 / 0.01	0.0 / 0.0
Supply Chains	**SDT elements** from Students' activities in the first week	Autonomy	**0.48 / 0.09**	0.23 / 0.04	0.04 / 0.04
		Competence	**0.33 / 0.12**	0.19 / 0.05	0.0 / 0.01
		Relatedness	**0.23 / 0.11**	0.02 / 0.04	0.0 / 0.01
	Success measures from Students' activities between week 2 to the last week	Completion Ratio	**0.63 / 0.43**	0.49 / 0.42	0.03 / 0.13
		Correct Answers Ratio	**0.4 / 0.33**	0.32 / 0.31	0.0 / 0.02
		Reply Ratio	**0.04 / 0.11**	0.0 / 0.03	0.0 / 0.02
The Mind is Flat	**SDT elements** from Students' activities in the first week	Autonomy	**0.4 / 0.08**	0.21 / 0.04	0.05 / 0.04
		Competence	**0.34 / 0.1**	0.18 / 0.03	0.0 / 0.01
		Relatedness	**0.21 / 0.1**	0.01 / 0.03	0.0 / 0.01
	Success measures from Students' activities between week 2 to the last week	Completion Ratio	**0.56 / 0.36**	0.45 / 0.37	0.01 / 0.08
		Correct Answers Ratio	**0.41 / 0.3**	0.33 / 0.3	0.0 / 0.04
		Reply Ratio	**0.03 / 0.06**	0.0 / 0.01	0.0 / 0.0
Open Innovation in Business (OI)	**SDT elements** from Students' activities in the first week	Autonomy	**0.37 / 0.08**	0.18 / 0.04	0.04 / 0.03
		Competence	**0.26 / 0.09**	0.07 / 0.04	0.0 / 0.01
		Relatedness	**0.24 / 0.09**	0.02 / 0.04	0.0 / 0.01
	Success measures from Students' activities between week 2 to the last week	Completion Ratio	**0.55 / 0.39**	0.45 / 0.41	0.01 / 0.04
		Correct Answers Ratio	**0.27 / 0.23**	0.19 / 0.22	0.0 / 0.02
		Reply Ratio	**0.02 / 0.11**	0.0 / 0.0	0.0 / 0.0

Table 5 (continued)

LMPCC	**SDT elements** from Students' activities in the first week	Autonomy	**0.35 / 0.07**	0.18 / 0.02	0.04 / 0.04
		Competence	**0.2 / 0.09**	0.1 / 0.03	0.0 / 0.01
		Relatedness	**0.16 / 0.08**	0.01 / 0.02	0.0 / 0.0
	Success measures from Students' activities between week 2 to the last week	Completion Ratio	**0.68 / 0.37**	0.53 / 0.38	0.02 / 0.12
		Correct Answers Ratio	**0.4 / 0.24**	0.33 / 0.28	0.0 / 0.05
		Reply Ratio	**0.02 / 0.1**	0.0 / 0.01	0.0 / 0.0

Competence constructs are less correlated. On the other hand, the Competence construct, in the Supply Chains and LMPCC courses, is the most correlated construct with the Answer Ratio measure, and the Autonomy is the most correlated construct with the Completion Ratio measure.

Please note that this is even more important, as SDT are measured as early variables, potentially usable for prediction (week 1, as said) and success measures are collected week 2 till the last week. Thus, we can arguably claim that *SDT motivational theory constructs can be used as early informer for success towards the end of the course*. Therefore, it is an important result that we could confirm, via a data-intensive approach, that most motivated and engaged students, **as defined by the SDT motivational theory**, turn out to be the most successful. Interestingly, our clustering succeeded to showcase this, by grouping these students in cluster 1. Likewise, cluster 2 has naturally resulted in assembling the intermediate students, who have statistically significantly ($p < 0.05$) lower results in terms of both SDT constructs as well as success (see Appendix F). Cluster 3 gathers, on the other hand, a large number of users who are not very engaged, as per SDT parameters (which is not that surprising, considering clustering was done based on SDT parameters), but, importantly, nor are they very successful (as per our three success measures). A good example of this can be found in the 'Babies in Mind' course, as the mean of cluster 1 was 0.37 in Autonomy, 0.2 in Competence and 0.16 in Relatedness; the Completion rate was 0.55. On the other hand, cluster 2 reported lesser values than cluster 1 in all SDT parameters (0.19 in Autonomy, 0.09 in Competence and 0.02 in Relatedness; with a Completion rate of 0.4 (0.15 less than cluster 1)). We can notice the same pattern for all other courses (see Table 5).

Appendix E further shows the mean, standard deviation and maximum values for the three SDT construct over six courses. For all students, the Autonomy construct has the highest mean score (ranging from 0.133 to 0.096). The autonomy mean score in Table 5 for students in cluster 1 (ranging from 0.48 and 0.35) represents a high degree of autonomy, compared to students in cluster 2 and 3. The Competence construct ranked as the second highest, with a mean score ranging from 0.102 to 0.036. Finally, the Relatedness construct had the lowest mean score (from 0.021 to 0.01).

A follow-up analysis additionally shows that there is a significantly high correlation between Autonomy and Competence (ranging from 86 to 93), and a lesser correlation between Relatedness and both Competence or Autonomy (ranging from 57 to 84) over the six courses (see Appendix A2). Section 6 further discusses these findings.

As previously mentioned, the clusters were created based on the SDT constructs (Autonomy, Competence and Relatedness) from students' activities in the first week (analysing basically if these newly proposed early engagement parameters would be good potential early predictors of success). Appendix F shows the results of the statistical analysis of the success measures for the highest two clusters in terms of SDT (cluster 1 versus cluster 2). We can see that a significant difference exists for all success measures between the highest two clusters for all six courses, meaning their differences are not due to chance.

Figure 4a-f allows for further visual analysis of the results, by showing the 3D-plots of the 6 courses, the relevance of SDT constructs being clearly visualised in the plots for each cluster. Clusters are well separated, with cluster 1 containing the higher SDT value values (in green), cluster 2 the intermediate (in yellow) and cluster 3 the low values (in red). Cluster 1 contains the most motivated and engaged students whereas, cluster 3 identified the very low engagement students. Please note that cluster 3 contains students with very low SDT values and their data points are very close to each other. The visualisation for the students with similar SDT values are overlapping giving a 'feel' that the red cluster is smaller, whereas it is the largest.

To better understand the usefulness of the SDT constructs with respect to the impact on the success measures, the centroids of each cluster are further represented as a point in the radar plot in Fig. 5 (a-f; clusters with the same colouring convention as in Fig. 4). It is clear that the students with higher values of the SDT features in the first week activities have a higher chance to be the most successful. Indeed, the wider spread of the green cluster 1 area shows for all SDT values (Autonomy, Competence and Relatedness) show a respective widespread for the success values (Answer Rate and Completion Rate). Similarly, the yellow cluster 2 is wider for all these when compared with the red cluster 2, for which all these values (SDT and success) are so low, it almost appears as a small dot.

We have repeated the experiment for one course "The mind is flat" using only one dimension of the SDT constructs ("Relatedness") and compare it with the results obtained by using all SDT constructs. The clusters with one construct look worse than using all SDT constructs. The results showed that 93% of students were clustered in cluster 3 (was 65% when we used all SDT constructs), which means that a lot of high achievers' students in cluster 1 and 2 moved to cluster 3. Therefore, the mean completion rate for Cluster 3 increased from 0.01 to 0.15 please see Appendix K.

5.3 Machine learning prediction

Table 6 shows the performance of the predictive models, evaluated by the Balanced Accuracy score (Brodersen et al. 2010), a commonly used metric for binary classification of unbalanced datasets (see Sect. 4.9). Moreover, several measures, such as Precision, Recall and F1-Score have been also used for a complete evaluation of the prediction performance (full results provided in Appendix B, due to the extensive size of the table). In addition, AppendixI shows more results for using tenfold cross-validation by combining all datasets (Runs) together for each course to predict Active and Non-Active Students in week 2.
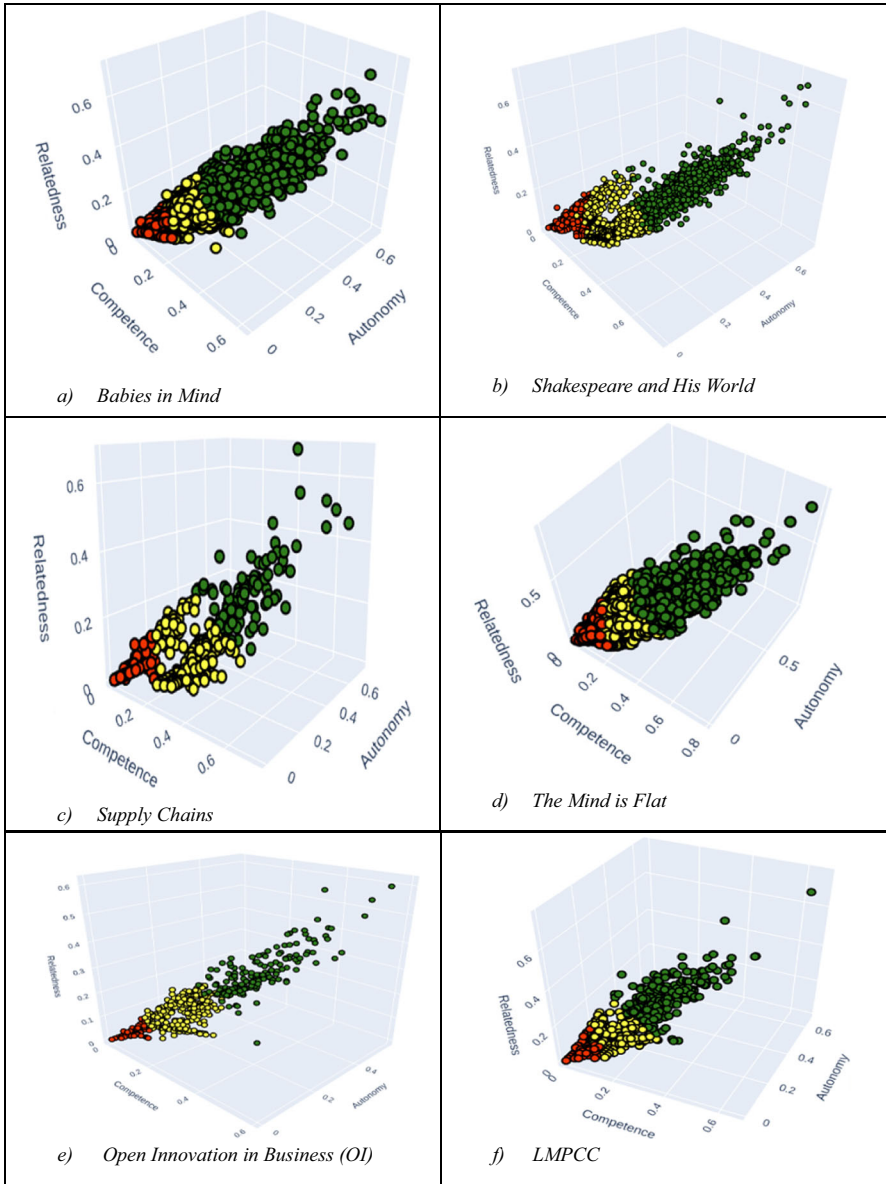
Fig. 4 a–f The three clusters mapped onto the Self-Determination Theory (SDT) (Cluster 1: green; Cluster 2: yellow; Cluster 3: red). (Color figure online)
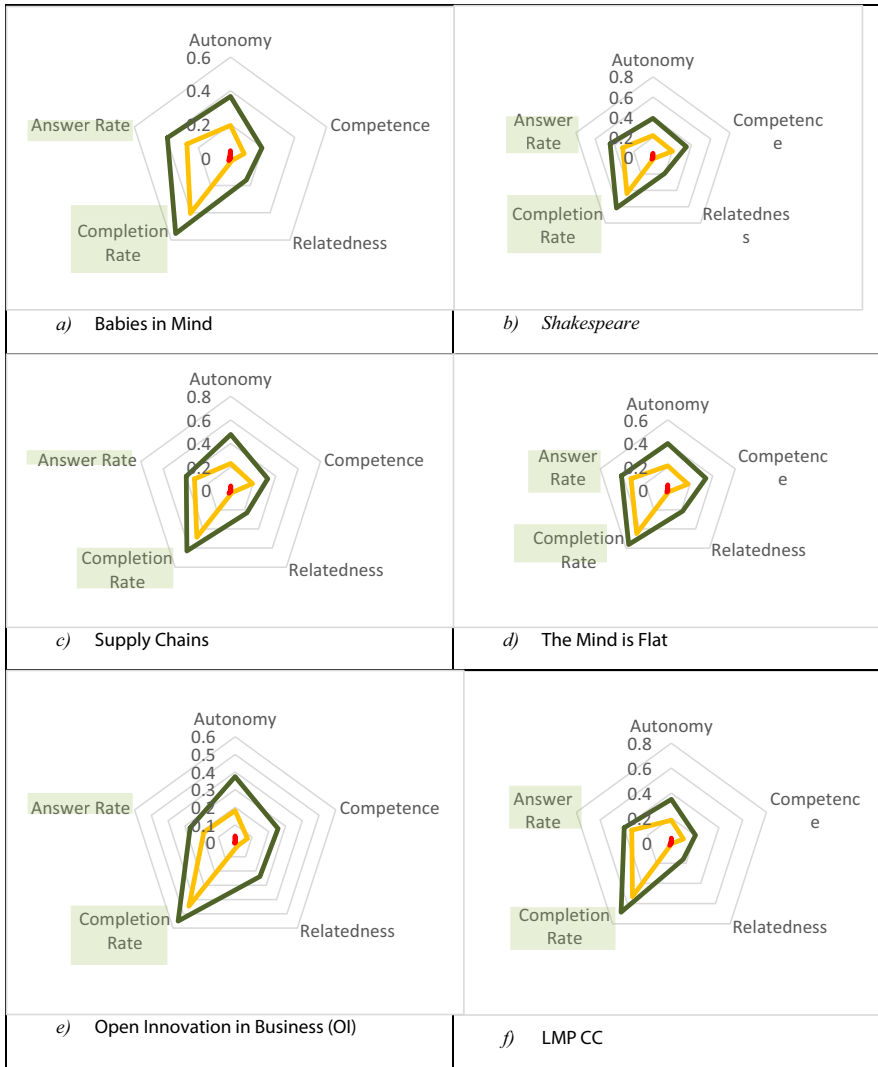
Fig. 5 a–f Values of the SDT features versus success measures (Completion Ratio and Correct Answers Ratio (Cluster 1: green; Cluster 2: yellow; Cluster 3: red). (Color figure online)

In general, all algorithms achieved good results, indicating that, regardless of the employed model, the SDT constructs extracted from the first week in this study proved to be powerful in predicting Active and Non-Active students in the second week. Whilst all models' performances are generally relatively good; the most robust model is the ExtraTrees model, as it outperforms in two courses: 'The Mind is Flat' 91.70%, and 'Leading and Managing People-Centred Change (LMPCC)' 90.13%.

Additionally, we have computed the Gini index (GI) (Dorfman 1979) for the 'winning' ExtraTrees algorithm, to evaluate the feature importance of each feature used to

Table 6 Prediction of Active and Non-Active Students in week 2 based on week 1 SDT constructs

Courses	AdaB	ExtrTr	GBoost	KNN	LR	RF	XGBoost
Babies in mind	82.87%	80.33%	62.48%	75.13%	**83.01%**	82.87%	67.50%
Shakespeare	87.11%	87.29%	**87.33%**	80.57%	87.24%	87.26%	87.24%
Supply chains	91.35%	92.46%	92.73%	83.74%	90.74%	92.11%	**92.73%**
The mind is flat	91.11%	**91.70%**	91.16%	86.54%	91.31%	91.65%	91.17%
Open innovation in business (OI)	**90.09%**	89.75%	84.70%	87.39%	83.35%	88.07%	89.08%
(LMPCC)	89.86%	**90.13%**	89.63%	87.12%	88.18%	89.57%	89.32%

predict non-active students in the following week (see Appendix C). Briefly, results show that Competence is ranked as the most important construct in the classification (importance value ranging from 0.43 and 0.50). Autonomy is ranked as the second important construct (importance value ranging from 0.27 to 0.34). Finally, the Relatedness construct is ranked as the least important factor (importance value ranging from 0.20 to 0.23).

Appendix J shows the prediction results of Active and Non-Active Students in week two based on one of the SDT constructs (the Relatedness); the prediction accuracies were considerably lower (between 50% and 63%) compared to using all SDT constructs (between 67% and 92%), showing that the three constructs are necessary together for prediction.

6 Discussion

In this paper, we propose the *Engage Taxonomy*, a mapping of 'bottom-up' MOOC data to 'top-down' high level concepts from motivational theories. To create it, we have mapped the engagement and motivation concepts and the potential indicators within MOOCs, with the help of three experts, onto these theories. Moreover, the Fleiss' Kappa agreement test has shown a high rate of agreement (substantial agreement) between experts (see Sect. 4.4 and Table 2). However, this process illustrated here is provided to researchers to further improve upon and explore. Indeed, a similar mapping can be extended to incorporate further MOOC variables, if other MOOCs provide additional behavioural meta-data. Additionally, our weighted model as in Sect. 3.4 can be further improved: weight optimisation sought—e.g. searching optimal values between [0,1]; or even proposing negative values. For instance, negative replies are considered here to influence positively the relatedness (in the sense of 'any news is good news', and any interaction and replies is affecting the relatedness). However, another model may consider this relation as a negative one. We also envision optimisation in the sense of more complex, non-linear models. Our methodology can be seen as a shell to be applied to different MOOCs, or onto different motivational theories (as already

started in Table 2, where we have not just SDT, explored further in-depth in this paper, but also the Drive Theory, Engagement Theory and Process of Engagement Theory).

In terms of MOOC completion, although about 10% of participants complete their courses in MOOC, clearly many more students quit over time. Appendix D presents the number of remaining students over time, showing that the curve of dropouts is itself dropping speedily. Hence, participants are most likely to drop out in the first few weeks. Therefore, identifying those students at an early stage is important, to provide early intervention, to keep the engagement going. In this study, we used SDT constructs extracted from the first week as input features for machine learning. This provides an opportunity to deal with at-risk students at an early stage (week 2, which we identified as a critical period). Future work will explore how using week-by-week prediction affects the prediction accuracy and gain (in terms of number of students dropping out at later stages).

The correlation results between the SDT constructs and the success measures (Appendix A1) point to Relatedness being the best construct to measure the Reply Rate of students, as it shows higher correlation values in the six courses. Both Autonomy and Competence constructs have similar correlation patterns with success measures. For the Completion Rate, the Autonomy was the most correlated construct in four courses and Competence was the highest in two courses. Finally, Competence was the most correlated construct to the Answer Rate in four courses.

This finding led us to further explore the direct correlation among the SDT constructs. The three SDT constructs are thought to represent different traits conceptually, so some independence is expected. This is confirmed by Fig. 1, in that at least one feature uniquely maps to each SDT construct. However, the linear trend of the distribution of students in Fig. 4 suggests that the Autonomy, Competence, and Relatedness dimensions from the expert mapping may be correlated. An additional analysis (Appendix A2), shows that a significantly high correlation between Autonomy and Competence, ranging between (0.86 and 0.93). Returning to Fig. 1, this corresponds to experts allocating sometimes the same feature to more than one Engagement dimension. Whilst these findings correspond with the literature, which shows a positive correlation between Autonomy and Competence (Wangwongwiroj and Bumrabphan 2021; Vlachopoulos and Michailidou 2006; Qin 2021; Gangire et al. 2021), data-driven approaches such as ours may be further explored to feed back to the theories and potential further improvement thereof.

Furthermore, it can be seen from the data in Table 5 that there is some consistency in the way the clusters appear, regardless of the course. As the bold, highest values for the means show, cluster 1, for all 6 analysed courses, tends to have the highest level of Autonomy, Competence and Relatedness (so SDT values), on one hand, as well as the best distribution of student success, taking into consideration the success measures from students' activities (*Completion Ratio, Correct Answers Ratio* and *Reply Ratio)* between week 2 to the last week presented. We can observe that cluster 1 comprises the high achievers, cluster 2 contains the intermediate students, and cluster 3 comprises the students who probably end up dropping out. This is clearly seen in Fig. 5a-f, which shows the centroids of each cluster as a point in the radar plot. In other words, the students with higher Autonomy, Competence and Relatedness in week 1 tend to be the students with higher success measures in later weeks.

Thus, we can say relatively confidently that our process extracts in cluster 1 the most motivated and engaged students, who also turn out to be the most successful. Interestingly, for all these parameters, the mean is statistically significantly higher than for the next best cluster ($p < 0.05$) see Appendix F.

The following cluster, Cluster 2, shows lower means, less engagement and motivation. Cluster 3 gathers a large number of users who are not very engaged, as per SDT parameters, nor are they very successful (as per our three success measures). The mean silhouette coefficients in all courses range from 68 to 78, which shows that our early collected SDT features worked as expected.

Figure 4 further supports these findings and shows that our SDT features can be used to separate the students into three distinct clusters, with distinct success, as per Fig. 5.

Our proposed SDT-based approach has thus been validated by leading to semantically relevant clusters. Indeed, we have clearly confirmed, with this method, facts known from literature (but from theory only, from small-scale studies, mainly using face-to-face data) via our *large-scale study on online data*, from different angles—such as the fact that engaged students have higher success (and have found them all as members of cluster 1). We have also identified the very low engagement students (cluster 3), who also were confirmed to be the least successful. Interestingly, we have identified via cluster 2 students who have some good results, but perhaps lower motivation. This is a very interesting find, because it may show students who would have the potential to complete, to succeed, but may fail, as being less motivated. Thus, some intervention towards motivating these students would have a better chance of an effect than on those in, e.g., cluster 3. Whereas cluster 1 students may come with intrinsic motivation and do not need much 'hand-holding'.

Analysing the outliers (see Appendix G), we have noticed some nuances in the students' behaviours related to the SDT constructs and success measures. It turns out that, contrary to the general trend, there are not only the high-achievers from cluster 1 who have higher Autonomy, Competency and Relatedness, the intermediate students from cluster 2 who have intermediate values of the SDT features and the non-completers from cluster 3 who have low values of the SDT features. In fact, there are also students who have intermediate or low Relatedness and were assigned to cluster 2 or cluster 3, but are high-achievers, according to the success measures. Such students belong to a group that are not engaged in participating in forums or commenting on the pedagogical materials, but are still committed to learning the course content and completing it. These trends can be seen for the other SDT features, that is, a student might not have a high Autonomy or Competence, but be a high- or intermediate achiever, as per our success measures. On the other hand, there are a few students who have high values for the SDT constructs, but do not achieve good results. Nevertheless, this is not the general pattern, but the exception.

We have two possible explanations about these outliers. First, we are collecting data from a very early stage (only the first week data) for each course and, hence, it is natural to have outliers, since student behaviour might change during the course. That is, a student might begin the course with positive attitudes and behaviours—however, end up failing or dropping out due to personal reasons or something that we cannot control. The opposite is possible as well, students might start with apparently bad attitudes and

behaviours, e.g. due to personal problems, but may succeed in concluding the course, due some attitude or circumstances change after the first week.

The second potential reason is that we are dealing with big data and, thus, it is typical to find behaviours that do not follow the general trend. Indeed, this is a characteristic of the human-being, as reported by many authors (Hawkins 1980; Rambo-Hernandez and Warne 2015). This leads to the need of further adaptive systems in MOOCs, to consider such nuances of learning and engaging and how they can influence students' achievement. For further work, we can explore finding optimal subsets (or supersets) of features which would express the SDT (as currently we use the whole set of available features). We could also explore if there are essential features, which lead to a high drop in prediction power for the success variables, and optional features, which only lead to minimal increase in success.

It can be seen from the data in Table 6 that the SDT values (Autonomy, Relatedness and Competence) could be used directly, as good indicators for prediction to enable early interventions for students at-risk (*Non-Active student*) in the following week. The seven classical machine learning algorithms achieved relatively good results (with highest values in bold). Moreover, semi-automatic predictions can be employed, where the SDT-based algorithm informs teachers about non-active, at risk students, employing additionally the teacher's own experience to alleviate for the fluctuations in algorithmic precision. Furthermore, advanced data mining techniques, such as deep learning models may be used; we have not applied them in this study, due the low number of input features (three features), as deep machine learning models are used to find complex and hidden correlations in large input spaces and datasets (Wischmeyer and Rademacher 2020). Alternatively, different definitions of active students may lead to different results, and other predictions, such as completion of course, could be further attempted based on the SDT mapping proposed (Monllaó Olivé 2020, Rawat 2021, Alamri 2019, Tóth 2018; Kameas 2021, Alamri 2021); however, these are beyond the scope of the current paper.

Further to note in terms of the prediction, that whilst the Autonomy construct has the highest mean score (ranging from 0.133 to 0.096) compared to the Competence (ranging from 0.102 to 0.036) and Relatedness (ranging from 0.021 to 0.01) (Appendix E), the Competence construct is ranked as the most important construct in the classification to predict non-active students in the following week (see Appendix C; task definition in 4.9).

Additionally, on a final note on SDT, only three coarse-grained constructs were considered and mapped. Further mappings and evaluations can look into more refined model mapping. For instance, Anderson's ACT-R theory (Anderson et al. 1997) is a good role model for how finer-grained connections can be made between a psychological theory and student behaviour that can inform e-learning technology. However, whilst the results may seem 'commonsensical', in that motivated students would not drop in such high numbers as non-motivated ones, please note that these theories or their constructs had not been evaluated via data-driven approaches previous to this work. These are all interesting avenues for future research, however go beyond the scope of the current paper.

Indeed, mapping large-scale student behaviour onto motivational theories opens the way to inform student models and create appropriate pedagogical interventions to

improve students' outcomes. For instance, students could be brought from cluster 3 to cluster 2, or the desirable cluster 1, by appropriate recommendations. This leads to further avenues of research, bringing together *measurable, data-driven metrics for engagement*, and *classical adaptive learning*. The ultimate goal of such data-driven approaches is not the data processing per se, but their direction of travel, from a current bottom-up cutting-edge approach, to embrace the large body of existent top-down, classical AI approaches, i.e. adaptation informed by pedagogists, psychologist and other specialists. We strongly believe the future will bring these two, seemingly disparate approaches, together, towards much richer solutions.

Finally, however, any research on MOOCs and their engagement needs to note the caveat that MOOCs are not just for traditional students, and many working professionals use them to touch up on certain skills or to explore new areas of knowledge. Once that goal is accomplished, which may occur before the natural end of a registered course, these individuals may quit, having learned and met their goals. The balance between motivation and success would need to take this further into account. Indeed, conducting a pre-survey is one way to identify the students who do not intend to complete the whole course. However, the response rates in MOOCs are generally lower for surveys. For example in (Mihalec-Adkins et al. 2016), only 1,624 completed responses, out of 22,000 students who were enrolled. Therefore, it is likely that MOOC statistics derived from surveys with low response rates would not accurately reflect the real population. Other methods may need to be devised to extract these 'hidden agendas'.

6.1 Threats to validity and future works

As in any study, there are threats to validity, especially when breaking new ground. For instance, we have opted to analyse motivation from the point of view of the SDT theory. Whilst motivation is arguably best analysed from solid theoretical foundations, there are many other motivational theories out there. We have opted for SDT as being one of the most well-known and the use of SDT has become commonplace in the educational domain (Zhou 2016). However, for future research, we will explore others as well (as we show here, we have already mapped *Drive, Engagement Theory, Process of Engagement* in our Engage Taxonomy). This approach can also help in validating theories from a data-intensive point of view, which is interesting for the future. Next, the mapping of student behaviour onto the theory has been done here also for the first time, to our knowledge. Whilst we were careful on checking our mapping with the help of several experts in education and motivational theories, it is possible that we may have missed something (either a construct not being included where needed, or a construct not needed but included—similarly to recall and precision in information retrieval). The current results seem to point to our mapping being successful. However, as this large-scale evaluation of the SDT, and mapping of metrics over engagement theories, is a promising direction of research, this opens the way for further analysis and possible extension of these findings, including increasing the accuracy of the prediction, or looking into ways in which adaptation or intervention built on the motivation-based prediction outcomes (and especially, prediction errors,

such as false positives or false negatives) further influence student success. Other possible avenues would be automatic mapping (to replace expert mapping). or leveraging these features to improve the machine learning models' performance, such as building a hierarchical model based on SDT-feature mapping to predict student engagement. Finally, the Engage Taxonomy can be further refined and improved, including via testing the robustness of the trained machine learning model performance to different definitions of 'success' (e.g. different thresholds).

7 Conclusion

In this paper, we propose a way to confirm the SDT theory (and potentially any motivational theory) via practical experimentation, which we believe is ground-breaking, as it has not been done before in the MOOC online context at this scale. For this purpose, we have proposed a novel, systematic way of analysing engagement, starting from multimodal tracking parameters, following established engagement and motivational theories. We proposed a concrete mapping between the tracking parameters and four of the most used theories of, or related to, engagement in digital systems, generating the *Engage Taxonomy*. We have also showcased *how such a mapping can be put into practice*, by analysing the engaged and disengaged MOOC student behaviours, in relation to the SDT theory. We have *clustered students based on their engagement*, and analysing them via the *connection to their success*. This connection shows that the *results support the SDT theory*, along with its dimensions of *Autonomy*, *Relatedness* and *Competence*. It thus validates the fact that mapping onto concrete features extracted from tracking student behaviour provides reliable, measurable (and thus directly comparable) variables, tested against independent success variables for the student. These findings are based on a large scale, data-driven study, where similar consistent results were obtained over several runs of the courses. This clearly supports a theoretically-rooted approach on how to characterise engaged and disengaged MOOC student behaviours and exploring what triggers and promotes MOOC students' interest and engagement. We have further *used these extracted SDT constructs directly as early predictors of Active versus Non-Active students*, showing *successful results with several machine learning methods*.

Concluding, we have created a mapping methodology of engagement, including the whole process from design, mapping, measuring, and evaluation, which can be further applied not only to MOOCs and e-learning systems when exploring engagement along the SDT constructs, but also in terms of the mapping of other engagement theories, such as Drive (Pink 2011), Engagement (Kearsley and Shneiderman 1998), and Process of Engagement (O'Brien and Toms 2008) onto data-intensive applications.

Appendix A: Pearson correlation coefficient

(1) Correlation between the SDT constructs and the success measures over the 6 courses

		Completion rate	Answer rate	Reply rate
Shakespeare	Autonomy	0.60	0.56	0.17
	Competence	**0.61**	**0.59**	0.22
	Relatedness	0.29	0.29	**0.36**
Babies in mind	Autonomy	**0.57**	**0.50**	0.24
	Competence	0.54	0.49	0.31
	Relatedness	0.30	0.29	**0.32**
LMPCC	Autonomy	0.68	0.62	0.15
	Competence	0.66	**0.63**	0.21
	Relatedness	0.26	0.23	**0.30**
OI	Autonomy	**0.62**	**0.57**	0.14
	Competence	0.57	0.54	0.24
	Relatedness	0.29	0.29	**0.25**
Supply Chains	Autonomy	**0.64**	0.59	0.18
	Competence	0.63	**0.65**	0.18
	Relatedness	0.28	0.26	**0.25**
TMF	Autonomy	0.64	0.60	0.26
	Competence	**0.66**	**0.64**	0.29
	Relatedness	0.26	0.25	**0.40**

(2) Correlation between SDT constructs (Autonomy, Competence and Relatedness) over the 6 courses

		Autonomy	Competence	Relatedness
Shakespeare	Autonomy	–	0.93	0.60
	Competence	0.93	–	0.57
	Relatedness	0.60	0.57	–
Babies in mind	Autonomy	–	0.87	0.73
	Competence	0.87	–	0.75
	Relatedness	0.73	0.75	–
LMPCC	Autonomy	–	0.89	0.61
	Competence	0.89	–	0.62
	Relatedness	0.61	0.62	–
OI	Autonomy	–	0.86	0.75
	Competence	0.86	–	0.84
	Relatedness	0.75	0.84	–
Supply Chains	Autonomy	–	0.87	0.69
	Competence	0.87	–	0.56
	Relatedness	0.69	0.56	–
TMF	Autonomy	–	0.92	0.62
	Competence	0.92	–	0.58
	Relatedness	0.62	0.58	–

Appendix B: Prediction of active and non-active students in week 2 measured over the 6 courses

BA: Balanced Accuracy, ACC: Accuracy, PR: Precision, Re: Recall, F1: F1-Score. The bold values show the overall results for both classes (0 and 1).

Course	Classifier	Active 1/Non-Active 0	Precision (%)	Re (%)	F1 (%)	**ACC** (%)	**BA** (%)
The mind is flat	KNeighbors	0	94.60	91.69	93.12	**89.43**	**86.54**
		1	73.39	81.40	77.19		
	LogisticRe	0	98.01	89.05	93.31	**90.04**	**91.31**
		1	70.64	93.57	80.51		
	XGBClassif	0	97.65	90.04	93.69	**90.53**	**91.17**
		1	72.29	92.31	81.08		

Course	Classifier	Active 1/Non-Active 0	Precision (%)	Re (%)	F1 (%)	**ACC** (%)	**BA** (%)
	AdaBoostCl	0	97.84	89.21	93.33	**90.04**	**91.11**
		1	70.82	93.01	80.41		
	GradientBo	0	97.69	89.88	93.62	**90.44**	**91.16**
		1	72.00	92.45	80.96		
	ExtraTrees	0	98.14	89.41	93.57	**90.41**	**91.70**
		1	71.41	93.99	81.16		
	RandomFore	0	98.06	89.60	93.64	**90.50**	**91.65**
		1	71.73	93.71	81.26		
Supply chains	KNeighbors	0	94.25	93.73	93.99	**90.16**	**83.74**
		1	71.95	73.75	72.84		
	LogisticRe	0	98.47	87.74	92.80	**88.81**	**90.74**
		1	62.50	93.75	75.00		
	XGBClassif	0	98.81	90.46	94.45	**91.28**	**92.73**
		1	68.47	95.00	79.58		
	AdaBoostCl	0	98.22	90.19	94.03	**90.60**	**91.35**
		1	67.27	92.50	77.89		
	GradientBo	0	98.81	90.46	94.45	**91.28**	**92.73**
		1	68.47	95.00	79.58		
	ExtraTrees	0	98.80	89.92	94.15	**90.83**	**92.46**
		1	67.26	95.00	78.76		
	RandomFore	0	98.52	90.46	94.32	**91.05**	**92.11**
		1	68.18	93.75	78.95		
Open innovation in business (OI)	KNeighbors	0	94.98	88.93	91.85	**88.16**	**87.39**
		1	72.03	85.86	78.34		
	LogisticRe	0	92.58	87.92	90.19	**85.64**	**83.35**
		1	68.42	78.79	73.24		
	XGBClassif	0	96.03	89.26	92.52	**89.17**	**89.08**
		1	73.33	88.89	80.37		
	AdaBoostCl	0	98.43	84.23	90.78	**87.15**	**90.09**
		1	66.90	95.96	78.84		
	GradientBo	0	92.78	90.60	91.68	**87.66**	**84.70**
		1	73.58	78.79	76.10		
	ExtraTrees	0	96.39	89.60	92.87	**89.67**	**89.75**
		1	74.17	89.90	81.28		
	RandomFore	0	95.34	89.26	92.20	**88.66**	**88.07**
		1	72.88	86.87	79.26		
Shakespeare	KNeighbors	0	87.88	84.06	85.93	**81.72**	**80.57**
		1	70.99	77.09	73.91		
	LogisticRe	0	96.58	80.09	87.57	**84.90**	**87.24**
		1	70.58	94.39	80.77		
	XGBClassif	0	96.17	80.83	87.84	**85.14**	**87.24**

Course	Classifier	Active 1/Non-Active 0	Precision (%)	Re (%)	F1 (%)	**ACC** (%)	**BA** (%)
		1	71.20	93.65	80.89		
	AdaBoostCl	0	96.45	80.02	87.47	**84.78**	**87.11**
		1	70.47	94.19	80.62		
	GradientBo	0	96.33	80.74	87.85	**85.17**	**87.33**
		1	71.17	93.92	80.97		
	ExtraTrees	0	96.47	80.38	87.69	**85.02**	**87.29**
		1	70.84	94.19	80.86		
	RandomFore	0	96.49	80.30	87.65	**84.98**	**87.26**
		1	70.76	94.22	80.82		
Leading and managing people-centred change (LMPCC)	KNeighbors	0	95.36	85.76	90.31	**86.48**	**87.12**
		1	69.22	88.47	77.67		
	LogisticRe	0	96.95	83.63	89.80	**86.05**	**88.18**
		1	67.21	92.73	77.94		
	XGBClassif	0	97.73	84.04	90.37	**86.85**	**89.32**
		1	68.20	94.61	79.27		
	AdaBoostCl	0	99.07	81.86	89.64	**86.11**	**89.86**
		1	66.13	97.87	78.93		
	GradientBo	0	98.24	83.40	90.21	**86.71**	**89.63**
		1	67.64	95.86	79.32		
	ExtraTrees	0	98.66	83.40	90.39	**86.98**	**90.13**
		1	67.87	96.87	79.81		
	RandomFore	0	97.99	83.90	90.40	**86.91**	**89.57**
		1	68.16	95.24	79.46		
Babies in mind	KNeighbors	0	91.26	81.88	86.31	**79.19**	**75.13**
		1	48.35	68.37	56.64		
	LogisticRe	0	97.51	73.59	83.88	**77.34**	**83.01**
		1	46.47	92.43	61.85		
	XGBClassif	0	86.96	88.45	87.70	**80.12**	**67.50**
		1	50.00	46.55	48.21		
	AdaBoostCl	0	97.76	72.43	83.21	**76.58**	**82.87**
		1	45.64	93.32	61.30		
	GradientBo	0	84.70	91.77	88.09	**80.12**	**62.48**
		1	50.00	33.18	39.89		
	ExtraTrees	0	95.60	74.48	83.73	**76.80**	**80.33**
		1	45.58	86.19	59.63		
	RandomFore	0	97.76	72.43	83.21	**76.58**	**82.87**
		1	45.64	93.32	61.30		

Appendix C: The constructs importance values for the extra trees algorithm

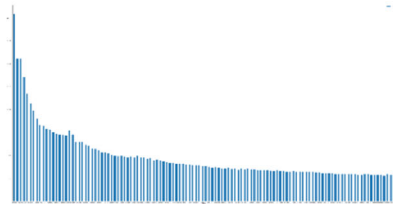|	Autonomy	Competence	Relatedness
Shakespeare	0.31	0.48	0.20
Babies in mind	0.34	0.43	0.23
LMPCC	0.29	0.48	0.23
OI	0.33	0.45	0.22
Supply chains	0.32	0.47	0.21
TMF	0.27	0.50	0.22

Appendix D: Remaining students over time in different courses (a-f)
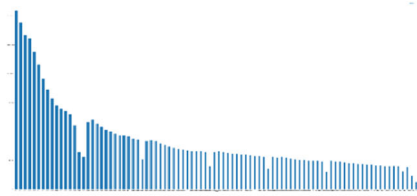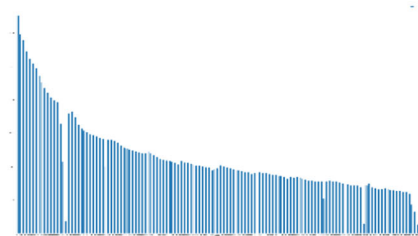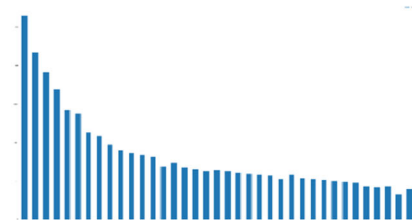

a) The mind is flat
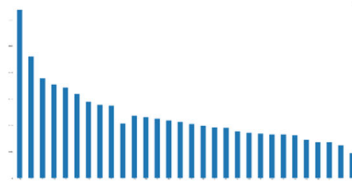
b) Lmpcc
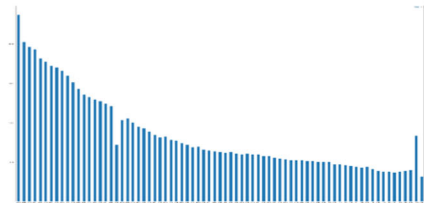

c) Supply chains

d) Shakespeare

e) OI

f) Babies in mind

Appendix E: Mean, standard deviation and maximum value for the 3 SDT constructs

		Autonomy	Competence	Relatedness
Shakespeare	Mean	0.13	0.10	0.01
	Std	0.11	0.11	0.04
	Max	0.73	0.74	0.70
Babies in mind	Mean	0.13	0.04	0.01
	Std	0.10	0.06	0.05

		Autonomy	Competence	Relatedness
	Max	0.67	0.62	0.70
LMPCC	Mean	0.10	0.04	0.01
	Std	0.09	0.06	0.04
	Max	0.66	0.66	0.75
OI	Mean	0.09	0.03	0.02
	Std	0.10	0.07	0.06
	Max	0.56	0.59	0.61
Supply chains	Mean	0.09	0.05	0.01
	Std	0.11	0.10	0.05
	Max	0.74	0.73	0.68
TMF	Mean	0.11	0.06	0.01
	Std	0.10	0.10	0.04
	Max	0.82	0.79	0.82

Appendix F: Statistical significance analysis ($p < 0.05$) of the difference between the highest two clusters (cluster1 vs cluster2)

Course	Completion ratio	Correct answers ratio	Reply ratio
Babies in mind	*4.30947e-54*	*1.97756e-46*	*0.0*
Shakespeare	*1.35573e-69*	*7.74394e-53*	*0.0*
Supply chains	*0.004673*	*0.02695*	*2.45989e-30*
The mind is flat	*1.34367e-24*	*1.54078e-23*	*0.0*
Open innovation in business (OI)	*6.440039e-4*	*5.668406e-06*	*8.801509e-63*
LMPCC	*4.544427e-15*	*0.000159*	*9.613372e-74*

The appendix F shows all p-values < 0.05 so all differences are statistically significant

Appendix G: Number and percentage of the outlier students for each SDT element in each cluster

Values greater than the third quartile (Upper Bound (Q3 + (1.5 * IQR)) or less than first quartile (Lower Bound (Q1—(1.5 * IQR)) are considered outliers. Where Q1 is the middle number between the lowest and the median values of the dataset. Q3 is the middle number between the median and the maximum value in the dataset.

Number and percentage of the outlier students for each SDT element in each cluster.

	Course	Cluster	AUT	COM	REL
[name-removed1] university	Babies in mind	Cluster 1	N = 53-%2.5	N = 127-%6.18	N = 92-%4.4
		Cluster 2	N = 1902-%18.7	N = 439-%4.32	N = 599-%5.8
		Cluster 3	N = 0-%0.0	N = 496-%3.55	N = 484-%3.4
	Shakespeare	Cluster 1	N = 18-%1.2	N = 44-%3.02	N = 47-%3.2
		Cluster 2	N = 829-%6.7	N = 579-%4.72	N = 2750-%22.4
		Cluster 3	N = 86-%0.5	N = 573-%3.64	N = 418-%2.6
	Supply chains	Cluster 1	N = 0-%0.0	N = 3-%3.12	N = 6-%6.2
		Cluster 2	N = 68-%10.7	N = 40-%6.33	N = 120-%19.0
		Cluster 3	N = 239-%10.9	N = 57-%2.60	N = 114-%5.2
	The mind is flat	Cluster 1	N = 14-%0.9	N = 27-%1.91	N = 45-%3.1
		Cluster 2	N = 1221-%9.7	N = 953-%7.61	N = 2579-%20.5
		Cluster 3	N = 191-%0.7	N = 1069-%4.11	N = 897-%3.4
[name-removed2] university	Open innovation in business (OI)	Cluster 1	N = 0-%0.0	N = 6-%3	N = 4-%0.2.4
		Cluster 2	N = 102-%13	N = 22-%3	N = 88-%11

Course	Cluster	AUT	COM	REL
	Cluster 3	N = 0-%0.0	N = 41-%2.2	N = 44-%2.3
LMPCC	Cluster 1	N = 4-%1.2	N = 10-%3.18	N = 15-%4.7
	Cluster 2	N = 220-%9.7	N = 104-%4.61	N = 281-%12.4
	Cluster 3	N = 227-%5.6	N = 69-%1.72	N = 76-%1.9

Appendix H: Example of classifying sentiment of students' comments using TextBlob

Comments	Sentiment
"Amazing and interesting hypothesis. With this article I really learnt something new today! Thank you!"	Positive comment
"In China, I am not able to use it. And i feel sorry."	Negative comment
"I've loved this course. It's been well designed and easy to keep up with. Thank you!"	Positive comment
"I dont agree with this..i think its a wrong starting point for this course! even the trend line(i mean the linearity in graph) is not correct just looking at it"	Negative comment

Appendix I: Prediction of active and non-active students in week 2 based on week 1 SDT constructs (tenfold cross validation); evaluated by the Balanced Accuracy score

Courses	AdaB	ExtrTr	GBoost	KNN	LR	RF	XGBoost
Babies in mind	83.93%	81.90%	70.13%	70.34%	84.17%	83.93%	72.69%

Courses	AdaB	ExtrTr	GBoost	KNN	LR	RF	XGBoost
Shakespeare	86.76%	86.67%	87.01%	83.81%	86.84%	86.87%	86.96%
Supply chains	91.30%	91.12%	88.88%	86.93%	91.22%	91.32%	90.69%
The mind is flat	90.08%	89.99%	90.13%	85.36%	90.23%	90.07%	90.51%
Open innovation in business (OI)	90.74%	90.14%	87.52%	87.24%	88.04%	90.27%	91.36%
(LMPCC)	89.55%	89.34%	88.97%	84.89%	87.59%	89.36%	89.57%

Appendix J: Prediction of active and non-active students in week 2 based on one of SDT construct (Relatedness); evaluated by the balanced accuracy score

Courses	AdaB	ExtrTr	GBoost	KNN	LR	RF	XGBoost
Babies in mind	50.00%	58.21%	55.53%	55.22%	58.85%	60.47%	55.45%
Shakespeare	63.00%	61.44%	62.94%	61.89%	62.43%	63.00%	63.00%
Supply chains	55.26%	56.97%	55.77%	56.22%	57.13%	57.12%	54.90%
The mind is flat	59.81%	59.70%	59.51%	58.60%	61.55%	59.70%	59.75%
Open innovation in business (OI)	50.68%	56.07%	54.55%	51.86%	56.24%	56.24%	51.35%
(LMPCC)	57.27%	56.40%	55.95%	55.68%	58.62%	56.38%	55.65%

Appendix K: *k*-means clustering for The Mind is Flat course using only one dimension of the SDT constructs ("Relatedness")

			Cluster 1	Cluster 2	Cluster 3
			Mean/SD	Mean/SD	Mean/SD
The Mind is Flat	Relatedness from Students' activities in the first week		**0.21/0.1**	0.01/0.03	0.0/0.01
	Success measures from students' activities between week 2 to the last week	Completion ratio	**0.55/0.37**	0.43/0.39	0.15/0.29
		Correct answers ratio	**0.41/0.31**	0.31/0.31	0.1/0.22
		Reply ratio	**0.04/0.08**	0.01/0.03	0.0/0.0

References

Adamopoulos, P.: 2013. What makes a great MOOC? An interdisciplinary analysis of student retention in online courses.

Adnan, M., Habib, A., Ashraf, J., Mussadiq, S., Raza, A.A., Abid, M., Bashir, M., Khan, S.U.: Predicting at-risk students at different percentages of course length for early intervention using machine learning models. IEEE Access **9**, 7519–7539 (2021)

Alamri, A. A. M., Cristea, A., Pereira, F. D., Oliveira, E., Shi, L., Stewart, C.: Predicting MOOCs dropout using only two easily obtainable features from the first week's activities. *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)* (2019)

Alamri, A. S. Z. T., Cristea, A. I., Stewart, C., Pereira, F. D.: MOOC next week dropout prediction: weekly assessing time and learning patterns. *Intelligent tutoring systems (ITS 2021)*

Alamri, A. S. Z., Cristea, A. I., Senthilnathan, G., Shi, L., Stewart, C.: Is MOOC learning different for dropouts? A visually-driven, multi-granularity explanatory ML approach. In: *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)* (2020)

Alarcon, G.M., Edwards, J.M.: The relationship of engagement, job satisfaction and turnover intentions. Stress Health **27**, e294–e298 (2011)

Alshehri, M., Alamri, A., Cristea, A.I., Stewart, C.D.: Towards designing profitable courses: predicting student purchasing behaviour in MOOCs. Int. J. Artif. Intelligence Educ. **31**, 215–233 (2021)

Alshehri, M., Foss, J., Cristea, A. I., Kayama, M., Shi, L., Alamri, A., Tsakalidis, A. On the need for fine-grained analysis of gender versus commenting behaviour in MOOCs. In: Proceedings of the 2018 the 3rd international conference on information and education innovations, p. 73–77 (2018)

Amirkhan, J.H.: Criterion validity of a coping measure. J. Personal. Assess. **62**, 242–261 (1994)

An, S., Liu, W., Venkatesh, S.: Fast cross-validation algorithms for least squares support vector machine and kernel ridge regression. Pattern Recogn. **40**, 2154–2162 (2007)

Anchalia, P. P., Roy, K.: The k-nearest neighbor algorithm using MapReduce paradigm. In: 2014 5th International conference on intelligent systems, modelling and simulation, p. 513–518. IEEE (2014)

Anderson, J.R., Matessa, M., Lebiere, C.: ACT-R: a theory of higher level cognition and its relation to visual attention. Hum. Comput. Interaction **12**, 439–462 (1997)

Arroyo, I., Ferguson, K., Johns, J., Dragon, T., Meheranian, H., Fisher, D., Barto, A., Mahadevan, S., Woolf, B. P.: Repairing disengagement with non-invasive interventions. AIED, p. 195–202 (2007)

Bakharia, A: Towards cross-domain MOOC forum post classification. In: Proceedings of the third (2016) ACM conference on learning@ scale, p. 253–256 (2016)

Barak, M., Watted, A., Haick, H.: Motivation to learn in massive open online courses: examining aspects of language and social engagement. Comput. Educ. **94**, 49–60 (2016)

Benesty, J., Chen, J., Huang, Y., Cohen, I.: Pearson correlation coefficient. Noise reduction in speech processing, Springer (2009)

Breiman, L.: Random forests. Mach. Learn. **45**, 5–32 (2001)

Brinton, C.G., Chiang, M., Jain, S., Lam, H., Liu, Z., Wong, F.M.F.: Learning about social learning in MOOCs: from statistical analysis to generative model. IEEE Trans. Learn. Technol. **7**, 346–359 (2014)

Brodersen, K. H., Ong, C. S., Stephan, K. E., Buhmann, J. M.: The balanced accuracy and its posterior distribution. In: 2010 20th International conference on pattern recognition. p. 3121–3124. IEEE (2010)

Chen, W., Brinton, C.G., Cao, D., Mason-Singh, A., Lu, C., Chiang, M.: Early detection prediction of learning outcomes in online short-courses via learning behaviors. IEEE Trans. Learn. Technol. **12**(1), 44–58 (2018)

Chen, W., Brinton, C.G., Cao, D., Mason-Singh, A., Lu, C., Chiang, M.: Early detection prediction of learning outcomes in online short-courses via learning behaviors. IEEE Trans. Learn. Technol. **12**(1), 44–58 (2018)

Chen, T., Guestrin, C.: Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. p. 785–794 (2016)

Chirkov, V.I.: A cross-cultural analysis of autonomy in education: a self-determination theory perspective. Theory Res. Educ. **7**, 253–262 (2009)

Cocea, M. Learning engagement: what actions of learners could best predict it? AIED. Citeseer, p. 683–684 (2007)

Courtney, A: Self-determination theory of motivation: why intrinsic motivation matters (2018)

Creveling, C.M., Hambleton, L., Mccarthy, B.: Six sigma for marketing processes: an overview for marketing executives, leaders, and managers (2006)

Cristea, A.I.A.A., Kayama, M., Stewart, C., Alshehri, M., Shi, L.: Earliest predictor of dropout in MOOCs: a longitudinal study of futurelearn courses. In: Proceedings of the 27th international conference on information systems development: designing digitalization, ISD (2018)

Davis, H., Leon, K.D.M., Vera, M., White, S. MOOCs for universities and learners. In: An analysis of motivating factors (2013)

De Barba, P.G., Kennedy, G.E., Ainley, M.D.: The role of students' motivation and participation in predicting performance in a MOOC. J. Comput. Assist. Learn. **32**, 218–231 (2016)

De Freitas, S.I., Morgan, J., Gibson, D.: Will MOOCs transform learning and teaching in higher education? Engagement and course retention in online learning provision. Br. J. Educ. Technol. **46**, 455–471 (2015)

Deci, E.L., Ryan, R.M.: Intrinsic motivation and self-determination in human behavior. Springer (2013)

Deng, R., Benckendorff, P., Gannaway, D.: Learner engagement in MOOCs: Scale development and validation. Br. J. Edu. Technol. **51**, 245–262 (2020)

Dorfman, R.: A formula for the gini coefficient. In: The review of economics and statistics, pp. 146–149. JSTOR, Ann Arbor (1979)

Duncan, N.J., Strevens, C., Field, R.: Resilience and student wellbeing in higher education: a theoretical basis for establishing law school responsibilities for helping our students to thrive. Eur. J. Legal Educ. **1**, 83–115 (2020)

Dutta, R.: To Find the best-suited model for sentiment analysis of real-time twitter data. In: International conference on innovative computing and communications, pp. 445–452. Springer (2021)

Evans, P.: Self-determination theory: an approach to motivation in music education. Music. Sci. **19**, 65–83 (2015)

Ferguson, R., Clow, D., Beale, R., Cooper, A. J., Morris, N., Bayne, S., Woodgate, A.: Moving through MOOCS: pedagogy, learning design and patterns of engagement. In: European conference on technology enhanced learning, p. 70–84. Springer (2015)

Fleiss, J.L., Levin, B., Paik, M.C.: The measurement of interrater agreement. Stat. Methods Rates Proportions **2**, 22–23 (1981)

Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci. **55**, 119–139 (1997)

Friedman, J.H.: Greedy function approximation: a gradient boosting machine. Ann. Stat. **29**(5), 1189–1232 (2001)

Gangire, Y., Da Veiga, A., Herselman, M.: Assessing information security behaviour: a self-determination theory perspective. Inf. Comput. Sec. **29**(4), 625–646 (2021)

Gardner, J., Brooks, C.: Student success prediction in MOOCs. User Model. User-Adap. Inter. **28**, 127–203 (2018)

Gerber, Z., Anaki, D.: The role of self-compassion, concern for others, and basic psychological needs in the reduction of caregiving burnout. Mindfulness **12**, 741–750 (2021)

Geurts, P., Ernst, D., Wehenkel, L.: Extremely randomized trees. Mach. Learn. **63**, 3–42 (2006)

Gryllos, P., Makris, C., Vikatos, P.: Marketing campaign targeting using bridge extraction. In: Proceedings of the symposium on applied computing. p. 1045–1052 (2017)

Hagger, M.S., Hamilton, K.: General causality orientations in self-determination theory: Meta-analysis and test of a process model. Eur. J. Personal. **35**, 710–735 (2021)

Hawkins, D.M.: Identification of outliers. Springer, Cham (1980)

Hofer, J., Busch, H.: Satisfying one's needs for competence and relatedness: consequent domain-specific well-being depends on strength of implicit motives. Personal. Soc. Psychol. Bull. **37**, 1147–1158 (2011)

Jackson, G. T., Graesser, A. C., Mcnamara, D. S.: What students expect may have more impact than what they know or feel. In: AIED. p. 73–80 (2009)

Jiang, C.Q., Loerbroks, A., Lam, K.-B.H., Bosch, J.A., Thomas, G.N., Zhang, W.S., Cheng, K.K., Lam, T.H., Adab, P.: Mental health and asthma in China: the Guangzhou biobank cohort study. Int. J. Behav. Med. **20**, 259–264 (2013)

Jordan, K.: Massive open online course completion rates revisited: assessment, length and attrition. Int. Rev. Res. Open Distrib. Learn. **16**, 341–358 (2015)

Kameas, G.K.T.P.S.K.C.P.A.: Interpretable Models for early prediction of certification in MOOCs: a case study on a mooc for smart city professionals. IEEE Access **9**, 165881–165891 (2021)

Kearsley, G., Shneiderman, B.: Engagement theory: a framework for technology-based teaching and learning. Educ. Technol. **38**, 20–23 (1998)

Khodeir, N.A.: Bi-GRU urgent classification for MOOC discussion forums based on BERT. IEEE Access **9**, 58243–58255 (2021)

Moreno-Marcos, P. M., Alario-Hoyos, C., Muñoz-Merino, P. J., Estévez-Ayres, I., Kloos, C. D.: Sentiment analysis in MOOCs: a case study. In: 2018 IEEE global engineering education conference (EDUCON). p. 1489–1496 (2018)

Kuh, G.D.: What we're learning about student engagement from NSSE: benchmarks for effective educational practices. Chang. Mag. High. Learn. **35**, 24–32 (2003)

Kumar, K. S., Semwal, V. B., Tripathi, R. C.: Real time face recognition using adaboost improved fast PCA algorithm (2011). *arXiv preprint* arXiv:1108.1353

Lan, M., Hew, K.F.: Examining learning engagement in MOOCs: a self-determination theoretical perspective using mixed method. Int. J. Educ. Technol. High. Educ. **17**, 1–24 (2020)

Langdon, J., Webster, C., Hall, T., Monsma, E.: A self-determination theory perspective of student performance at the end of a volleyball unit in compulsory high school physical education. Sport Sci. Pract. Asp. **11**, 5–16 (2014)

Lavigne, G.L., Vallerand, R.J., Miquelon, P.: A motivational model of persistence in science education: a self-determination theory approach. Eur. J. Psychol. Educ. **22**, 351–369 (2007)

Littlejohn, A., Hood, N., Milligan, C., Mustain, P.: Learning in MOOCs: motivations and self-regulated learning in MOOCs. Internet High. Educ. **29**, 40–48 (2016)

Lohar, P., Xie, G., Bendechache, M., Brennan, R., Celeste, E., Trestian, R., Tal, I.: Irish attitudes toward COVID tracker app & privacy: sentiment analysis on Twitter and survey data. In: The 16th international conference on availability, reliability and security. p. 1–8 (2021)

Loizzo, J., Ertmer, P.A., Watson, W.R., Watson, S.L.: Adult MOOC learners as self-directed: perceptions of motivation, success, and completion. Online Learn. **21**, n2 (2017)

Marczewski, A.: Even Ninja monkeys like to play. Lond. Blurb Inc **31**, 541–553 (2015)

Mazini, M., Shirazi, B., Mahdavi, I.: Anomaly network-based intrusion detection system using a reliable hybrid artificial bee colony and AdaBoost algorithms. J. King Saud Univ. Comput. Inf. Sci. **31**, 541–553 (2019)

Mihalec-Adkins, B., Hicks, N., Douglas, K. A., Diefes-Dux, H., Bermel, P., Madhavan, K.: Surveying the motivations of groups of learners in highly-technical STEM MOOCs. In: 2016 IEEE frontiers in education conference (FIE), p. 1–6. IEEE (2016)

Mishina, Y., Murata, R., Yamauchi, Y., Yamashita, T., Fujiyoshi, H.: Boosted random forest. IEICE Trans. Inf. Syst. **98**, 1630–1636 (2015)

Mohamed, U.U.E., Salleh, N.: Measuring the success of massive open online courses: a mixed-method case study. In: 2021 IEEE Asia-Pacific conference on computer science and data engineering (CSDE), p. 1–5. IEEE (2021)

Monllaó Olivé, D., Huynh, D.Q., Reynolds, M., Dougiamas, M., Wiese, D.: A supervised learning framework: using assessment to identify students at risk of dropping out of a MOOC. J. Comput. High. Educ. **32**, 9–26 (2020)

Moreno-Marcos, P.M., Pong, T.-C., Munoz-Merino, P.J., Kloos, C.D.: Analysis of the factors influencing learners' performance prediction with learning analytics. IEEE Access **8**, 5264–5282 (2020)

Moreno-Marcos, P. M., Alario-Hoyos, C., Munoz-Merino, P. J., Estevez-Ayres, I., Kloos, C. D.: Sentiment analysis in MOOCs: a case study. In: IEEE global engineering education conference, EDUCON. p. 1489–1496 (2018)

Moreno-Marcos, P. M., Alario-Hoyos, C., Muñoz-Merino, P. J., Estévez-Ayres, I., Kloos, C. D.: Sentiment analysis in MOOCs: A case study. In: 2018 IEEE global engineering education conference (EDUCON), p. 1489–1496. IEEE (2018)

Moreno-Murcia, J.A., Gimeno, E.C., Hernández, E.H., Belan-Do Pedreño, N., Marín, J.J.R.: Motivational profiles in physical education and their relation to the theory of planned behavior. J. Sports Sci. Med. **12**, 551 (2013)

Mubarak, A.A., Han, C., Zhang, W.: Prediction of students' early dropout based on their interaction logs in online learning environment. Interact. Learn. Environ. **3**, 1–20 (2020)

Nam, S., Frishkoff, G., Collins-Thompson, K.: Predicting students disengaged behaviors in an online meaning-generation task. IEEE Trans. Learn. Technol. **11**, 362–375 (2017)

Nevo, B.: Face validity revisited. J. Educ. Meas. **22**, 287–293 (1985)

O'brien, H.L., Toms, E.G.: What is user engagement? A conceptual framework for defining user engagement with technology. J. Am. Soc. Inf. Sci. Technol. **59**, 938–955 (2008)

Oyelade, O., Oladipupo, O. O., Obagbuwa, I. C.: Application of k means clustering algorithm for prediction of students academic performance (2010). *arXiv preprint* arXiv:1002.2425

Pardo, A., Han, F., Ellis, R.A.: Combining university student self-regulated learning indicators and engagement with online learning events to predict academic performance. IEEE Trans. Learn. Technol. **10**, 82–92 (2016)

Patro, S., Sahu, K. K.: Normalization: a preprocessing stage (2015). *arXiv preprint* arXiv:1503.06462

Phillips, J.M.: Strategies for active learning in online continuing education. J. Contin. Educ. Nurs. **36**, 77–83 (2005)

Pink, D.H.: Drive: the surprising truth about what motivates us. Penguin (2011)

Qin, Y.: Attractiveness of game elements, presence, and enjoyment of mobile augmented reality games: The case of Pokémon go. Telemat. Inform. **62**, 101620 (2021)

Rambo-Hernandez, K.E., Warne, R.T.: Measuring the outliers: an introduction to out-of-level testing with high-achieving students. Teach. Except. Child. **47**, 199–207 (2015)

Rana, S., Garg, R.: Application of hierarchical clustering algorithm to evaluate students performance of an institute. In: 2016 Second international conference on computational intelligence & communication technology (CICT), p. 692–697. IEEE (2016)

Rawat, S., Kumar, D., Kumar, P., Khattri, C.: A systematic analysis using classification machine learning algorithms to understand why learners drop out of MOOCs. Neural Comput. Appl. **33**, 14823–14835 (2021)

Rawlings, J.O., Pantula, S.G., Dickey, D.A.: Applied regression analysis: a research tool. Springer, Cham (1998)

Shen, B., Mccaughtry, N., Martin, J., Fahlman, M.: Effects of teacher autonomy support and students' autonomous motivation on learning in physical education. Res. Q. Exerc. Sport **80**, 44–53 (2009)

Shernoff, D.J., Ruzek, E.A., Sinha, S.: The influence of the high school classroom environment on learning as mediated by student engagement. Sch. Psychol. Int. **38**, 201–218 (2017)

Shi, L., Cristea, A. I.: In-depth exploration of engagement patterns in MOOCs. In: International conference on web information systems engineering, p. 395–409. Springer, 2018b

Shi, L., Cristea, A.I.: Demographic indicators influencing learning activities in MOOCs: learning analytics of FutureLearn courses. Association for Information Systems, Atlanta, Georgia (2018)

Shi, L., Cristea, A. I., Toda, A. M., Oliveira, W.: Revealing the hidden patterns: a comparative study on profiling subpopulations of MOOC students (2020). *arXiv preprint* arXiv:2008.05850

Shorfuzzaman, M., Hossain, M.S., Nazir, A., Muhammad, G., Alamri, A.: Harnessing the power of big data analytics in the cloud to support learning analytics in mobile learning environment. Comput. Hum. Behav. **92**, 578–588 (2019)

Standage, M., Duda, J.L., Ntoumanis, N.: A test of self-determination theory in school physical education. Br. J. Educ. Psychol. **75**, 411–433 (2005)

Stocks, T., Renders, C., Bulk-Bunschoten, A., Hirasing, R., Van Buuren, S., Seidell, J.: Body size and growth in 0-to 4-year-old children and the relation to body size in primary school age. Obes. Rev. **12**, 637–652 (2011)

Stone, J.E.: Self-determination theory and MOOC enrollment motivation: validation of the online learning enrollment intentions scale. Oklahoma State University, Stillwater, OK (2021)

Sunar, A.S., White, S., Abdullah, N.A., Davis, H.C.: How learners' interactions sustain engagement: a MOOC case study. IEEE Trans. Learn. Technol. **10**, 475–487 (2016)

Sunar, A.S., White, S., Abdullah, N.A., Davis, H.C.: How learners' interactions sustain engagement: A MOOC case study. IEEE Trans. Learn. Technol. **10**, 475–487 (2017)

Korosi, G., Esztelecki, P., Farkas, R, Tóth, K.: Clickstream-based outcome prediction in short video MOOCs. In: 2018 International conference on computer, information and telecommunication systems (CITS), p. 1–5 (2018)

Vasconcellos, D., Parker, P.D., Hilland, T., Cinelli, R., Owen, K.B., Kapsal, N., Lee, J., Antczak, D., Ntoumanis, N., Ryan, R.M.: Self-determination theory applied to physical education: a systematic review and meta-analysis. J. Educ. Psychol. **112**, 1444 (2020)

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. In: Advances in neural information processing systems, p. 30 (2017)

Vlachopoulos, S.P., Michailidou, S.: Development and initial validation of a measure of autonomy, competence, and relatedness in exercise: the basic psychological needs in exercise scale. Meas. Phys. Educ. Exerc. Sci. **10**, 179–201 (2006)

Vyas, J. D., Han, M.: Understanding the mobile game app activity. In: Proceedings of the 2019 ACM Southeast conference, p. 206–209 (2019)

Walker, E., Ogan, A.: We're in this together: intentional design of social relationships with AIED systems. Int. J. Artif. Intell. Educ. **26**, 713–729 (2016)

Wang, X., Yang, D., Wen, M., Koedinger, K. & Rosé, C. P.: Investigating how student's cognitive behavior in MOOC discussion forums affect learning gains. In: International educational data mining society (2015)

Wangwongwiroj, T., Bumrabphan, K.: Self-determination theory: statistical correlations between motivational regulations and basic psychological needs. Int. J. High. Educ. Pedag. **2**, 53–58 (2021)

Wen, M., Yang, D., Rosé, C.: Linguistic reflections of student engagement in massive open online courses. In: Proceedings of the international AAAI conference on web and social media. p. 525–534 (2014)

Willms, J.D.: Student engagement at school. A sense of belonging and participation. Organisation for Economic Co-operation and Development, Paris (2003)

Wischmeyer, T., Rademacher, T.: Regulating artificial intelligence. Springer, Cham (2020)

Witchel, H. J.: Engagement: the inputs and the outputs: conference overview. In: Proceedings of the 2013 inputs-outputs conference: an interdisciplinary conference on engagement in HCI and performance.p. 1–4 (2013)

Xu, R., Wunsch, D.: Survey of clustering algorithms. IEEE Trans. Neural Netw. **16**, 645–678 (2005)

Zhang, C., Ma, Y.: Ensemble machine learning: methods and applications. Springer, Cham (2012)

Zhou, M.: Chinese university students' acceptance of MOOCs: a self-determination perspective. Comput. Educ. **92**, 194–203 (2016)

Alexandra I. Cristea is Professor, Head of the Artificial Intelligence and Human Systems (AIHS) group and Deputy Head at the Computer Science Department, Durham University; Honorary Professor at the Computer Science Department, Warwick University. Her research includes web science, learning analytics, user modelling and personalisation, semantic web, social web, authoring, with over 300 papers on these subjects (over 4000 citations on Google Scholar, h-index 35). She was classified within the top 50 researchers in the world in the area of educational computer-based research according to Microsoft Research. Prof. Cristea has been highly active and has an influential role in international research projects. She has led various projects, has been keynote/invited speaker, organiser, co-organizer, panelist and program committee member of various conferences in her research field. She is a member of the editorial board of the IEEE Transactions on Learning Technologies, executive peer reviewer of the IEEE LTTF Education Technology and Society Journal, Associate Editor of Frontiers in Artificial Intelligence.

Ahmed Alamri is an assistant professor at Jeddah University's Department of Computer Science. He earned his PhD from Durham University in 2023, specialising in Educational Data Mining. Additionally, he holds a master's degree in Internet and Security and a bachelor's degree in Information and Communication Technology. Ahmed's keen interest lies in educational data mining and learning analytics, utilising machine learning to forecast students' future activities.

Mohammed Alshehri is a lecturer at Jeddah University's Business department. He's currently pursuing a Ph.D. in educational data mining and e-commerce, using machine learning to predict students' purchasing behaviour and boost MOOCs revenue. With a Master's in information system management and a Bachelor's in Management Information System.

Filipe Dwan Pereira received the B.S. degree in Computer Science from Federal University of Roraima and the M.S. degree in Computer Science from Federal University of Amazonas, where he is currently pursuing the Ph.D. degree in Artificial Intelligence applied to Education. Since 2013, he has been a Auxiliar Professor with the Department of Computer Science, Federal University of Roraima. His area of research covers education data mining, learning analytics, artificial intelligence, machine learning, big data, computing in education, and information systems.

Armando M. Toda is a postdoctoral research associate at Durham University. Former lecturer at University of Sao Paulo. His research interests are: Computers in Education, Gamification, Human-Computer Interaction, Artificial Intelligence in Education, Software Engineering, Educational Data Mining, and Computational Thinking.

Elaine Harada T. de Oliveira holds a diploma degree in Computer Science from the University of São Paulo, Brazil. She finished her PhD at the Graduate Program in Informatics in Education from the Federal University of Rio Grande do Sul, Brazil, in 2011. Since 2002, she has been a professor at the Institute of Computing at the Federal University of Amazonas. She is an associate editor of the Brazilian Journal of Computers in Education (2019–2021). Her present research is focused on studying students' behaviour as they learn how to program, using learning paths and data-driven approaches. The data is collected by the interaction of the students with a self-devised Online Judge and Learning Management Systems. The goals of her research are to predict outcomes, help decision making and provide adaptive learning through Learning Analytics.

Craig Stewart has worked in the areas of HCI, Serious Games, Personalised Systems, Multimedia and Cultural Studies in research and education for over 25 years. Dr Stewart's doctoral research (entitled A Cultural Education Model: Design and Implementation of Adaptive Multimedia Interfaces in eLearning) consists of examining the effect that technology enhanced learning is having on cultural education and how HCI influences this.

Authors and Affiliations

**Alexandra I. Cristea[1] · Ahmed Alamri[5] · Mohammed Alshehri[6] ·
Filipe Dwan Pereira[2] · Armando M. Toda[3] · Elaine Harada T. de Oliveira[4] ·
Craig Stewart[1]**

Ahmed Alamri
asalamri4@uj.edu.sa

Mohammed Alshehri
malshehri@uj.edu.sa

Filipe Dwan Pereira
filipe.dwan@ufrr.br

Armando M. Toda
armando.toda@usp.br

Elaine Harada T. de Oliveira
elaine@icomp.ufam.edu.br

Craig Stewart
craig.d.stewart@durham.ac.uk

[1]   Department of Computer Science, Durham University, Lower Mountjoy, South Road,
       Durham DH1 3LE, UK

[2]   Federal University of Roraima, Boa Vista, Brazil

[3]   University of Sao Paulo, Sao Paulo, Brazil

[4]   Federal University of Amazonas, Manaus, Brazil

[5]   University of Jeddah, College of Computer Science and Engineering, Department of Information
       Systems and Technology, Jeddah, Saudi Arabia

[6]   University of Jeddah, College of Business, Department of Management Information Systems,
       Jeddah, Saudi Arabia