

RESEARCH

Open Access



Modeling students' behavioral engagement through different in-class behavior styles

Samuel Gomes^{1,2*} , Luis Costa³, Carlos Martinho^{1,2}, João Dias^{1,4,5}, Geraldo Xexéo³ and Ana Moura Santos²

Abstract

Background The growing necessity of providing better education, notably through the development of Adaptive Learning Systems (ALSs), leveraged the study of several psychological constructs to accurately characterize learners. A concept extensively studied in education is engagement, a multidimensional construct encompassing behavioral expression and motivational backgrounds. This metric can be used to not only guide certain pedagogic methodologies, but also to endow systems with the right tutoring techniques. As such, this article aims to inspire improved teaching styles and automatic learning systems, by experimentally verifying the influence of in-class behaviors in students' engagement.

Results Over 16 math lessons, the occurrence of students' and instructors' behaviors, alongside students' engagement estimates, were recorded using the COPUS observation protocol. After behavior-profiling the classes deploying such lessons, significant linear models were computed to relate the frequency of the students' or instructors' behaviors with the students' engagement at different in-class periods. The models revealed a positive relation of students' initial individual thinking and later group activity participation with their collective engagement, as well as a positive engagement relation with the later application of instructor's strategies such as giving feedback and moving through class, guiding on-going work.

Conclusions The results suggest the benefit of applying a workshop-like learning process, providing more individual explanations and feedback at the beginning of an interaction, leaving collective feedback and students' guidance of on-going work for later on. Based on the relations suggested by our models, several guidelines for developing ALSs are proposed, and a practical illustrative example is formulated.

Keywords Behavioral engagement, STEM Education, COPUS Protocol, Adaptive Learning Systems

Introduction

During the last years, research has identified the necessity of developing systems that are able to automatically assign or adapt learning content to the students' needs and style of learning (Bagheri, 2015; Hwang et al., 2013). Systems like these are particularly important given that students may not autonomously choose content based on their optimal way of learning, relying instead on external factors such as how their interaction was "interesting," "relaxing," "easy to use," and "conforming to previous experiences" (Hwang et al., 2013). As a response to this demand, the field of Intelligent Tutoring Systems, or ITSs (Bagheri 2015), contributed to the development of Adaptive Learning

*Correspondence:

Samuel Gomes
samuel.gomes@tecnico.ulisboa.pt

¹ INESC-ID, Rua Alves Redol, 9, 1000-029 Lisboa, Portugal

² Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais 1, 1049-001 Lisboa, Portugal

³ Universidade Federal do Rio de Janeiro, Rua Antônio Barros de Castro, 119, CEP: 21941-853 Cidade Universitária, Rio de Janeiro, Brazil

⁴ CCMAR, Campus de Gambelas, 8005-139 Faro, Portugal

⁵ Faculdade de Ciências e Tecnologia, Universidade do Algarve, Campus de Gambelas, 8005-139 Faro, Portugal

Systems (ALSs). ALSs can be defined as an evolution from traditional ITSs that, instead of acknowledging learners as passive receivers of information, attempt to tailor learning material and goals based on the learners' preference and knowledge. The adaptive learning and training field is ever-growing, notably when leveraged by the interactivity of serious games, now extremely widespread and researched in institutions around the globe (Susi et al., 2007). ALSs were inspired by the proliferation of online and digital platforms, which allow more detailed learner modeling and integration, e.g. by endowing learners with an easy and personalized access to learning content. For instance, while some learning systems may integrate learners by giving them customization choices, others may automatically build learner models by silently monitoring subjects' task completion or problem solving strategies. In fact, although research has defended that collective instruction may bring more challenges than one-to-one tutoring (Bloom 1984), recent work regarding the use of technology in education is considering to model learners at a group/team level as a path to aid collective learning (Bonner et al., 2016; Gomes et al., 2019), which reflects the continuous growth of research in learner modeling and integration.

As Murphy et al. review, there is value in further exploring how different teacher characteristics and teaching styles may support learner motivation, one of the constructs of engagement (Murphy et al., 2019). This type of exploration is particularly relevant for automatic adaptive systems, in the sense that they may be informed by such constructs in order to make decisions. Various possibilities have been considered for measuring engagement, from more objective metrics such as re-playability (Catarino & Martinho 2019; Sarkar et al., 2017), dialogue patterns (Novielli 2010), or as an individual's involvement in a task via the time spent on-task (Järvelä et al., 2008; Määttä et al., 2012; Seifert & Beck 1984), to more psychology-driven metrics such as the students' motivations for achieving, and how positive learners feel while working on tasks. Overall, engagement can be embraced as a complex, multidimensional concept, encompassing behavioral, emotional, cognitive, and social connotations (Fredricks et al., 2016), thus advances in the study of this human trait and how it influences education are crucial. In particular, we believe that the assessment of the more objective, behavioral facets of engagement, notably the observation of how the emergence of several instructors' and students' in-class behaviors can influence the engagement of students, may help to improve existent teaching styles and guide the development of more advanced ALSs capable of enhancing learning through the promotion of the right behavior and feedback at the right moments.

Keeping these thoughts in mind, we aim to approach the following research questions in the current paper:

RQ1: *How is students' engagement influenced by the emergence of different styles of behavior during a lesson?*

RQ2: *What are the practical implications of this influence for improving education methodologies and ALSs?*

In order to explore these questions, we observed students' behavioral engagement in two math subjects that take part in the first year of a computer science STEM (Science, Technology, Engineering, and Math) degree of one of our institutions. We selected these subjects and the STEM domain, given the relevance of mathematics for the development of students' competences in first engineering degrees (Flegg et al., 2012), and because STEM has extensively been the target of analysis of research approaching the use of students' and instructors' behaviors to profile learning (Lund et al., 2015; Smith et al., 2013, 2014; Stains et al., 2018; Tomkin et al., 2019). In specific, we measured the behavior of both students and instructors resorting to a protocol known as COPUS, the Classroom Observation Protocol for Undergraduate STEM (Smith et al., 2013). This protocol allows an observer to reliably characterize how faculty and students are spending their time in the classroom by encompassing a list of representative students' and instructors' in-class behaviors. After gathering data from multiple in-class behaviors, during a whole semester, we managed to construct significant linear models relating the frequency of some students' and instructors' behaviors with the students' average engagement at different periods, thus providing some answers to our research questions.

Background

This work draws on two major concepts, explored in the background. Firstly, we look into how research defines and studies engagement, notably academic engagement, as well as the constructs associated to this complex psychological concept. Secondly, we investigate COPUS, in particular its premises and how it was applied to profile classes and evaluate the suitability of pedagogic methodologies.

Correlates of engagement

Although there is a lack of understanding to what constitutes engagement, this human characteristic has been associated with positive changes in skills and abilities, and greater psychological adjustment during college years (Wilson et al., 2015). As commented while introducing this work, various possibilities have been considered for measuring engagement, e.g. considering re-playability, dialogue patterns, or on-task time. However, engagement can also be connected to the students' motivations for achieving, and how positive they

felt while working on tasks. In fact, holistically speaking, psychology-driven research characterizes engagement as a multidimensional construct encompassing both behavioral and subjective definitions. Notably, research on academic engagement distinguishes between behavioral engagement—the positive conduct of students, their adherence to academic tasks and school-related activities; emotional engagement—the students’ affective reactions to academic experiences, measured in terms of positive or negative emotions; and cognitive engagement—connected to the investment in learning, self-regulation, being strategic, desiring to go beyond the requirements, and valuing challenge (Fredricks et al., 2004). More so, each dimension of engagement can vary qualitatively, for instance behavioral engagement may encompass doing an assigned task, or joining a student council; or motivation can have intrinsic and extrinsic motives, depending if it is influenced from internal constructs, such as a search for autonomy and competence through play, exploration, and curiosity, or external constructs such as incentives or pressure (Furrer & Skinner 2003; Ryan & Deci 2020). Other engagement classifications further expand engagement theory (Fredricks et al., 2016), notably social-behavioral engagement as a style of engagement accounting for the key role of social interaction in supporting learning in small groups, and that associates with affect (Linnenbrink-Garcia et al., 2011), agentic engagement relating to the students’ constructive contribution into the flow of the instruction they receive, e.g. by offering input, expressing a preference, or asking a question (Reeve & Tseng 2011), and volitional engagement connected to the enactment of intentions, separated from motivation, which relates to the building of intentions (Filsecker & Kerres 2014).

Despite harder to measure, intrinsic factors are usually considered as better predictors of positive engagement and learning performance than external factors. Thus, one limitation of this work is that it confines itself to the extrinsic nature of in-class behaviors, and how they are observed, not going deeper into the students’ subjective minds. Even so, in our opinion, the positive engagement relations observed along the emergence of some in-class behaviors may, at the same time, connect to the existence of intrinsic benefits, for instance, more active students’ behavioral engagement may be linked to a higher emotional predisposition to actually learn more about that subject. Extrinsic motivation in the form of introjected regulation, in which behavior is regulated by internal rewards of self-esteem for success and through avoidance of anxiety, shame, or guilt for failure (Ryan & Deci 2020), can help to justify this interpretation.

In the present work, we measured engagement through the appraisal of an observer who verified and recorded

the overall behavioral in-class engagement at specific periods (see Fig. 1). Indeed, the use of externally filled reports as indicators of learners’ engagement is not new (Rudolph et al., 2001). Even though self-reported assessments would allow the appraisal of intrinsic (e.g. emotional or cognitive) aspects of engagement, this evaluation method could easily disrupt the normal lesson flow, and would render difficult the appraisal of collective engagement due to the nonexistence of a criterion baseline. Besides, although one’s confidence about their ability in a domain (self-concept) or a given task (self-efficacy), and one’s attributed task value (the utility attributed as a justification of doing a task), seem to be needed for students to engage in STEM content and have career aspirations on these fields, they are not stable assessment traits, possibly influenced by demographic attributes such as gender (Murphy et al., 2019).

In addition, several works showed that it is crucial for faculty and administrators to promote a sense of relatedness in students, i.e., their need to feel that they are connected to their community (e.g. peers, teachers), feeling loved, cared for, and important, in order to develop their academic engagement, personal well-being, and self-determination (Deci & Ryan 2013; Murphy et al., 2019; Wang & Holcombe 2010; Wilson et al., 2015). For instance, it is recognized that the sense of relatedness between students and instructors may influence academic engagement, achievement, school liking, and school belonging (Furrer and Skinner 2003; King 2015;

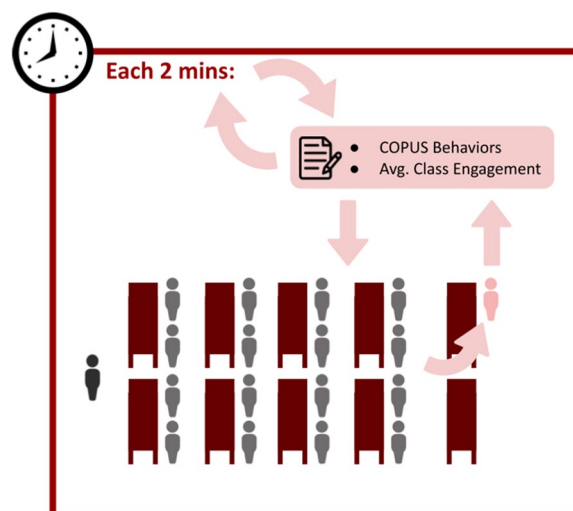


Fig. 1 Schematic representation of our data collection procedure. An observer sat discretely behind the class so that the data collection process was not disruptive, and at the same time all of the students were assessed while performing the observations. At each 2 min of the lesson, the observer estimated, via an observation form, which COPUS behaviors were present, as well as what level of engagement the class presented

O'Connor & McCartney 2007; Wang et al., 2018). In fact, the theory of self-determination defends the existence of three fundamental, interconnected, needs that are required to achieve psychological well-being—autonomy, competence, and relatedness; thus inciting extensive research on how these needs are met from the point of view of students and instructors (Ryan & Deci 2020). In our opinion, a sense of relatedness can be generally fulfilled by some instructors' in-class conduct, such as kindly clarifying doubts by answering a question to the whole class, or even more by discussing a certain topic with an individual student or a small group. Autonomy is usually enforced at the same time, while instructors attempt to understand, acknowledge, and be responsive to students' perspectives and unique challenges; and competence may work together with the other two characteristics whenever a good academic environment is met, i.e. that renders adequate challenges and feedback, and opportunities for growth (Ryan & Deci 2020).

In fact, the relations between the belonging to either class, academic major, or university communities, and engagement were already tested across STEM classes in five education institutions (Wilson et al., 2015), revealing that class belonging most consistently linked to positive engagement, contrasting to major belonging that only linked to engagement for some of the schools, and university belonging, which was the least consistently linked trait. Specifically, greater class belonging was linked to higher levels of reported student attendance across all five schools, and less negative emotional engagement at four of the five schools. In our opinion, the fact that the perceived sense of belonging to a class community more consistently related to positive engagement than the sense of belonging to the community of a particular academic major or university, renders valuable the observation of the students' engagement and behavior via an in-class setting.

COPUS as a behavior profiling tool

This work aims to relate the emergence of several in-class behaviors with observable students' engagement changes, particularly in STEM (math) classes. As such, we applied the COPUS methodology (Smith et al., 2013), which encompasses a set of attributes to characterize the behavior of both students and instructors during undergraduate STEM classes (see Table 1). Because of its discrete and extensive behavior categorization, we valued COPUS in relation to broader teaching practices indicators (Wiemann & Gilbert 2014). From now on, we will treat COPUS students' and instructors' categories of behavior as SB and IB.

The protocol was developed with several objectives in mind: (i) to characterize the general state of teaching

and learning; (ii) to provide feedback to instructors who desire information about how they and their students are spending class time; and (iii) to identify faculty professional development needs. Over the years, it has proved itself as a tool to reliably obtain behavioral data from classroom observations (Lund et al., 2015; Stains et al., 2018; Tomkin et al., 2019), by which we deemed this protocol as suitable for the goals of the present study. A benefit that differentiates this observation protocol from others is that it is easy to learn and requires little observer training (2 h or less). Using this protocol, an observer collects data by evaluating, for each two minutes of a lesson, which students' and/or instructors' behaviors are present (see Fig. 1). This iterative data collection allows the analysis of in-class behaviors not only at the global level, but also to observe their evolution during a lesson.

Some recent research by Lund et al. and Stains et al. used the data collected from multiple class observations (Lund et al.: 73 faculty members and 269 individual class periods; Stains et al.: 548 faculty members and 2008 classes) to further cluster the complex array of COPUS behaviors into more simple categories, for an easier profiling of students' and instructors' behavior and ease of analysis (Lund et al., 2015; Stains et al., 2018). In specific, Lund et al. clustered the COPUS behaviors into four categories, reflecting different teaching styles—Lecturing, Socratic, Peer Instruction, and Collaborative Learning; and Stains et al. clustered the COPUS variables into three teaching style categories—Didactic, Interactive Lecture, and Student-Centered. As acknowledged by Tomkin et al., (2019), these categories have remarkably similar traits. Firstly, lecturing, socratic, and didactic lecturing styles involve large portions of instructor lecturing (at least 80%) and student work is negligible, averaging 10% or less. Secondly, peer instruction and interactive lecture display concordant instructor behaviors, i.e. lecturing, on average, between 55% and 76% of the periods in the former case and in roughly 75% of the periods in the latter, and students' group work (with student response systems or without) averaging between a quarter to a half of the observed periods. Finally, both the collaborative learning and the student-centered learning categories share low levels of lecturing (50% of periods or less) and high levels of student work (around 50%). In fact, Tomkin et al. managed to relate the application of active learning with instructors' participation in communities of practice (Scaled Agile Framework 2021), by using a collapsed categorization of the COPUS behaviors (Smith et al., 2014) to discriminate between active and passive learning practices. In summary, the COPUS SB code L was labeled as passive, and CG, WG, Prd, WC, and SP were labeled as active; and IB codes Lec, D/V were labeled as passive, and FUp, PQ, and CQ were labeled as active. We believe that

Table 1 A list of all COPUS attributes. Adapted from: (Smith et al., (2013); Tomkin et al., (2019))

	Code	Description	Collapsed categories	Active(A)/ Passive(P)	
Students are doing...	L	Listening to instructor/taking notes, etc.	Receiving (R)	P	
	AnQ	Student answering a question posed by the instructor with the rest of class listening	Talking to class (STC)	-	
	SQ	Student asks question		-	
	WC	Engaged in whole class discussion by offering explanations, opinion, judgment, etc. to whole class, often facilitated by instructor		A	
	SP	Presentation by student(s)		A	
	Ind	Individual thinking/problem solving. Only mark when an instructor explicitly asks students to think about a clicker question or another question/problem on their own	Working (SW)	-	
	CG	Discuss clicker question in groups of 2 or more students		A	
	WG	Working in groups on worksheet activity		A	
	OG	Other assigned group activity, such as responding to instructor question		-	
	Prd	Making a prediction about the outcome of a demo or experiment		A	
	TQ	Test or quiz		-	
	W	Waiting (instructor late, working on fixing AV problems, instructor otherwise occupied)	Other (OS)	-	
	O	Other-explain in comments		-	
	Instructors are doing...	Lec	Lecturing (presenting content, deriving mathematical results, presenting a problem solution, etc.)	Presenting (P)	P
		RtW	Real-time writing on board, doc. projector, etc. (often checked along with Lec)		-
D/V		Showing or conducting a demo, experiment, simulation, video, or animation		P	
FUp		Follow-up/feedback on clicker question or activity to entire class	Guiding (G)	A	
PQ		Posing a non-clicker question to students (non-rhetorical)		A	
CQ		Asking a clicker question (mark the entire time the instructor is using a clicker question, not just when first asked)		A	
AnQ		Listening to and answering student questions with entire class listening		-	
MG		Moving through class guiding ongoing student work during active learning task		-	
1o1		One-on-one extended discussion with one or a few individuals, not paying attention to the rest of the class (can occur with AnQ, but should not occur with MG)		-	
Adm		Administration (assign homework, return tests, etc.)	Administration (A)	-	
W		Waiting when there is an opportunity for an instructor to be interacting with or observing/listening to student or group activities and the instructor is not doing so	Other (OI)	-	
O		Other-explain in comments		-	

this active/passive categorization (included in Table 1) can help us to more easily describe the origin of our data, notably the different education strategies employed in our observed classes.

Although much research was dedicated to the interpretation and validation of the COPUS protocol, we still have to consider that its present version captures some general actions of both instructors and students, and therefore it is not specific enough to judge *per se* the quality of those actions for enhancing learning, by which those assumptions have to be built upon the context where the observations were performed. For example, COPUS detects that an activity of clarifying nature was triggered, but it fails to make any further evaluation on

the formative quality of the behavior, that is, if the clarifying feedback was complete enough to clarify the doubts. In fact, McConnell and colleagues were unable to detect differences in the formative assessment practices enacted in class sessions that were categorized to be in the three different styles defined by Stains et al.—Didactic, Interactive Lecture, and Student-Centered (McConnell et al., 2021). We believe that this contextual need imposed while analyzing the COPUS variables may be crucial to understand if learning gains are achieved after applying certain in-class practices. Even so, as a first approach, we will try to maintain the descriptive nature of COPUS, generally interpreting the relation between engagement and the observation of in-class behaviors without

making any specific formative assessment about the content at hand. Nevertheless, future follow-ups of this work may apply a more granular criteria assessing formative constructs.

Measuring behavioral engagement

Alongside the behaviors being triggered, Smith et al. also proposes the measurement of engagement from a behavioral perspective, which can be applied in order to judge the effectiveness of different instructional activities (Smith et al., 2013). However, as reviewed before, engagement encompasses a wide range of constructs and its emergence is heavily context-related, thus its reliable annotation alongside in-class behaviors is difficult and must be approached systematically. Given this, prior to the in-class observations, several discussions were conducted between our observer and the instructors in order to classify students’ actions as indicative of behavioral engagement. We found it relevant to conduct these discussions as both our observer and the target instructors of this study have more than fifteen years of experience in the field of STEM education. Concordant with research on behavioral engagement measurement ((Lane & Harris 2015), see Table 2), we concluded that, in our case, examples of actions that reflected the presence of engagement were looking at notes, doing exercises, and listening (relates to “Listening” and “Writing”), brief interaction with colleagues/instructor after an in-class question (relates to “Engaged student interaction” and “Engaged interaction with instructor”), or even looking

at the phone for a few minutes after the teacher recommended a website or app (relates to “Engaged computer use”); and that, conversely, a lack of students’ engagement was reflected by irrelevant distractions such as frequent laughs, and music or game sounds (relates to all disengaged states except “Settling in/packing up”).

Considering such a classification, we then followed the proposed way to measure this attribute (Smith et al., 2013), notably by using a scale endowed with three levels—low, medium, and high; while assuming that there was low collective engagement whenever a small fraction of students (0–20%) was engaged; there was medium collective engagement whenever substantial fractions of students were engaged and substantial fractions of students were not engaged; and that there was high collective engagement whenever a large fraction of students ($\geq 80\%$) was engaged.

Method

Our experiments encompassed observational data from 16 lessons—13 lessons of the Linear Algebra: LA (Instituto Superior Técnico, 2019b), and 3 lessons of the Differential and Integral Calculus: DIC (Instituto Superior Técnico, 2019a) courses, conducted in Instituto Superior Técnico (2017). In fact, the STEM content of first-year LA has been considered to generalize well before, i.e. more likely to show greater similarities of learner objectives among schools than other disciplines (Seifert & Beck 1984). We assumed the same for DIC due to its structure and covered materials.

Table 2 Descriptions of students’ in-class behaviors that indicate they are engaged or disengaged. Adapted from: Lane and Harris (2015)

Engaged	Disengaged
Listening: Student is listening to the lecture. Eye contact is focused on the instructor or activity and the student makes appropriate facial expressions, gestures, and posture shifts (i.e. smiling, nodding in agreement, leaning forward)	Settling in/packing up: Student is unpacking, downloading class material, organizing notes, finding a seat, or packing up and leaving classroom
Writing: Student is taking notes on in-class material, the timing of which relates to the instructor’s presentation or statements	Unresponsive: Student is not responsive to the lecture. Eyes are closed or not focused on instructor or lecture material. Student is slouched or sleeping, and student’s facial expressions are unresponsive to instructor’s cues
Reading: Student is reading material related to the class. Eye contact is focused on and following the material presented in lecture or pre-printed notes. When a question is posed in-class, the student flips through their notes or textbook	Off-task: Student is working on homework or studying for another course, playing with phone, listening to music, or reading non-class-related material
Engaged computer use: Student is following along with the lecture on computer or taking class notes in a word processor or on the presentation. Screen content matches lecture content	Disengaged computer use: Student is surfing web, playing a game, chatting online, or checking the e-mail.
Engaged student interaction: Student discussion relates to class material. Student verbal and nonverbal behavior indicates he or she is listening or explaining lecture content. Student is using hand gestures or pointing at notes or the screen	Disengaged student interaction: Student discussion does not relate to class material
Engaged interaction with instructor: Student is asking or answering a question or participating in an in-class discussion	Distracted by another student: Student is observing other student(s) and is distracted by an off-task conversation or by another student’s computer or phone

At the time of the data collection, the LA program at Instituto Superior Técnico included mastery of matrix operations and factorization, methods for solving systems of linear equations, knowledge of vector spaces and linear transformations, eigenvectors and eigenvalues, and diagonalization techniques; and the DIC program included mastery of concepts of limits, derivatives, integral operations, and applications such as the convergence of sequences and series.

The data captured in our observations refers to 13 lessons from the same professor and 3 other lessons from two other professors. It is worth to note that the professor from which we observed 13 lessons co-authored this work. The monitored classes encompassed 5 theoretical classes (4 from LA and 1 from DIC) and 11 problem-solving classes (9 from LA and 2 from DIC).

Given that we considered the LA and DIC courses as our target of analysis, our sample included students from the first university year, who were aged around 18 years old. The same students attended both the LA and DIC classes, excluding a few students who were repeating one of the courses. As such, the sample covered roughly 100 students, where approximately 89% of the participants were male, and 11% were female. In addition, because there were multiple problem-solving shifts each week, the student attendance in theoretical classes (around 100 students) was higher than in problem-solving classes (around 30 students). Another difference is that the theoretical classes were by definition 50 min long and the problem-solving classes were 80 min long. At the start of the course, the students were informed about the experiment, and invited to sign a compulsory consent form, confirming their predisposition to participate in our data collection. Even so, since the analysis was made at a collective behavioral level, no personal or confidential information was extracted from the individual participants, and there were no potential risks and no anticipated benefits to each of them.

The students who participated in our experiment were monitored for each 2 min of each lesson, during 50 min, by an expert in the computer science field familiarized with the COPUS protocol (the first 50 min were considered for the case of the problem-solving classes). It is worth to note that the expert co-authored this work. This observer sat behind the students in a discrete way, so that data collection did not disrupt the normal lesson flow. As explained earlier, we relied on the observer's perception of different students' conduct considered to demonstrate the presence or lack of engagement, and used an engagement scale endowed with three levels—low, medium, and high, following the COPUS guidelines.

In total, over the 16 lessons, we collected 400 engagement datapoints and 1816 behavior datapoints (some

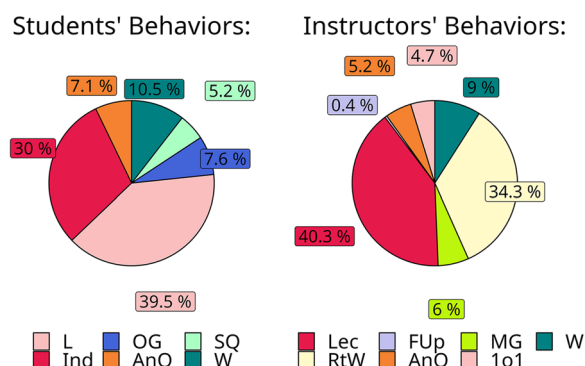


Fig. 2 Pie charts demonstrating the distribution of the COPUS behaviors over the observed theoretical classes

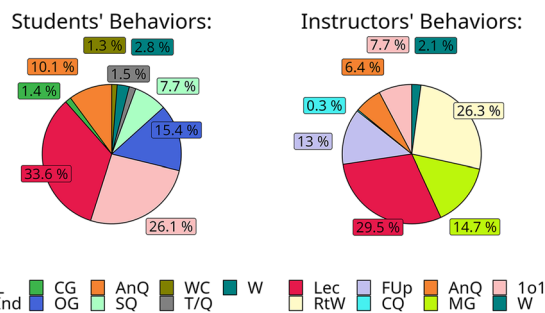


Fig. 3 Pie charts demonstrating the distribution of the COPUS behaviors over the observed problem-solving classes

timesteps encompassed more than one behavior, but only one engagement datapoint), from which 922 were SB and 894 were IB. 443: 210 (SB) + 233 (IB) of the behavior data were extracted from theoretical classes, and 1373: 712 (SB) + 661 (IB) were extracted from problem-solving classes. Even so, 15 of the 400 datapoints were left blank due to external factors such as the class starting late, and therefore were discarded from our analysis.

Figures 2 and 3 plot the distribution of the behavior datapoints collected over the theoretical and problem-solving classes. After analyzing the distributions, we can verify that a big part of the COPUS behaviors were recorded during our data collection. From all of the recorded behaviors, individual thinking and listening (Ind and L) were by far the most frequent SB in both class types, and conversely, CG, T/Q, and WC had few collected datapoints. Lecturing, real-time writing on board, and guidance while moving through class (Lec, RtW, and MG) were among the most frequent IB across the two class styles, and conversely, CQ had few collected datapoints. On top of that, we can also verify the nonexistence of data for WG, Prd, SP, O, PQ, D/V, and Adm in our observations. By analyzing the origin of our behavior data, we can verify that, as expected due to their nature,

a wider range of students' active behaviors was observed in the problem-solving classes (i.e. CG and WC) (Lund et al., 2015; Stains et al., 2018), although L and Ind dominated the students' behaviors in both class styles. Even more, relating the IB footprints with existing categorizations (Tomkin et al., 2019), the theoretical classes were lecture-based, but still slightly interactive and promoted some peer instruction (Lec and RtW constituted 74.6% of the observations, although FUp, MG, and 1o1 were present and occupied a considerable portion—11.1%), and the problem-solving classes were interactive and promoted peer instruction, leaning toward the collaborative and student-centered methodology (Lec and RtW constituted only 55.8% of the observations, and FUp and CQ were present and occupied 13.3%).

To ensure the likelihood of the collected data, we held a discussion about the COPUS data collected and our behavior profiling choices at a national mathematics congress. The participants present at this session, mostly statisticians, did not refute the coherence of our profiles in respect to the ones currently assumed for similar types of STEM classes.

Data cleaning and synthesis

After profiling our data, we prepared it for our main tests. Our final database grouped the data by each lesson, thus presenting 16 entries. Each entry encompassed the frequencies of each measured behavior in three class periods (an Initial period—from the start of the class until minute 16; a Medium period—from minute 18 to minute 32; and a Later period—from minute 34 to minute 50), by which these dimensions were named, for example, as *L_{Init}*, *OG_{Med}*, or *AnQ_{Later}*. For each lesson, the final database also registered the average engagement experienced by the students for each period: *eng_{Init}*, *eng_{Med}*, *eng_{Later}*, computed after attributing the values 1, 2, or 3 in case the engagement was coded as low, medium, or high by our observer at each 2 min time step, and averaging those values for that period. While grouping the data, we considered that no behaviors were observed in the 15 missing

datapoints, and that those cases did not contribute to the average engagement of the students. The histograms displayed in Figs. 4, 5, and 6 plot the absolute frequencies of the COPUS behaviors for each of the considered in-class periods. As previously suggested, Ind and L were by far the most frequent SBs, and more predominant towards the later period. The students and instructors waited (W) frequently at the initial period, yet the same did not happen during the other time frames. Besides, AnQ and OG had a considerable frequency, similar among the three periods, although AnQ was slightly more frequent at the medium period, and OG was slightly more frequent at the later period. As for the IBs, both RtW and Lec presented high frequency, yet they were less frequent at the beginning. MG and FUp were present over all periods with considerable frequency, yet MG prevailed in the initial and later periods, and FUp prevailed at the beginning. The non-mentioned behaviors presented low frequency overall.

All of the COPUS variables were assumed to be quantitative scales, which allowed us to compute linear regressions considering either the different behaviors of students or instructors as predictors of the students' average engagement at each of the three class periods. Instead, we could try to build models relating all behaviors simultaneously, but we chose not to follow that direction of analysis due to some conceptual correlations between students' and instructors' variables that could

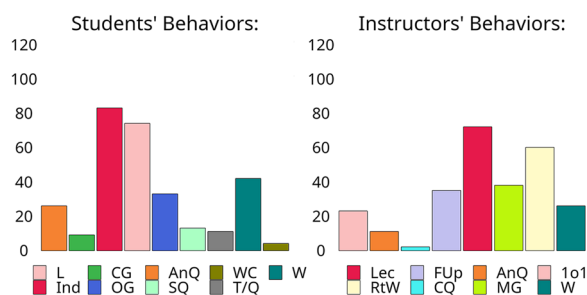


Fig. 4 Histograms of the COPUS behaviors over the *Init* period

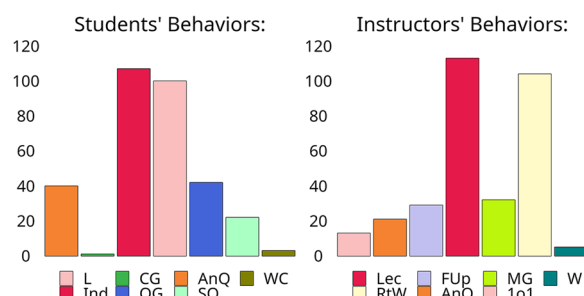


Fig. 5 Histograms of the COPUS behaviors over the *Med* period

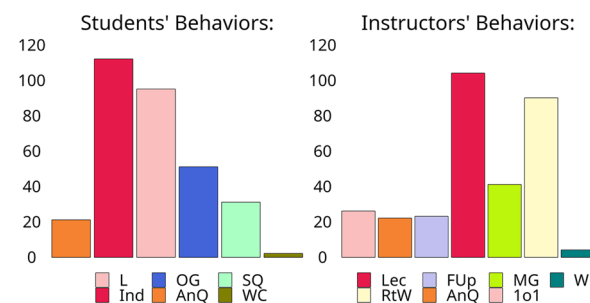


Fig. 6 Histograms of the COPUS behaviors over the *Later* period

ambiguate the statistical analysis and interpretation of the data. Further, although we could attempt to build separate models for theoretical and problem-solving classes, we assumed that said approach would render statistically weak models, given that the theoretical classes provided a more limited sample and behavior set (see Figs. 2 and 3). Consequently, six linear regressions were obtained resorting to a stepwise variable selection, which was examined and compared with forward and backward variable selection methods, meaning that after different combinations of behaviors were evaluated as predictors, only the most significant and scientifically explainable ones were selected for further scrutiny. All of the analyses resorted to the IBM SPSS Statistics, v. 26 software (IBM 2021), from which we used the default critical entry and removal F probabilities for variables selection of 0.05 and 0.1, respectively.

All linear model assumptions were assessed, notably the overall distribution of the data, and the normality, homogeneity, and independence of the errors. The first three assumptions were graphically validated, and the assumption of the independence of the errors was validated resorting to the Durbin–Watson statistic, as described by Marôco (Marôco 2021). With only one SB regressor per model (see next), the Durbin–Watson tests indicated non-autocorrelated errors for the SB regressions ($d = 1.968, 1.768, \text{ and } 1.562$, respectively). With one IB regressor in the first and last models, and two regressors in the second model (see next), the Durbin–Watson tests also indicated non-autocorrelated errors for the IB regressions ($d = 1.992, 1.822, \text{ and } 1.658$, respectively). Additionally, because the R^2 measure of a model is interpreted as the proportion of variance in the dependent variable that can be explained by the independent variables, we considered that $0 \geq R^2 > 0.33$ to be associated with low explanatory power, $0.33 \geq R^2 > 0.66$ to be associated with medium explanatory power, and $0.66 \geq R^2 \geq 1.0$ to be associated with high explanatory power.

Results

A first set of linear regressions was conducted to relate the students' behaviors with the average engagement values. The multiple linear regression regarding the *Init* period allowed us to identify initial individual thinking Ind_{Init} ($\beta = 0.838$; $t(14) = 5.749$; $p < 0.001$) as a significant predictor of eng_{Init} , defining the adjusted model: $\widehat{eng}_{Init} = 1.082 + 0.175 \times Ind_{Init}$. This model presents high explanatory power ($F(1,14)=33.050$; $p < 0.001$; $R_a^2=0.681$); the multiple linear regression regarding the *Med* period allowed us to identify other assigned group activities OG_{Med} ($\beta = 0.798$; $t(14) = 4.953$; $p < 0.001$) as a significant predictor of eng_{Med} , defining the adjusted model:

$\widehat{eng}_{Med} = 1.701 + 0.149 \times OG_{Med}$. This model presents medium explanatory power ($F(1,14)=24.536$; $p < 0.001$; $R_a^2=0.611$); the multiple linear regression regarding the *Later* period allowed us to identify other assigned group activities OG_{Later} ($\beta = 0.747$; $t(14) = 4.207$; $p < 0.01$) as a significant predictor of eng_{Later} , defining the adjusted model: $\widehat{eng}_{Later} = 1.412 + 0.126 \times OG_{Later}$. This model presents medium explanatory power ($F(1,14)=17.695$; $p < 0.01$; $R_a^2=0.527$).

A second set of linear regressions was conducted to relate the instructors' behaviors with the average engagement values. The multiple linear regression regarding the *Init* period allowed us to identify instructor waiting W_{Init} ($\beta = -0.746$; $t(14) = -4.189$; $p < 0.01$) as a significant predictor of eng_{Init} , defining the adjusted model: $\widehat{eng}_{Init} = 2.389 - 0.247 \times W_{Init}$. This model presents medium explanatory power ($F(1,14)=17.546$; $p < 0.01$; $R_a^2=0.525$); the multiple linear regression regarding the *Med* period allowed us to identify instructor follow-up/feedback FUp_{Med} ($\beta = 0.516$; $t(13) = 2.896$; $p < 0.05$) and guidance by moving through class MG_{Med} ($\beta = 0.472$; $t(13) = 2.647$; $p < 0.05$) as significant predictors of eng_{Med} , defining the adjusted model: $\widehat{eng}_{Med} = 1.559 + 0.130 \times FUp_{Med} + 0.150 \times MG_{Med}$. This model presents medium explanatory power ($F(2,13)=10.333$; $p < 0.01$; $R_a^2=0.554$); the multiple linear regression regarding the *Later* period allowed us to identify instructor guidance by moving through class MG_{Later} ($\beta = 0.770$; $t(14) = 4.510$; $p < 0.001$) as a significant predictor of eng_{Later} , defining the adjusted model: $\widehat{eng}_{Later} = 1.333 + 0.187 \times MG_{Later}$. This model presents medium explanatory power ($F(1,14)=20.337$; $p < 0.001$; $R_a^2=0.563$).

Discussion and practical considerations

To experimentally verify the influence of different in-class behaviors in the overall collective engagement of the students in a class, we observed 16 STEM (in particular math) lessons using the COPUS protocol, and selected linear models significantly relating the frequency of the students' or the instructors' behaviors with the students' average engagement at three class periods: *Init*, *Med*, and *Later*. The models highlighted five variables of interest (two SBs: Ind and OG; and three IBs: W, FUp, and MG). As previously analyzed, although these behaviors were observed across all styles of classes, FUp data was more frequent at the beginning, and Ind and OG were more frequent towards the final period. W prevailed at the beginning of the class, and MG data was frequent overall, with slightly lower values during the medium period. Nonetheless, all models encompassed frequently observed variables (Figs. 4, 5, and 6), and therefore were considered statistically relevant.

The obtained models allow us to better understand how student engagement can be influenced and/or predicted by the instructors' and students' behavior across different lesson periods. In the beginning of a lesson, the existence of individual thinking (Ind), in which a student individually responds to an instructor question, is positively correlated with engagement, and making students wait instead of interacting with them (W) is negatively correlated. During the medium stage of a lesson, students' collective-driven participation, e.g. when answering shared questions (OG), and the instructors' movement through the class guiding on-going student work (MG) together with follow-up discussion to students' questions and answers (FUp) are positively correlated with engagement. Finally, at the later period, student's shared participation (OG) and instructors' movement through class (MG) are each positively correlated with engagement, but there is no longer the need for follow-up discussions (FUp) to achieve engagement. In our opinion, these relations suggest that an ideal learning scenario may combine the benefits of theoretical and problem-solving activities, and assume the form of a workshop which stimulates more individual thinking at the start, and later makes use of group questions and feedback. Extrapolating this thought, students may be more self-centered and reflective at the beginning of an interaction extending beyond a single lesson (a long-term project, or a subject as a whole), while at the end they may value more the instructor's attention, guidance, and relatedness. However, care is needed to adapt this knowledge to broader time frames, as it requires validation through longitudinal studies that consider broader activities spanning through several classes.

Next, let us further interpret the applicability of our models in an ALS, providing some design guidelines useful for future reference. The fact that the models include both behaviors of students and instructors is important. Notably, the relations verified between behaviors and engagement are relevant because they inform designers of an ALS what are the best strategies to achieve engagement in tasks deployed across different points in time. Yet, some of the students' variables also depend on the instructor's behavior, for instance the existence of individual thinking as annotated by COPUS can only occur if the instructor asks individual questions to students, and the participation in group activities relies on the plural nature of the questions posed by the instructor.

Given this, the students' behavior models determined in the previous section suggest that to achieve/maintain high engagement at the early stages, the instructor should take some opportunities to ask questions and provide feedback to individual students, and afterwards

opt for collective questions and feedback, by which we can write the first ALS design guideline.

Design Guideline 1 (DG1)

To maintain engagement, an ALS should deploy individual-directed questions and feedback at early stages of the interaction, leaving group-directed questions and feedback for later on.

Upon further scrutiny, the models that show correlation between students' behavior and engagement can also be used as direct estimators of engagement. If after posing a question to a student (or group) there is no answer, this is a signal that engagement has not increased, and as such the ALS should try to mitigate the lack of engagement, e.g. by finding a new responder, or by rephrasing the original question. We can then register a second ALS design guideline.

Design Guideline 2 (DG2)

An ALS can use the presented student models to compute a rough estimation of engagement. This can be achieved by computing the Ind_{Init} , OG_{Med} , and OG_{Later} behavior frequencies, obtained by counting how many questions are answered by a single student or by the spokesperson of a group of students, respectively, and using those frequencies as in the students' linear models:

$$\widehat{eng}_{Init} = 1.082 + 0.175 \times Ind_{Init}$$

$$\widehat{eng}_{Med} = 1.701 + 0.149 \times OG_{Med}$$

$$\widehat{eng}_{Later} = 1.412 + 0.126 \times OG_{Later}$$

Furthermore, the instructor-directed models indicated that high engagement can be maintained by (expectedly) minimizing initial delays (W_{Init}), and by valuing discussions and follow-ups (FUp_{Med}) as well as students' guidance (MG_{Med} and MG_{Later}) further on. In specific, the models suggest that guidance of on-going work is needed through medium and later stages, although later stages disregard follow-ups as a measure to maintain engagement. Follow-ups can emerge naturally in discussions which require bi-directional interaction between students and the system, and students' guidance may be achieved by triggering automatic feedback prompts and/or hints about the students' performance or process (Bonner et al., 2016; Gabelica et al., 2012). This leads to the formulation of a third ALS design guideline.

Design Guideline 3 (DG3)

To maintain high engagement, it is important that an ALS is as proactive as possible (minimizes students' inactive periods) at early stages of the interaction, and that, after the early period of interaction, it stimulates follow-up discussion and students' guidance. In specific, at transitive moments, an ALS should consider both of these strategies, and at later stages it can exclusively provide guidance of on-going work, e.g. through feedback prompts and/or hints.

We will now further explore the application of guidelines DG1–DG3. Research on feedback provided to teams in higher education (Gabelica et al., 2012) covers topics such as the feedback level—whether feedback displays metrics of individuals (individual-level feedback) or the whole team (team-level feedback), and the setting in which the feedback is delivered—if the feedback is given to a team or provided to individuals in isolation. We consider that both individual-directed and group-directed feedback interventions can be deployed to guide the completion of a task by a collective, independently of its style, i.e. we can produce individual-level feedback for the participants of a group active task, or a discussion encompassing collective feedback with team metrics for the members of a task valuing individual completion. Based on recent research (Gomes et al., 2020; Alves et al., 2020), we speculate that individual-oriented tasks can (even in a group/class context) reward individual focus and target self-improvement, contrasting with group active tasks such as altruistic tasks rewarding full others-driven task completion, mutual-help tasks that reward all the task members equally, or even competitive tasks in which individual reward is provided to the detriment of the others' reward. In fact, this aspect is further approached by some serious game research, which considers gameplay-shaping elements, such as rewards and scores, as separate from interactivity-shaping game constructs such as feedback (Zea et al., 2009). We believe that, by applying DG1–DG3 to construct feedback interventions (which can be deployed as questions, discussions, and follow-ups), we can use an ALS to deploy tasks with different aims (e.g. an individual task targeting self-improvement, or a group task targeting cooperation), while at the same time taking into account the general knowledge extracted from our models to keep students engaged along the completion of those tasks. This leads us to propose a fourth design guideline.

Design Guideline 4 (DG4)

DG1–DG3 may be applied to construct ALS feedback interventions that can in turn be deployed to inform students during the completion of a task, independently of its style, for instance, individual-directed questions and feedback can be provided in tasks valuing group completion, and group-directed questions, discussions, and follow-ups can be prompted in tasks valuing individual completion.

Further, Gabelica et al. align their practical implications of feedback with our time-based analysis and premises, arguing that in educative contexts, teachers should continuously observe their students when engaged in a team task, allowing themselves to provide suited and timely feedback. One of the main benefits of applying our guidelines to an ALS is that it can be used to automatically intervene and deploy engagement-inducing feedback through

the course of a whole interaction, instead of the feedback being singularly given at the end of the interaction (as what usually happens in education for harder-to-respond in-class doubts or when an assignment only provides a final grade). Figure 7 illustrates the integration of the guidelines approached here in the operation of an ALS. From the beginning of the interaction until the last stage is finished, the ALS estimates the students' learning states, and provides them different styles of tasks (area A), to which it integrates feedback based on our engagement models (area B). More specifically, individual-directed questions are generated if the current moment is at the early stage of interaction (area B.1), group-directed discussions are generated if the current moment is at the medium stage (area B.2), and group-directed questions are generated if the current moment is at the later stage (area B.3). While formulating the feedback, the system verifies if engagement increased since the last intervention so that it decides to rephrase questions or to propose other discussion topics. To better materialize the benefits and concepts approached here, we will end our discussion by providing an illustrative example of applying an ALS endowed with this operation.

Illustrative example

This section presents a hypothetical use-case for our models—using an ALS to guide a 50-minute LA workshop class. Like in our observations, we assume that this ALS observes and updates its state every 2 min, and that the early stage of our interaction comprehends minutes 1 to 16, the medium stage comprehends minutes 18 to 32, and the later stage comprehends minutes 34 to 50. We will assume that this ALS models students' learning states and uses such information to propose learning materials (tasks), while giving students' feedback about their performance and/or work process (following the operation of Fig. 7). Now, let us exemplify the operation of our ALS, by considering how, at different moments, different levels of feedback can be integrated in the following LA exercise, adapted from Instituto Superior Técnico (2001) (for ease of comprehension, this assumes that the same task is chosen by the ALS at different periods).

Example of an LA Exercise

Calculate, if possible, $A+B$, $B+C$, $2A$, AB , BA , and CB :

$$A = \begin{bmatrix} 1 & 4 & \sqrt{2} \\ -2 & 1 & 3 \end{bmatrix}$$

$$B = \begin{bmatrix} 1 & 2 & \pi \\ \sqrt{3} & -1 & 2 \\ 0 & 1 & -1 \end{bmatrix}$$

$$C = \begin{bmatrix} 3 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

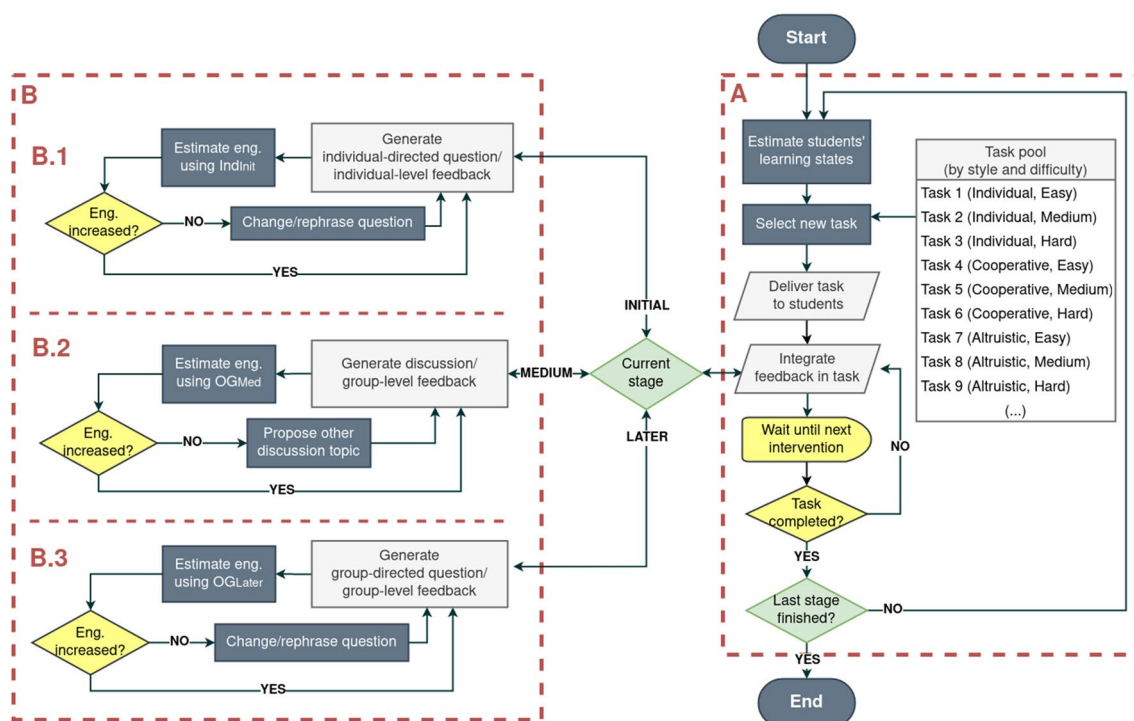


Fig. 7 Flow chart depicting the operation of an ALS following our guidelines. From the beginning of the interaction until the last stage is finished, the ALS measures the students' learning states, and provides them different styles of tasks (area A), to which it integrates feedback based on our engagement models (area B) according to the current adaptation stage (initial: area B.1; medium: area B.2; or later: area B.3). While formulating the feedback, the system verifies if engagement increased since the last intervention so that it decides to rephrase questions or to propose other discussion topics

Example of an LA Exercise

Solution:

Applying the matrix sum and multiplication rules, we get the following possible results (the others are not well defined):

$$B + C = \begin{bmatrix} 4 & 2 & \pi \\ \sqrt{3} & -3 & 2 \\ 0 & 1 & 4 \end{bmatrix}$$

$$2A = \begin{bmatrix} 2 & 8 & 2\sqrt{2} \\ -4 & 2 & 6 \end{bmatrix}$$

$$AB = \begin{bmatrix} 1 + 4\sqrt{3} & -2 + \sqrt{2} & \pi + 8 - \sqrt{2} \\ -2 + \sqrt{3} & -2 & -1 - 2\pi \end{bmatrix}$$

$$CB = \begin{bmatrix} 3 & 6 & 3\pi \\ -2\sqrt{3} & 2 & -4 \\ 0 & 5 & -5 \end{bmatrix}$$

During the early stage of our interaction (minutes 1 to 16), the ALS will prioritize individual-level questions and feedback (DG1), focusing on the frequency of students' individual interventions as an estimator of engagement (DG2). For instance, it can ask individual members of our simple task what formula they can apply to calculate $B+C$, or provide them with such a formula, even if the grade of that particular problem depends on the contribution of all group members. During this phase, it is

expected that the frequency and timeliness of each task is higher, minimizing waiting periods (DG3). Advancing to the medium stage (minutes 18 to 32), the system will use discussions (implying follow-ups) and team-level feedback (DG1 and DG3), for instance by opening a forum for justifying whether and how different matrices can be summed/multiplied, and it will estimate, whenever possible, the engagement of the class via the frequency of the students' collective-driven participation (DG2), e.g. when a student provides a valid input to the discussion such as presenting/explaining a formula, or contributes to solving a matrix like CB . For the ALS to follow and guide students, these group discussions may further be complemented with hints, e.g. the ALS showing how matrix multiplication is applied to obtain some elements of the CB matrix. In fact, team-level discussions can be used even if each student's grade for that problem depends only on their individual performance, or if the discussions are shown in each students' display instead of a shared display. Advancing to the later stage (minutes 34 to 50), the system will promote more direct group feedback and guidance (no need for discussions implying follow-ups), e.g. through group questions such as whether $A+B$ can be solved, and activity-completion hints such as

showing some elements of the AB matrix. When the lesson reaches its end at the 50-minute mark, the ALS finishes its last stage and stops the adaptation process.

Conclusion

The goal of this study is to experimentally verify the influence of different in-class behaviors in students' engagement, hoping to inspire improved education methodologies and Adaptive Learning Systems. In order to approach such question, we collected students' and instructors' behavior data, as well as students' engagement data from 16 different STEM (math) lessons, using the COPUS observation protocol.

After analyzing the behavior footprints of our lessons, we selected several linear models significantly relating the initial, medium, and later in-class frequency of either students' or instructors' behaviors with students' average engagement. The models revealed benefits of initial students' individual thinking, and later group participation, as well as the guidance of instructors at later periods, suggesting the benefit of applying a workshop-like education methodology deployed with such a structure. Given these trends, we provided not only some guidelines on how to apply this knowledge in an ALS, but also materialized these guidelines through a more objective automatic process endowed with the proposed pedagogic methodology. In specific, based on our results, we suggested that, at an early stage, an ALS may prioritize the deployment of individual questions and feedback, and focus on the frequency of students' individual input as an estimator of engagement. Advancing to the medium stage, the system can deploy follow-up discussions and collective feedback, and estimate, whenever possible, the engagement of the class via the frequency of the students' answers in such discussions. At the later stage, the system may ask group questions, and use students' participation on those questions to estimate class engagement. Additionally, we argued how an ALS can consider both individual-directed and group-directed feedback interventions in tasks with different settings, i.e. individual-directed questions and feedback can be provided in tasks valuing group completion, and group-directed questions, discussions, and follow-ups can be prompted in tasks valuing individual completion.

Limitations and future work

Even with interesting results, some limitations of this work have yet to be pointed. Firstly, our models were obtained using the observation of math classes, which narrows the scope of our results to this STEM area. Thus, the generalization of the ideas here expressed to other areas requires observations from a more varied subject

set. Besides, opposed to more general COPUS research (Smith et al., 2014), only one observer collected the data, which may have somewhat biased data collection. Also, the observation of only the first 50 min of each class may have inhibited some relations, mainly during problem-solving classes which were by definition 80 min long. The possible existence of the same relations in a more sparse number of intervals or a longer-than-a-lesson project could probably allow the computation of more accurate models. Another aspect which possibly influenced our models was the fact that our classes were not endowed with the same number of samples, i.e. while our theoretical classes presented an attendance of around 100 students, only around 30 students attended each problem-solving class. Besides, there were considerably more male than female students. Given this, in the future, we may re-iterate our tests with longer and more attendance-consistent and gender-consistent observations. The fact that the observations relied predominantly on classes from one instructor (13 lessons from the same professor and 3 other lessons from two other professors) may have also impacted the quality of our data collection, although this limitation was rendered non-critical due to the fact that the data analysis was done at the level of students' and instructors' behaviors, i.e. even though we only have 3 lessons from other professors, we still registered the emergence of multiple behaviors in those cases, as a consequence of different class circumstances. Overall, as the sparsity of the data sample in terms of demography and institutions is a recurrent concern in education research (Murphy et al., 2019; Wilson et al., 2015), longitudinal studies encompassing several school years, more subjects and instructors, or more education institutions, may be developed to further validate or disprove the models presented and discussed here. Ultimately, we hope that the application of these models to an ALS may unleash the practical value of our proposal, and that our analysis opens an exploration path for the deployment of more sophisticated learning processes.

Abbreviations

ALS	Adaptive Learning System
ITS	Intelligent Tutoring System
STEM	Science, Technology, Engineering, and Math
COPUS	Classroom Observation Protocol for Undergraduate STEM
LA	Linear Algebra
DIC	Differential and Integral Calculus

Acknowledgements

We want to thank all of the professors who taught the classes that we analyzed, and that supported our ideas. We also want to thank all the external people who gave us feedback, notably when we discussed our ideas at the presented national mathematics congress. Finally, we especially want to thank Marta Couto who helped us to plan our statistical analysis, and Prof. Warren for his availability and interest in discussing some aspects related to the COPUS protocol.

Author Contributions

SG contributed to the formulation of the research plan, analyzed the data, made the figures, and contributed to the writing and editing of the manuscript. LC contributed to the formulation of the research plan, performed the course observations, analyzed the data, and contributed to the writing and editing of the manuscript. JD analyzed the data, and contributed to the writing and editing of the manuscript. CM contributed to the writing and editing of the manuscript, GX contributed to the writing and editing of the manuscript, and AMS contributed to the writing and editing of the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with references SFRH/BD/143460/2019 and UIDB/50021/2020. The work was also supported by funds from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)—Brasil under Grant Finance Code 001.

Availability of data and materials

The dataset and outputs supporting the conclusions of this article are available in the *Modeling Students' Behavioral Engagement Through Different In-class Behavior Styles* repository, <https://osf.io/you2xw/>.

Declarations

Ethics approval and consent to participate

At the start of the course, the students and instructors were informed about the experiment, and invited to sign a compulsory consent form, confirming their predisposition to participate in our data collection. Even so, as the analysis was made at a collective behavioral level, no personal or confidential information was extracted from the individual participants, and there were no potential risks and no anticipated benefits to each of them. In addition, as this work comprehends an analytical study on the observational in-class data relating to the learning process, and in which the data was collected by one of the paper authors, we followed the self-assessment check-list proposed by the EU [1], that reflects that there is no need of further ethical approval in cases such as this.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 11 April 2022 Accepted: 10 February 2023

Published online: 11 March 2023

References

- Alves, T., Gomes, S., Dias, J., & Martinho, C. (2020). The Influence of Reward on the Social Valence of Interactions. *IEEE Conference on Games (CoG), 2020*, 168–175. <https://doi.org/10.1109/CoG47356.2020.9231853>
- Bagheri, M. M. (2015). Intelligent and adaptive tutoring systems: How to integrate learners. *International Journal of Education, 7*(2), 1–16.
- Bloom, B. S. (1984). The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring. *Educational Researcher, 13*(6), 4–16. <https://doi.org/10.3102/0013189X013006004>
- Bonner, D., Gilbert, S., Dorneich, M. C., Winer, E., Sinatra, A. M., Slavina, A., MacAllister, A., & Holub, J. (2016). The challenges of building intelligent tutoring systems for Teams. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 60*(1), 1981–1985. <https://doi.org/10.1177/1541931213601451>
- Catarino, J., & Martinho, C. (2019). Procedural Progression Model for Smash Time. *IEEE Conference on Games (CoG), 2019*, 1–8. <https://doi.org/10.1109/CIG.2019.8847964>
- Deci, E. L., & Ryan, R. M. (2013). *Intrinsic motivation and self-determination in human behavior*. New York: Springer.
- Filsecker, M., & Kerres, M. (2014). Engagement as a Volitional Construct: A Framework for Evidence-Based Research on Educational Games. *Simulation & Gaming, 45*(4–5), 450–470. <https://doi.org/10.1177/1046878114553569>
- Flegg, J., Mallet, D., & Lupton, M. (2012). Students' perceptions of the relevance of mathematics in engineering. *International Journal of Mathematical Education in Science and Technology, 43*(6), 717–732. <https://doi.org/10.1080/0020739X.2011.644333>
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*(1), 59–109. <https://doi.org/10.3102/00346543074001059>
- Fredricks, J. A., Filsecker, M., & Lawson, M. A. (2016). Student engagement, context, and adjustment: Addressing definitional, measurement, and methodological issues [Special Issue: Student engagement and learning: theoretical and methodological advances]. *Learning and Instruction, 43*, 1–4. <https://doi.org/10.1016/j.learninstruc.2016.02.002>
- Furrer, C., & Skinner, E. (2003). Sense of relatedness as a factor in children's academic engagement and performance. *Journal of Educational Psychology, 95*(1), 148.
- Gabelica, C., den Bossche, P. V., Segers, M., & Gijsselaers, W. (2012). Feedback, a powerful lever in teams: A review. *Educational Research Review, 7*(2), 123–144. <https://doi.org/10.1016/j.edurev.2011.11.003>
- Gomes, S., Alves, T., Dias, J., & Martinho, C. (2020). Reward-Mediated Individual and Altruistic Behavior. In *Videojogos, the 12th international conference on videogame sciences and arts*.
- Gomes, S., Dias, J., & Martinho, C. (2019). GIMME: Group Interactions Manager for Multiplayer sERious games. *IEEE Conference on Games (CoG), 2019*, 1–8. <https://doi.org/10.1109/CIG.2019.8847962>
- Hwang, G.-J., Sung, H.-Y., Hung, C.-M., & Huang, I. (2013). A Learning Style Perspective to Investigate the Necessity of Developing Adaptive Learning Systems. *Journal of Educational Technology & Society, 16*(2), 188–197.
- IBM. (2021). SPSS Software | IBM [Accessed: 2021-12-27]. <https://www.ibm.com/analytics/spssstatistics-software>
- Instituto Superior Técnico. (2001). Lista2.dvi- Lista2.pdf [Accessed: 2023-03-16]. <https://www.math.tecnico.ulisboa.pt/amoura/AL/Lista2.pdf>
- Instituto Superior Técnico. (2017). Técnico Lisboa- Engenharia, Arquitetura, Ciência e Tecnologia [Accessed: 2023-01-16]. <https://tecnico.ulisboa.pt/en/>
- Instituto Superior Técnico. (2019a). Initial Page . Differential and Integral Calculus I [Accessed: 2023-11-16]. <https://fenix.tecnico.ulisboa.pt/disciplinas/CDI54179577/2019-2020/1-semester>
- Instituto Superior Técnico. (2019b). Initial Page . Linear Algebra [Accessed: 2023-11-16]. <https://fenix.tecnico.ulisboa.pt/disciplinas/AL291795147/2019-2020/1-semester>
- Järvelä, S., Veermans, M., & Leinonen, P. (2008). Investigating student engagement in computer-supported inquiry: a process-oriented analysis [Publisher: Springer]. *Social Psychology of Education, 11*(3), 299.
- King, R. B. (2015). Sense of relatedness boosts engagement, achievement, and well-being: A latent growth model study. *Contemporary Educational Psychology, 42*, 26–38. <https://doi.org/10.1016/j.cedpsych.2015.04.002>
- Lane, E. S., & Harris, S. E. (2015). A New Tool for Measuring Student Behavioral Engagement in Large University Classes. *Journal of College Science Teaching, 44*(6), 83–91.
- Linnenbrink-Garcia, L., Rogat, T. K., & Koskey, K. L. (2011). Affect and engagement during small group instruction [Students' Emotions and Academic Engagement]. *Contemporary Educational Psychology, 36*(1), 13–24. <https://doi.org/10.1016/j.cedpsych.2010.09.001>
- Lund, T. J., Pilarz, M., Velasco, J. B., Chakraverty, D., Rosploch, K., Undersander, M., & Stains, M. (2015). The Best of Both Worlds: Building on the COPUS and RTP Observation Protocols to Easily and Reliably Measure Various Levels of Reformed Instructional Practice [PMID: 25976654]. *CBE-Life Sciences Education, 14*(2), 18. <https://doi.org/10.1187/cbe.14-10-0168>
- Määttä, E., Järvenoja, H., & Järvelä, S. (2012). Triggers of students' efficacious interaction in collaborative learning situations [Publisher: Sage Publications Sage CA: Los Angeles, CA]. *Small Group Research, 43*(4), 497–522.
- Maróco, J. (2021). Análise de equações estruturais: Fundamentos teóricos, software & aplicações (3). ReportNumber, Lda.
- McConnell, M., Boyer, J., Montplaisir, L. M., Arneson, J. B., Harding, R. L., Farlow, B., & Offerdahl, E. G. (2021). Interpret with Caution: COPUS Instructional Styles May Not Differ in Terms of Practices That Support Student Learning. *CBE—Life Sciences Education, 20* (2), ar26.

- Murphy, S., MacDonald, A., Wang, C. A., & Danaia, L. (2019). Towards an understanding of STEM engagement: A review of the literature on motivation and academic emotions. *Canadian Journal of Science, Mathematics and Technology Education*, 19(3), 304–320.
- Novielli, N. (2010). HMM modeling of user engagement in advice-giving dialogues [Publisher: Springer]. *Journal on Multimodal User Interfaces*, 3(1–2), 131–140.
- O'Connor, E., & McCartney, K. (2007). Examining Teacher-Child Relationships and Achievement as Part of an Ecological Model of Development. *American Educational Research Journal*, 44(2), 340–369. <https://doi.org/10.3102/0002831207302172>
- Reeve, J., & Tseng, C.-M. (2011). Agency as a fourth aspect of students' engagement during learning activities. *Contemporary Educational Psychology*, 36(4), 257–267. <https://doi.org/10.1016/j.cedpsych.2011.05.002>
- Rudolph, K. D., Lambert, S. F., Clark, A. G., & Kurlakowsky, K. D. (2001). Negotiating the Transition to Middle School: The Role of Self-Regulatory Processes. *Child Development*, 72(3), 929–946. <https://doi.org/10.1111/1467-8624.00325>
- Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology*, 61, 101860. <https://doi.org/10.1016/j.cedpsych.2020.101860>
- Sarkar, A., Williams, M., Deterding, S., & Cooper, S. (2017). Engagement Effects of Player Rating System-Based Matchmaking for Level Ordering in Human Computation Games. in *Proceedings of the 12th International Conference on the Foundations of Digital Games*. <https://doi.org/10.1145/3102071.3102093>
- Scaled Agile Framework. (2021). Communities of Practice-Scaled Agile Framework [Accessed: 2021-12-27]. <https://www.scaledagileframework.com/communities-of-practice/>
- Seifert, E. H., & Beck, J. J., Jr. (1984). Relationships between task time and learning gains in secondary schools [Publisher: Taylor & Francis]. *The Journal of Educational Research*, 78(1), 5–10.
- Smith, M. K., Jones, F. H. M., Gilbert, S. L., & Wieman, C. E. (2013). The Classroom Observation Protocol for Undergraduate STEM (COPUS): A New Instrument to Characterize University STEM Classroom Practices [eprint: doi: 10.1187/cbe.13-08-0154]. *CBE-Life Sciences Education*, 12 (4), 618–627. <https://doi.org/10.1187/cbe.13-08-0154>
- Smith, M. K., Vinson, E. L., Smith, J. A., Lewin, J. D., & Stetzer, M. R. (2014). A Campus-Wide Study of STEM Courses: New Perspectives on Teaching Practices and Perceptions [PMID: 25452485]. *CBE-Life Sciences Education*, 13 (4), 624–635. <https://doi.org/10.1187/cbe.14-06-0108>
- Stains, M., Harshman, J., Barker, M. K., Chasteen, S. V., Cole, R., DeChenne-Peters, S. E., Eagan, M., Esson, J. M., Knight, J. K., Laski, F. A., et al. (2018). Anatomy of STEM teaching in North American universities. *Science*, 359(6383), 1468–1470.
- Susi, T., Johannesson, M., & Backlund, P. (2007). Serious Games : An Overview (tech. rep. HS-IKI-TR-07-001). University of Skövde, School of Humanities and Informatics. Institutionen för kommunikation och information.
- Tomkin, J. H., Beilstein, S. O., Morphew, J. W., & Herman, G. L. (2019). Evidence that communities of practice are associated with active learning in large STEM lectures. *International Journal of STEM Education*, 6(1), 1–15.
- Wang, C., Harrison, L. J., McLeod, S., Walker, S., & Spilt, J. L. (2018). Can teacher-child relationships support human rights to freedom of opinion and expression, education and participation? [PMID: 29215309]. *International Journal of Speech-Language Pathology*, 20(1), 133–141. <https://doi.org/10.1080/17549507.2018.1408855>
- Wang, M.-T., & Holcombe, R. (2010). Adolescents' Perceptions of School Environment, Engagement, and Academic Achievement in Middle School. *American Educational Research Journal*, 47(3), 633–662. <https://doi.org/10.3102/0002831209361209>
- Wieman, C., & Gilbert, S. (2014). The Teaching Practices Inventory: A New Tool for Characterizing College and University Teaching in Mathematics and Science [PMID: 25185237]. *CBE-Life Sciences Education*, 13 (3), 552–569. <https://doi.org/10.1187/cbe.14-02-0023>
- Wilson, D., Jones, D., Bocell, F., Crawford, J., Kim, M. J., Veilleux, N., Floyd-Smith, T., Bates, R., & Plett, M. (2015). Belonging and academic engagement among undergraduate STEM students: A multi-institutional study. *Research in Higher Education*, 56(7), 750–776.
- Zea, N. P., Sánchez, J. L. G., Gutiérrez, F. L., Cabrera, M. J., & Paderewski, P. (2009). Design of educational multiplayer videogames: A vision from

collaborative learning [Designing, modelling and implementing interactive systems]. *Advances in Engineering Software*, 40(12), 1251–1260. <https://doi.org/10.1016/j.advengsoft.2009.01.023>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)