Utilizing Multimodal Data Through fsQCA to Explain Engagement in Adaptive Learning

Zacharoula Papamitsiou, Ilias O. Pappas, Kshitij Sharma, and Michail N. Giannakos

***Abstract*—Investigating and explaining the patterns of learners' engagement in adaptive learning conditions is a core issue towards improving the quality of personalized learning services. This study collects learner data from multiple sources during an adaptive learning activity, and employs a fuzzy set qualitative comparative analysis (fsQCA) approach to shed light to learners' engagement patterns, with respect to their learning performance. Specifically, this study measures and codes learners' engagement by fusing and compiling clickstreams (e.g., response time), physiological data (e.g., eye-tracking, EEG, electrodermal activity) and survey data (e.g., goal-orientation) to determine what configurations of those data explain when learners can attain high or medium/low learning performance. For the evaluation of the approach, an empirical study with 32 undergraduates was conducted. The analysis revealed six configurations that explain learners' high performance and three that explain learners' medium/low performance, driven by engagement measures coming from the multimodal data. Since fsQCA explains the outcome of interest itself, rather than its variance, these findings advance our understanding on the combined effect of the multiple indicators of engagement on learners' performance. Limitations and potential implications of the findings are also discussed.**

***Index Terms*—Adaptive learning, engagement, fsQCA, multimodal data, multimodal learning analytics, performance.**

I. Introduction

SUCCESSFUL learning and important educational outcomes such as persistence in learning, confidence, and academic achievement, have been linked to learners' engagement [1]–[3]. Interdisciplinary researchers in digital learning settings use the term "engagement" to refer to learners' effort and time investment, as well as persistence in learning [1], [4], and to synopsize learners' conscious, intrinsically motivated, and active involvement with the learning tasks [5]. Engagement has been widely conceptualized as a multidimensional construct, involving individuals' ability to implicate in ongoing learning processes; this ability is depicted through actual interaction between engagement objects and subjects [1], [4], [6]. A substantial number of factors has been associated with this term across studies, including participation; degree of interaction; commitment; response times; attention; goal-orientation; enjoyment; frustration; attitudes; depth of information processing stemming from self-regulated learning and more [7]–[9]. Researchers have classified those factors into three dimensions of engagement: thoughts (e.g., perceptions of self-regulation, motivation, goal-orientation) [10], [11], feelings (e.g., boredom, frustration, enjoyment) [9], [11], and behaviors (e.g., attention, effort, response time) [11], [12].

The overall benefits raising from learners' engagement in their own learning are widely acknowledged and include discouraging dropout, reducing procrastination, increasing attendance, improving self-regulation, fostering community, to name a few [2], [4], [13]. The success of the learning experience has been strongly associated with learners' engagement patterns observed during their interactions with the learning activities tasks [8], [13], [14]. Previous research has shown that learners' successful engagement in learning activities is primarily determined by their level of motivation and self-regulation [15]. Specifically, motivation is often seen in literature as an impetus for engagement in learning [16], whereas adaptive self-regulated learning is integral to learners' engagement [17]. Adaptive learning environments help learners initiate self-regulated learning processes in order to align with their own goal-orientation [18], which in turn can explain the reasons for students' engagement in a task [17]. In other words, adaptive learning settings are inherently supportive to learners' motivation; when learning occurs in such contexts, adaptation positively impacts learners' engagement [19], [20].

A. Adaptive Learning and Learner's Engagement: Motivating Engagement through Personalization and Feedback

Adaptivity and adaptive learning environments are in the epicentre of the digital learning research community. The 2018 NMC Horizon Report [21] highlighted the emergence of the systems that focus on timely providing the best possible tailored support to learners [22]. The support is mostly delivered as adaptive content/activities [23], as group or individualized recommendations [24], as analytics dashboards and open learner models [25], or by adjusting learning design to meet learners' abilities [26], to cater to personalized learning needs.

A variety of computer-based adaptive learning and assessment environments, such as intelligent tutoring systems and computerized adaptive tests, have been designed to detect, promote, and support learning engagement [13], [27]–[29]. In these settings, learners' engagement reflects how actively involved they are with the adaptive tasks [5], whereby active involvement results in significant benefits, such as improving problem-solving skills, increasing attendance and attention, improving self-regulation, guiding and facilitating autonomous learning decisions [13], [30], [31], to name a few.

What all aspects of engagement have in common when learning occurs in adaptive learning contexts is the underlying

Zacharoula Papamitsiou, Kshitij Sharma and Michail N. Giannakos are with the Norwegian University of Science and Technology, Trondheim,7034, Norway (email: zacharoula.papamitsiou@ntnu.no; kshitij.sharma@ntnu.no; michailg@ntnu.no).

Ilias O. Pappas, is with the University of Agder, Kristiansand, 4630, Norway (email: ilias.pappas@uia.no).

learners' *intrinsic motivation* which is *amplified and encouraged via adaptation*: in fact, the connection between intrinsic motivation and engagement is catalyzed by the personalization of the learning experience and the provided feedback.

In adaptive learning and assessment contexts, the activities are tailored to fit learners' needs and mastery levels; motivational (mastery or performance) goal-orientation is a critical factor in learners' responsiveness to adaptation mechanisms, feedback and scaffolding [32]. The goal of adaptive learning activities is to encourage learners to stay motivated, i.e., to feed learners' intrinsic desire to learn. By deriving suitable adaptation mechanisms, the learning process is controlled in a way that meets learners' motivation (by considering the learner models), whereby motivation is often considered an impetus for engagement in learning [16]. Indeed, motivation and theories of goal-orientation can help to explain the reasons for students' engagement in a task [17] and provides a theoretical framework to understand and enhance students' adaptive patterns of learning engagement [15]. Achievement goals are considered a facet of motivation given that they provide a purpose or focus for the task and, as such, influence students' learning behaviors, emotions, thoughts, and performance [33].

Furthermore, in adaptive learning and assessment contexts, students' engagement in learning is enhanced by the provided feedback when students are encouraged (e.g, to correct errors and to receive award marks [34]). Receiving regular feedback from the adaptive learning system has been acknowledged for motivating students to keep trying and remaining engaged in their learning to improve their performance [35]. Yet, students' responsiveness to feedback usually depends on whether they consider it to convey information about their personal growth, or if it scaffolds their self-regulatory learning processes and understanding of the content [32], since adaptive self-regulated learning is integral to learners' engagement [17].

B. Motivation of the Research and Research Question

The recent technological developments in high-frequency data collection has opened-up an unparalleled opportunity to unobtrusively understand how humans engage and learn with technology and to use these insights to design systems that amplify learning. Wearable sensors, gesture and facial sensing, eye-tracking, among others, can help us enhance the way we collect and make sense of rich user data [36] in diverse learning contexts. Leveraging such multi-sensor data is, therefore, expected to allow us to more accurately explain learners' engagement in adaptive learning contexts as well, and to inform adaptive learning design decisions accordingly.

We conceptualize engagement at the adaptive activity level using three dimensions: thoughts, feelings, and behaviors [1], [8], [11]. Thoughts are learners' perceptions about their motivation to be engaged with the activity in terms of goal-expectations; feelings are learners' intensity of emotions during the activity; behaviors are learners' efforts in terms of cognitive load, response times, and arousal, as they try to process and answer the adaptive tasks. Former studies in adaptive learning contexts explored mostly the behavioral (as observed actions/reactions) and affective aspects of engagement (e.g., [37]–[39]), or the relation between motivation and cognitive behaviors of engagement (e.g., [13], [32], [40]).

Overall, previous studies conducted in adaptive learning settings did not investigate learners' engagement in a holistic manner (e.g., [38], [40]), i.e., they did not cover all the three dimensions. Furthermore, the existing studies used variance-based approaches for hypotheses testing (e.g., [11], [12], [41]). Such methods provide a single, best-fit solution that explains the variance in a part of the sample (i.e., not the sample itself), but they do not allow for the identification of *multiple solutions* that jointly can explain larger portion of the sample, or uncover the exact interrelations between the considered variables.

Given the multidimensionality of engagement as a construct itself, this study adopts a holistic conceptualization of engagement (i.e., thoughts, feelings and behaviors), builds on previous research on engagement-related processes that employed activity logs and/or multimodal data (e.g., [42]–[44]), and introduces configurations to capture complexity, extend current insights and enhance the results with contextual information: we posit that there is not one unique, optimal, configuration of engagement dimensions. Instead, multiple and equally effective configurations of causal conditions exist, which may include different combinations of these dimensions: depending on how they combine, they may or may not explain students' high or medium/low performance in an adaptive activity. High performance refers to the presence of a condition, whereas medium/low to the absence of the condition. The absence is examined as the negation of a condition (i.e., not present), thus we examine the non-high, i.e., medium/low performance.

Configurational analysis focuses on the asymmetric relations among the examined variables and the outcome of interest, which may be achieved through the different configurations. The principle of *causal asymmetry* suggests that the causes explaining the presence of an outcome are likely to be different from those explaining the absence of the same outcome [45]. Furthermore, the principle of *equifinality* suggests that multiple complex configurations of the same conditions may explain the same outcome [46]. For example, different analytics representing students' activity (e.g., response time), self-regulation skills (e.g., time management), and personal interests (e.g., satisfaction from content) can explain learning outcomes only if they are examined in combination [29].

To this end, this study posits that there is a synergy among thoughts, feelings and behaviors and how they are reflected on students' performance, in adaptive learning contexts. Therefore, this study addresses the following research question:

RQ: *What configurations of learners' thoughts, feelings and behaviors lead to high or medium/low learning performance in adaptive learning contexts?*

C. Contribution and Novelty

To fully understand engagement-related processes in adaptive learning contexts and improve the quality of personalized learning services, learner engagement data need to be collected in multiple modalities. Fusing data from diverse sources and determining which combinations of the considered data can describe the multiple aspects of learners' on-task engagement, is still an open issue [47]. This study contributes as follows:

- *Theoretical Contribution:* it is the first study in an adaptive learning context—to the best of our knowledge—that utilizes and fuses learners' information-rich physiological data (in multiple modalities, yet not exhaustively) with clickstreams and self-reported perceptions to *capture and explain the full-complexity* of engagement as on-task behaviors, feelings and thoughts;
- *Methodological Contribution:* it employs fuzzy-set Qualitative Comparative Analysis (fsQCA) with multimodal data for the first time—to the best of our knowledge: fsQCA allows to *identify specific configurations* (i.e., combinations/interrelations of variables that best explain parts of the outcome) that can act as necessary or sufficient conditions in explaining the outcome [45].

The aim of this study is not to model the learners (or their behavior) in adaptive learning contexts or to build improved learner models to guide adaptation *per se*. The goal is to identify important interrelations amongst engagement-related variables that can be generalized to all adaptive learning systems. Thus, building on multimodal data capacity to holistically capture engagement, and determining which are these configurations, along with which ones are indispensable, which ones are not needed and which ones suffice for the outcome to occur, will advance our knowledge on why learners act the way they do in adaptive tasks, and how their on-task engagement reflects *both* their learning and actions, resulting to a more concise and rigorous interpretation of their performance.

II. Background and Related Work

This study is founded on three pillars: (a) the complex construct of learning engagement; (b) the utilized multimodal data to capture its complexity; and (c) the data configurations, using appropriate analysis methods, and in particular fsQCA. Accordingly, this section (a) elaborates on the gaps in related work on learning engagement (Sections II.A and II.B), and (b) provides the background for understanding the capacity of multimodal data to address the existing gaps (Section II.C) and the appropriateness of the data analysis method to identify the interrelationships in the data (Section II.D).

A. Measurement of Engagement in Digital Learning Settings

The multidimensionality in the definition of engagement and the inclusion of diverse factors for its description led to lack of clear measures of engagement, and thus, to encompassing measures that—most of the times—are incomparable across studies [4], [47]. To overcome this obstacle, previous work on the study of engagement employed well-established survey-based methods, and many publications explored possible quantitative scales of self-report measures [10]. In particular, driven by the multidimensionality of the concept and its relation with self-regulation and motivation, researchers adopted items from those close research areas (e.g., Motivated Strategies for Learning Questionnaire, Learning and Study Strategies Inventory) (e.g., [2], [48]). Other researchers developed and evaluated their own scales (e.g., [6], [10], [11]).

In brief, previous research used retrospective engagement measurement techniques that relied mostly on data collected by observers (e.g., taking notes, keeping running records, filming video) and self-reported perceptions, and were employed at the course level [10], [11], [41]. Employing such instruments, prior research has provided evidence regarding students' self-perceptions on a number of cognitive, metacognitive, affective, and motivational beliefs of engagement *across contexts* [9], [11]. Despite that such methods are easy to administer, they hinder our understanding on the deployment of engagement: they lack unobtrusiveness and the potential to capture the complexity of learners' continues involvement with the activities, failing to provide clues about what is needed for enriching instruction to motivate and engage learners [8], [47].

B. Analytical Approaches for Understanding Engagement

Recently, technology has afforded us with new methods to measure student engagement (at the course or the activity level), in ways that are scalable and minimally disruptive to learning [12], [47]. Specifically, the digital footprints left during interactions with the learning systems have been exploited for investigating engagement patterns. Activity-based approaches were employed to measure engagement in different contexts using heuristic features constructed from learners' activity logs [2], [49]. For the measurement of on-task behavior and active participation, the researchers used indicators tracked by the learning environment (i.e., learning analytics) such as number and frequency of postings, responses, and views; number of learning resources (e.g., videos) accessed, quizzes taken and exercises solved; time-spent and frequency of dashboard views; response time-based effort on solving tasks; number of tasks/assignments completed and many more (e.g., [2], [29], [35], [49]). The objectives in these studies were beyond creating valid and accurate measures of learner engagement at the course or activity level; understanding *when*, *how* and *why* students engage more efficiently and perform better [2], [29], gaining valuable insights about *how* students "react" to the design of learning system for making accurate inferences of student needs, and informing teaching and task design [35], and preventing disengagement and off-task behaviors [49], are among the issues to address.

However, two important issues remain open in those studies: (a) using data only from the learning environment (i.e., from the log files of clickstreams) can reveal only behavioral traits of engagement, lacking the potential to explain motivational or emotional aspects; and (b) if those approaches are employed alone, they might lack contextual information necessary to infer engagement processes. Thus, for capturing the full-complexity of engagement, the creation of fine-grained personalized experiences and the consideration of additional proxies prompt for more sophisticated initiatives to rethink measures of learning and engagement [21], [47].

C. Multimodal Data for Understanding Engagement

Interaction between learners and contemporary technologies offers an opportunity for collecting rich and multimodal data [50] using multiple data channels from various sources and in different modalities, i.e., multimodal learning analytics (MMLA). Findings from a study that utilized MMLA to predict learning constructs [5] suggest that although *individual modalities* can be a good proxy for performance and engagement, *fusing features from different modalities* has the potential to increase consistency and prediction accuracy. Overcoming the difficulties in gathering, fusing, analysing and making sense of MMLA has the capacity to provide rich information about learners' engagement states and behaviors [47], [51]–[53] and to offer novel means to enhance the learning experiences [36], [50]. Research in the study of learners' engagement has shown that this construct can be operationalized in more sophisticated ways, including heart rate [54], cognitive load [55], tracking of gaze [43], [52], or facial expressions [53].

Physiological measures can provide continuous information about participants' cognitive-affective states through their *arousal* levels, including states "outside of awareness," i.e., not directly observable [56]. For instance, it has been found that Heart Rate and Heart Rate Variability (as indicators of mental effort) are influenced by and can be used to detect mental stress [57]. Furthermore, Galvanic Skin Response sensors were utilized to measure electrical conductance of the skin, i.e., an indicator of arousal, in an attempt to associate learners' arousal with engagement, in a physical space and in a distributed learning environment [44]. The results reported a non-engaging learning experience: both the arousal measurement and the participants' self-reported perception of engagement were negative, indicating disengagement. The relation between engagement and arousal was explored also during collaborative learning activities in classroom settings [54]. EDA, i.e., a proxy for arousal, and responsiveness to stimuli were measured unobtrusively and accurately using a research-quality multi-sensor apparatus. It was found that attention and engagement are positively correlated to arousal, yet task difficulty increases both cognitive and emotional arousal. Utilizing unobtrusive measurements (i.e., physiological data and clickstreams) instead of self-reports to study engagement was also explored in [58]. The goal was to provide timely feedback for learning support from learners' current physiological states, in self-regulated learning settings, where learners are generally highly engaged with the learning activities [59]. However, the scarcity of available data did not allow for further processing of the biosignals. Finally, it was found that using only arousal can lead to results that are comparable to the best models for engagement [5], indicating that wristband data can provide accurate ubiquitous measurements and insights.

Furthermore, electroencephalography (EEG) variables were sensitive to disengagement due to *cognitive load* and the high level of working memory load on difficult tasks [42], [60]. It was also found that the participants exhibited avoidance behavior (i.e., withdrawing) by reducing levels of mental effort on those tasks [42], but as skills increased, the levels of workload did not decrease accordingly [60]. The approach in [60] seems promising because they removed the limitation of dealing with cables and allow collecting data from multiple participants at the same time in a natural setting. In addition, an EEG-based detector of cognitive load was developed to make inferences about student engagement using "peripheral" measures [55]. The sensitivity of this detector was tested in an experimental manipulation of instructional difficulty embedded in an adaptive system, and it was found that the cognitive load detectors were highly attuned to the cognitive load differences during the scaffolding phase of the adaptation.

Moreover, computer vision techniques have been employed to detect students' *emotional aspects* of engagement with the activities from students' face videos. For instance, real-time automated recognition of emotions from students' facial expressions while students played cognitive training games was explored in [53]. The results suggested that machine learning methods could be used to develop a real-time automatic engagement detector with comparable accuracy to that of human observers, without the need for self-reports and questionnaires [53]. Attentional and emotional involvement with a task was also explored using facial expressions recognition with computer vision and affective computing in activities with writing tasks [61]. In [62], the authors also focused on predicting the affective states of their participants, using face videos, wristband data and ECG. An automated recognition of fine-grained facial expressions that occur during computer-mediated tutoring revealed that upper face movements were predictive of engagement, frustration, and learning [63].

Finally, [64] used speech, posture and gaze data to automatically detect the moments when students' expectations are likely to influence their engagement with the knowledge ("epistemological frames"), aiming to understand the depth of students' *attention* and engagement with the content, but they could not verify a direct relationship between the behavioral patterns in the multimodal data and "epistemological frames." Gaze fixations were also utilized along with log-file data to investigate the student's attention in the areas of interest and the sequences of students' engagement in cognitive and metacognitive self-regulated learning processes during learning in an adaptive cognitive tutor [65]. Attentional states during students' interaction with adaptive tutoring systems were also detected using gaze data (i.e., number of fixations and fixation duration) aiming to build an attention-aware system that could react to students' attentional failure (e.g., mind wandering) [66]. Although eye-movements were utilized to measure student engagement at a microlevel, however, it was reported that measuring mind wandering and emotional arousal via eye-tracking was not yet a mature procedure [43].

Apparently, physiological data has great potential to enhance the deeper understanding of engagement and explain multiple aspects of it. However, most of the related studies have employed a single modality (or a few), or they explored a single dimension of engagement, with typically small samples. Yet, only few studies were conducted in adaptive learning contexts. Therefore, additional work, contextualized in adaptive learning, and considering multiple data modalities is required.

D. FsQCA in Educational Contexts

It is common practice in Technology Enhanced Learning (TEL) research to pose questions/hypotheses and to conclude with the acceptance or rejection of the initial assumptions. To analyse data and examine net effects between variables, many studies from the TEL domain use variance-based approaches, such as analysis of variance (ANOVA) or multiple regression analysis (MRA). Conceptually, each statistical test is appropriate for hypotheses testing, offering a single concrete assessment (e.g., mean, p-value, etc.) to explain the observed outcome. Such methods lack the ability to consider the complexity inherent in the various characteristics (or combinations of characteristics) of different sub-groups *within* a sample.

FsQCA is based on the notion of *configurations*, i.e., different combinations of learners' characteristics, describing different behaviors. The technique can identify multiple unique configurations that take into consideration small learner groups and may jointly explain a much larger portion of the sample, whilst in variance-based approaches the model explains only a portion of the sample [67]. This is an important methodological difference: fsQCA can help us identify how to design learning technologies *for all* [68], [69], unlike variance-based methods that test competing models to identify the fittest. Thus, this is a promising approach, especially for "*learning at-scale*" studies, where frequentist-based approaches (e.g., ANOVA, MRA) might ignore some thousands of learners that have different (from the average learner) needs and belong on the "tail" of the sample. FsQCA allows us to consider this group and identify an optimal solution addressing its needs.

Recently, fsQCA has been used to address challenges in the educational domain. In detail, studies aim to unravel causal factors that affect students' intention to undertake computer science studies [70]; to feed state-of-the-art learning analytics systems and facilitate research in this domain [68]; to explain learning performance using students' behavioral data (e.g., response time) and perceptions (e.g., self-regulation, satisfaction from content) [29]; to make sense of diverse learning phenomena happening simultaneously using learning analytics coming from datasets consisted of learners with different needs, learning strategies, backgrounds and so on [69], to name a few. Other recent promising seminal work on applying fsQCA with learning analytics can be found in [71], [72].

Utilizing learner-data analytics and configurational analysis may offer valuable insights in understanding users of contemporary learning systems, in diverse contexts [29], [68], [72].

III. Methods

A. Study Context and Design

This study is contextualized in adaptive self-assessment settings. As already stated in the introduction, adaptive learning settings are inherently supportive to learners' motivation and adaptation positively impacts learners' engagement [19], [20]. Yet, the primary purpose of engaging students in self-assessment is to boost learning and achievement [73]. In fact, self-assessment leads students to a greater awareness, by training them to self-regulate their motivation and behavior, as well as by fostering reflection on their own progress in knowledge or skills, and finally, to deeper engaging with their learning and to understanding themselves as learners [74].

Furthermore, the nature of the tasks to be carried out influences students' engagement with the tasks [75]. In this study, figuring out the correct solution and responding to a set of adaptive multiple-choice quiz questions constitutes the learning task. Previous results indicate that adaptive quizzes can amplify students' motivation and engagement, and students perceive that adaptive quizzes support their learning [35].

The students took the adaptive self-assessment quizzes *individually*, at a University lab, especially equipped and organized for the needs of the experimental process, for approx. 45 mins each student, on October 2018. The study was conducted as part of the Web Technologies course (related to front-end development), at a European University. In this course, the instructor typically employed gamified quizzes (Kahoots!) in the beginning of each lecture, to link the previous lecture with the current one and engage the classroom in the lecture. When the quizzes are gamified, the students can be motivated to improve their learning performance by engaging in competitions that are exciting and fun [76]. Shifting the focus, the adaptive self-assessment quizzes were introduced in the middle of the semester to assist students' *independent* learning, and were designed with a focus on facilitating students' self-preparation before the final exams, and helping them track their progress and align with their learning goals, by providing adaptive content and immediate feedback. The feedback about the correctness of the response was provided along with the option to show the correct answer to the questions that the students had submitted a wrong answer, to initiate students' self-reflection and self-evaluation processes, and to amplify their engagement. The scores that the students achieved on the self-assessment tests had no participation to the student's final course grade (i.e., no rewards as external motivation).

During the study design, the decision to utilize a set of multiple-choice questions was grounded on previous research that demonstrates that students who take practice tests often outperform students in non-testing learning conditions such as restudying, practice, or filler activities. A recent meta-analysis examined the effects of practice tests versus non-testing learning conditions, and the results revealed that practice tests are more beneficial for learning than restudying and all other comparison conditions [77]. It is ground truth that retrieval practice (i.e., calling information to mind rather than rereading it or hearing it, in order to trigger "an effort from within" to induce better retention) is better at reinforcing knowledge than restudying information, and that testing is a good way to activate this retrieval process, i.e., the so called "*testing effect*" [78], [79]. Research has provided evidence that multiple-choice testing had the power to stabilize access to marginal knowledge, highlighting how relatively simple it is to reactivate and consolidate knowledge [80], and at the same time, a growing number of studies on this topic have reported robust benefits of testing on transfer of learning [81].

B. Participants, System, and Procedure

Overall, thirty-two (32) undergraduate students (15 females [46.9%] and 17 males [53.1%], aged 18–21 years-old [M = 19.24, SD = 0.831]) were enrolled in the online adaptive self-assessment procedure. Students' self-assessment and interaction data were collected with a system that consists of (a) a testing interface, (b) an adaptation mechanism, (c) a tracker that logs the students' interaction data, and (d) a database storing information about the students and test-items [82].

The testing interface (Fig. 1) displays the test-items delivered to students one-by-one according to the adaptation mechanism that selects the *next most appropriate item* to deliver to the student, driven by the correctness of the student's response to the previous item and the discrimination ability of the items (the details are available in the Appendix).

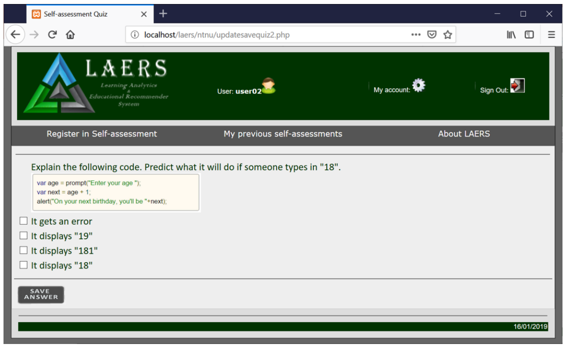

Fig. 1. The self-assessment interface.

Finally, the system also calculates and updates the test score for each student according to the correctness (0/1) of the student's answer on each item *i*. For the score computation, only the correct answers are considered, without penalizing the incorrect ones (i.e., no negative scores). Specifically, the incorrect answers participated inherently to formulating the "degree of difficulty" of the test, since the selection of the next item to deliver to students was guided by the correctness of the previous answer. Furthermore, due to the adaptive nature of the test, the students had to respond to and solve different number of items. Overall, a minimum of 10 and a maximum of 20 items were used to classify the students based on their diagnosed mastery level. To overcome these issues concerning the score computation, each student's *j* learning performance was calculated as: $\frac{\sum_{i=1}^{k} d_i z_i}{k}$, where *k* is the number of items and according to the correctness of the student's answer on each item *i*, with $z_i \in \{0, 1\}$ and the difficulty of the item, d_i. Each item had been previously weighted based on its difficulty level (see Appendix) and contributed differently to the overall score, ranging from 0.5 points (easy) to 1 point (medium) to 1.5 points (hard). The final score was in a 0–10 scale.

The participation to the procedure was optional. Prior to their participation, all students signed an informed consent form that explained to them the procedure and was giving the right to researchers to use the data collected for research purposes. After granting consent, the participants had to wear a wristband and an EEG cap and be connected to all the data collection devices (i.e., eye-tracker, wristband, EEG, cameras). Furthermore, the participants had to answer to a pre-test questionnaire that measures their goal-expectations. Then, the actual adaptive self-assessment test started and the students had to answer to the test items. Each item had two to four possible answers, but only one was the correct. In the end of the procedure, the test score was available to the students, along with their full-test results, including all the items they had answered to, their responses, the correctness of the responses, and the option to check the correct answer to the items that they had submitted wrong answers.

C. Data Collection and Measurements

The online self-assessment environment illustrated in Fig. 1 was employed to capture students' interactions with the adaptive learning system during the activity, according to the process described in the previous section. Based on the review of relevant literature, measurements commonly associated to the different aspects of learners' engagement, i.e., thoughts, feelings, and behaviors, were computed from the logged clickstreams (i.e., response time), as well as from the multiple physiological measurements collected with the respective devices (i.e., cognitive load, heart rate, attention, emotions) [43], [54], [62], [64]. Specifically, this study fuses multiple measurements computed from the EEG, eye-tracking, facial features, wristband data, and clickstreams. The measurements were as follows (the details can be found in Appendix):

a) response times: response times are by definition the time-spent the students constantly aggregate on answering the self-assessment items. In this study, they are distinguished according to the correctness of the submitted answer. In particular, total time to answer correctly (TTAC) and total time to answer wrongly (TTAW) are defined as the total time that students accumulatively spend on viewing the self-assessment items and answering correctly or wrongly respectively [84].

b) Arousal: For the calculation of arousal, four physiological measurements from the participants were computed using the wristband data, namely Heart Rate (HR), Blood Volume Pressure (BVP), Temperature, and EDA [54].

c) Cognitive load: Cognitive load was calculated from the EEG data-stream using band powers, i.e., frequency ranges that are strongly correlated with cognitive load [55].

d) Emotions: The facial data-stream was utilized to compute and model students' emotional intensity. Extracting emotions from the facial expressions is a common feature extraction technique [62]. Next, the average of the presence of high and low intensity emotions was computed.

e) Attention: eye-tracking data were utilized to compute attention. The most common practice to compute attention from students' gaze-patterns is to compute the average fixation duration during each sub-task in any give study [66].

Furthermore, since engagement has been found to be highly correlated with motivation (e.g., [4], [13], [32], [40]), for grading learners' motivation to take the self-assessment, we used goal-expectancy, i.e., a measure of self-preparation and goal-orientation particularized on Computer Based Assessment (CBA) procedures [84]. Goal-expectancy was measured via pre-test questionnaire in a 7 point Likert-like scale (1 = not at all to 7 = very much), by configuring the following items:

- *GE1*: I believe that I am well-prepared to take the CBA (i.e., I have read sufficiently for the test).
- *GE2*: I believe that attending the lectures helped me a lot to be prepared for the CBA.
- *GE3*: I believe that I can achieve high score in the CBA.

For all constructs, Cronbach's a was above 0.7, ensuring internal consistency, and GE was standardized in [0–1]. Table I summarizes all measurements considered in this study, along with a short description, engagement aspect the measurement is associated with in literature, type and value range.

D. Data Analysis

1) Introduction to fsQCA: To analyze the fused multimodal data, this study employs a configurational approach, namely fsQCA. The goal of fsQCA is to capture logically simplified statements describing different combinations (or configurations) of conditions and their interrelations that lead to a

TABLE I
MEASUREMENTS CONSIDERED IN THIS STUDY

Variable	Description	Engagement Aspect	Type	Value Range
Arousal	the state of physical vigilance or calmness	Behaviors [5], [44], [54], [58]	Composite—computed from wristband data	0—4
Cognitive Load	mental workload allocated on information processing	Behaviors [42], [55], [60], [83]	Composite—computed from EEG data	0—51
Emotional intensity	degree of externalization of affective state	Emotions [53], [61]–[63]	Composite—computed from facial expressions	0—1
Attention	on-task visual focus allocation	Behaviors [43], [64]–[66]	Composite—computed from gaze data	0—1
Total time to answer wrongly	response time aggregated on submitting wrong answers	Behaviors [7], [13], [29], [43]	Simple—computed from logged timestamped data	≥ 0 (msec)
Total time to answer correctly	response time aggregated on submitting correct answers			
Goal-expectancy	the motivation to take the self-assessment	Thoughts [4], [13], [32], [40]	Complex—computed from questionnaire data	0—1
Performance	the self-assessment test result	—	Computed from correct answers	1—10

specific outcome [85]. FsQCA views the variables as fuzzy-sets, allowing them to receive all values between 0 and 1.

In fsQCA the factors under examination are viewed as groups of interrelated structures, different from the typical variance-based approaches (e.g., ANOVA, MRA) that view them in isolation and in a competing environment while calculating net effects [45]. Specifically, the configurational approach of fsQCA can capture complex conditions existing in learning environments, and allows to go beyond commonly used variance-based methods, for two main reasons. First, it can lead to multiple solutions (i.e., patterns of factors or groups of factors) that explain the same outcome, compared to the single best solution of variance-based approaches. Second, it assigns a *truth value* to the data instead of a probability: once the analysis is complete, fsQCA allows for a follow-up "per case analysis"; the researcher can return to the cases, identify which ones are explained by each solution, and interpret each one using external information that are not included in the analysis (e.g., goal-orientation or other survey data), thus gaining a deeper understanding of the sample [85], [86].

To this end, the multiple solutions identified by fsQCA are combinations of variables in which they can act as either necessary or sufficient conditions for explaining the outcome. A variable in a solution might be present, absent (or negated), or not contributing at all at the solution (similarly with how variables behave in the real world). Also, a variable can have different roles in the different solutions, depending on how it combines with the other variables. FsQCA can answer the question as to which variables and their combinations are indispensable or sufficient for an outcome to occur, as well as which combinations are more (or less) important than others.

Furthermore, FsQCA can be used on very small to very large samples [86]. The sample size in this study (N = 32) is acceptable and supported by the method, which can work with samples of 12—50 cases. Specifically, it is suggested that "*Systematic comparison of causal connections across more than 10—12 cases becomes quite unwieldy without a tool for systematic comparison such as QCA; a deep, rich investigation [...] is still possible when examining 12—50 cases via QCA*" [86, p. 57]. FsQCA can also work well with different data types as long as each variable can be coded into a fuzzy set.

With fsQCA we can overcome limitations of frequentist-based approaches, complement them to better understand how the co-existence of different variables (e.g., multimodal data) can explain complex phenomena (e.g., learning experience).

2) Transforming data into fuzzy sets with fsQCA: First we transform the data into fuzzy sets, a process called *data calibration* in fsQCA and can be done within the fsQCA software. Instead of using probabilities (as in variance-based approaches), data are transformed into degrees of membership in the target set, indicating if and how much a case belongs into a specific set [45]. FsQCA computes the presence of a condition or its negation. For example, if for the variable "performance" we are interested in high values, then for the calibration process "high performance" means that the cases with high values on "performance" belong *fully* in the set. When a variable (or condition) is *fully* in the set it means that it is present, while when it is *not fully* in the set it means that it is not present (i.e., negation). The distinction highlights the asymmetric relations among the variables. For example, it is clear what values "high performance" includes, but its negation, i.e., "not high performance," includes everyone except those with "high performance"; "not high" may be translated as "medium/low" that is easier to interpret [87]; in symmetric relations the negation of high would be low.

In fuzzy sets, the 0–1 range indicates the degree that a variable belongs into the fuzzy-set, and all values need to be transformed in that rannge. A variable can be a full member of the set, a full non-member of a set, or in the middle (i.e., the intermediate set) being both a member of the fuzzy set and a non-member. In the 0–1 range, values 1, 0.5, and 0 indicate the full membership, intermediate membership, and full non-membership respectively. Calibration can be done either directly (i.e., the researcher chooses three qualitative thresholds to define membership levels in the fuzzy set) or indirectly (i.e., rescaling the measurements based on researcher's substantive knowledge of data and underlying theory) [85].

Choosing the three thresholds is a process that also involves the researcher's knowledge and there are different ways it can be done. First, the three thresholds can be 0.95 (full membership), 0.5 (cross-over point), and 0.05 (full non-membership) [45]. Percentiles can be used (i.e., 95%, 50%, 5%) to compute them. These percentiles can be changed accordingly if, for instance, the distribution of the dataset is skewed [70], [71]. In this study, the thresholds were chosen based on the 80%, 50%, 20%, and the calibration was performed using the dedicated function in fsQCA software which is based on the log odds of full membership, to fit the three breakpoints. Table II presents the thresholds for all engagement-related variables considered.

TABLE II
PERCENTILES CHOSEN FOR DATA CALIBRATION

Variable	**80%**	**50%**	**20%**
Arousal	0.428	0.355	0.297
Cognitive Load	21.71	18.78	12.80
Emotions	0.477	0.404	0.336
Attention	272.1	241.6	182.9
Total time to answer wrongly (TTAW)	246.6	180.5	103.2
Total time to answer correctly (TTAC)	357.2	214.7	99.10

3) Evaluation of the solutions and interpretations: Next, fsQCA creates a truth table of 2^k rows, where k represents the number of outcome predictors and each row represents each possible combination. The truth table is sorted based on frequency and consistency [45]. *Frequency* indicates the number of observations for every combination (i.e., how many cases in the sample are described by each combination). *Consistency* is the degree to which cases correspond to the set-theoretic relationships expressed in a solution [88], and indicates how strong a relationship is. A frequency cut-off point ensures that a minimum number of empirical observations is obtained. For small or medium samples a cut-off point of 1 is appropriate or 2 for larger samples (>150) [45]. All combinations smaller than this point are removed. Here, a frequency threshold of 1 is chosen (due to sample size N = 32). The consistency threshold is set at >.80, i.e., higher than the minimum recommended value of .75 [85]. Note that a lower consistency threshold may lead to identifying more necessary conditions, reducing type II errors (i.e., false negatives), but increasing type I errors (i.e., false positives) [89]. Combinations with consistency higher than the chosen threshold explain the outcome, while those with lower consistency do not explain the outcome.

FsQCA computes three solutions: complex, parsimonious, and intermediate. The complex presents all possible combinations of conditions, and can be very large and impractical to interpret. Thus, it is simplified to parsimonious and intermediate solution. The parsimonious is a version of the complex solution, presenting the most important conditions that cannot be left out from any solution, called “core conditions” [88]. The intermediate solution is part of the complex solution and includes the parsimonious ones. It is computed by employing counterfactual analysis on the complex and parsimonious solutions, including only theoretically plausible counterfactuals [45]. Conditions that appear in the intermediate solution but not in the parsimonious are called “peripheral conditions” [88]. The researcher may interpret either the parsimonious or the intermediate solution. The parsimonious focus on the core conditions (i.e., the most important ones), presented in the next section.

IV. RESULTS

A. Fuzzy Set Analysis of the Fused Multimodal Data

The findings from the fuzzy set analysis present combinations of the causal conditions that are sufficient in explaining learners’ high and medium/low performance (Table III). The solution presents only the core conditions. Black circles (●) represent the presence of a condition while crossed-out circles (⊗) its negation. A blank space means that a causal condition is not playing a role in the specific solution and may be either present or negated. Table III presents consistency values for each combination and for the overall solution, with all values being above the recommended threshold (>0.75). The overall solution coverage shows the extent that learners’ performance can be determined based on the identified configurations and is comparable to the R-square value. The overall solution coverage of .89 suggests that the solutions account for a substantial proportion of both high and medium/low performance.

For learners’ high performance, solutions S1–S6 present combinations for which the different factors can be present or absent, depending on how they combine with each other:

Solution 1 (S1): Students achieved high performance when they had high arousal, and did not spend a lot of time in the items that answered wrongly (low TTAW). In simple terms, when the body was in a state of “vigilance” (i.e., high heart rates and high blood volume pressure) instead of physical calmness or relaxation, the students could achieve high performance, if their response time to answer wrongly was low, i.e., they submitted wrong answers either rarely or quickly (which might indicate slipping a correct answer). The other physiological measurements (i.e., cognitive load, attention, emotion) did not play a role in this solution, which explains about half (53%) of the high-performers.

Solutions 2 and 3 (S2–S3): These solutions showcased that spending a lot of time to find the correct answers is important for a high performance, but not enough. This is an intuitive finding; it shows that students who give all their focus in correctly answering the items will achieve high performance. However, students who had studied sufficiently and submitted correct answers (i.e., high TTAC), also needed to remain either emotionally calm, i.e., preserve low intensity emotions (e.g., not excited) (*S2*), or physically calm, i.e., retain low arousal (*S3*), to achieve a high score. In this study, *S2* explains 48% of the high performers, whereas *S3* explains a slightly smaller sub-population of high-performers, since raw coverage is 43%. Cognitive load and attention did not play a role in *S2* and *S3*.

Solutions 4 and 5 (S4–S5): High performance was achieved when students exhibited high mental work and information processing (i.e., high cognitive load), and low intensity emotions (i.e., controlled their emotions), regardless of their response time. From that point on, students in *S4* (25% of the high performers) were physically calm (low arousal), whereas students in *S5* (24% of the high performers) had low attention.

Solution 6 (S6): The combination of high cognitive load and high attention with high response time to answer correctly led to high performance, explaining 37% of respective cases. In other words, the high performing students in this solution were mentally active and focused their attention on processing and solving the items, by allocating considerable amounts of time to understand them and submit the correct answers. Arousal and emotional intensity did not play a role in this solution.

The solutions for students’ medium/low performance (i.e., not high performance) are not the exact opposites of the ones explaining high performance (principle of causal asymmetry):

TABLE III
CONFIGURATIONS OF ENGAGEMENT-RELATED MULTIMODAL FACTORS (SOLUTIONS) FOR HIGH AND MEDIUM/LOW PERFORMANCE

	High Performance						Medium/Low Performance		
	S1	S2	S3	S4	S5	S6	S7	S8	S9
Arousal	●		⊗	⊗			●	⊗	⊗
Cognitive Load				●	●	●			
Emotion		⊗		⊗	⊗				●
Attention					⊗	●		●	
Total Time to Answer Wrongly (TTAW)	⊗						●		
Total Time to Answer Correctly (TTAC)		●	●			●		⊗	⊗
Raw coverage	0.53	0.48	0.43	0.25	0.24	0.37	0.56	0.27	0.38
Unique coverage	0.14	0.04	0.04	0.01	0.01	0.02	0.45	0.05	0.14
Consistency	0.91	0.82	0.95	0.91	0.85	0.94	0.86	0.83	0.94
Overall solution coverage	0.89						0.89		
Overall solution consistency	0.83						0.88		

Solution 7 (S7): Students who aggregated considerable response time to submit wrong answers had a medium/low performance, although their body physically reacted as it was experiencing high physical stimulus from the items, and they were not physically relaxed (i.e., high arousal). This behavior was observed in 56% of the medium/low performing students.

Solutions 8 and 9 (S8–S9): High emotional intensity (e.g., high astonishment, excitement, anger, frustration) or high attention led to medium/low performance when arousal and TTAC were on lower levels. Students who could not submit correct responses (low TTAC), even though they were physically calm (low arousal), either unsuccessfully tried to pay attention on the items (high attention) (*S8*), or they could not control their emotional reactions during answering the items (high emotional intensity) (*S9*). These solutions explain 27% and 38% of the medium/low performing students, respectively.

B. Enhancing the Solutions with Learners' Motivation

FsQCA allows the researchers to identify which specific cases in the sample are explained by each solution presented in Table III [86], by plotting each solution against the outcome. Producing plots in fsQCA is explained in detail in [87]. Each solution is computed separately to be used as input for the plot. Fig. 2 illustrates indicative plots for two solutions.

The Y–axis corresponds to the degree of membership of outcome (i.e., learners' high or medium/low performance) along with the consistency value for the specific solution. The solution (i.e., degree of membership of the causal recipe) is presented on the X–axis, along with the coverage, showing how many cases are explained by this solution. All values are within the 0–1 range. Each dot on the plot represents one case in the sample (i.e., one student). By clicking on a dot, the fsQCA software points to the case in the sample. The position of the dot in the plot allows us to understand how much a case belongs into the solution and also in the outcome. The higher the coverage and consistency are for a dot, the more the case belongs to the specific solution for the particular outcome.

In this study, the cases within the red rectangular shapes highlight students that have high values on the specific solutions and achieved high performance or medium/low performance (Fig. 2) (see supplementary material for all solutions).

Furthermore, some of the cases appear to more than one solutions. For example, case 13 appears to both *S1* and *S2* for high performing students. Similarly, case 11 is explained by *S2, S3, S4*, and *S5*. This means that for a specific student there is not always one single best solution; instead, multiple solutions exist that are a bit different but sufficient to lead to high performance. Also, the conditions are not competing with each other to offer the "best" solution, but instead, they are complementing each other in synergy into identifying the combinations that are sufficient to explain the outcome.

This also allows to better understand the blank spaces in the solutions, i.e., the causal conditions that *do not play a role* in specific solutions and may be either present or negated. A student may have high levels on multiple conditions, but *not all are necessary* to play a role for high (or medium/low) performance; only some of them, when combined, can be *sufficient* to explain the outcome. Thus, a student may appear in more than one solutions. This methodological characteristic of fsQCA will allow for making design decision at a later stage, according to the availability of learner-generated data.

After identifying the cases that can be explained by each solution (using the plots), we enhanced the solutions with

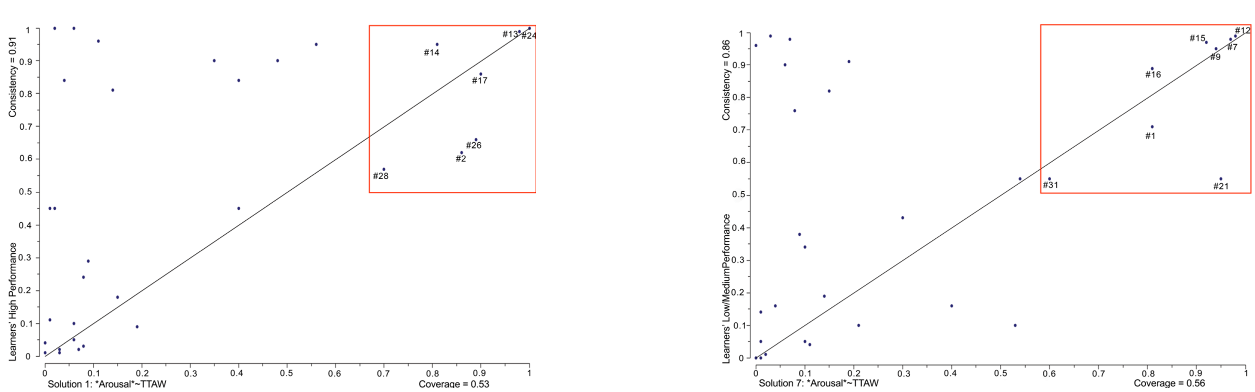

Fig. 2. Indicative solutions for high performance (left) and medium/low performance (right).

the self-reported goal-expectation, i.e., a measure of students' performance-oriented motivation to achieve particularized on assessment procedures [84]. Our objective was to better understand the identified engagement-related patterns based on data that originate from students' beliefs. Furthermore, according to literature, motivation has been strongly associated with engagement [2], [10], and thus, this information could be useful to explain where behaviors, feeling, and thoughts meet.

Specifically, motivation does not seem to play a role in solution *S1* (*M = 0.59, SD = 0.217*). High performers who have high arousal and low TTAW, have either high or low performance orientation from the self-assessment. Similarly, for solutions *S4* and *S5*, no motivation-based inferences on the engagement patterns can be made. In particular, the high performers identified by both solutions correspond to the same two cases (i.e., the same students). Those two students were cognitively engaged (high cognitive load) in the adaptive tests, and were physically and emotionally calm, but they had highly divergent expectations (*M = 0.69, SD = 0.438*).

On the contrary, the per case analysis shown that students explained by *S2* and *S6* were all highly motivated (*M = 0.81, SD = 0.158 - S2*; *M = 0.79, SD = 0.147 - S6*). In fact, those students who managed to control their emotion intensity (*S2*) or those who were strongly cognitively engaged and exhibited high attention (*S6*) appear to be motivated students (high goal expectations) who wanted to achieve high scores, and believed that they were well-prepared to take the test. Also, there were some students with slightly lower goal-expectancy (*M = 0.75, SD = 0.244*) who were physically calm and managed to control their emotions, and achieved high scores (*S3*).

Furthermore, the medium/low performers in *S7* are medium/low motivated (*M = 0.39, SD = 0.221*), having less achievement expectations in the self-assessment. The students in *S7* appear to be not well prepared (since they have high TTAW), and not physically relaxed (high arousal). Similarly, the physically calm students who achieved medium/low scores, exhibited moderate levels of motivation as well, either when they had high levels of emotional intensity (*M = 0.51, SD = 0.046; S9*) or high attention (*M = 0.59, SD = 0.080; S8*).

V. Discussion, Contributions, and Implications

Adaptive learning activities offer learners the opportunity of tailored experiences with significant personalized learning benefits. Adaptation has a positive impact on learners' self-regulation and engagement with the activities [19], [20], which is also reflected on learners' performance [14], [90].

The review of relevant literature revealed numerous engagement-related factors that are tracked by the learning environment (e.g., [2], [29], [35]), or are captured with specialized equipment (e.g., sensors) (e.g., [5], [53], [62], [66]), or are directly extracted as perceptions from the learners, mostly in relation to their motivation and self-regulation (e.g., [10], [13]). Researchers used diverse data sources to capture the multidimensionality of engagement, and understand and explain its role in the learning process (*why, how, and when*).

Given (a) the multidimensionality of engagement as a concept itself [4], [47]; (b) the difficulty in collecting, fusing and analysing data from multiple channels to deeper understand such complex mechanisms [50]; and (c) the limited results about engagement patterns in adaptive learning contexts (i.e., driven by the particularities of the context itself) [35], this study considers engagement as a complex process that involves learners' thoughts, behaviors, and feelings, and explores different combinations of engagement-related mutlimodal data to justify learning performance in adaptive self-assessment conditions. Furthermore, previous studies explored mostly symmetric relationships between the data, using variance-based approaches for hypotheses testing (e.g., [10]). The present study adopted the fsQCA [45] method for exploring multiple configurations of causal conditions which may include different combinations of the engagement-related multimodal variables. The results provided several interesting findings.

A. Insights and Implications from the Fused Multimodal Data

The engagement patterns identified for high performing students consisted of 6 configurations (solutions) of the multimodal factors (i.e., arousal, cognitive load, emotional intensiveness, attention, response time), The respective engagement configurations for medium/low performers were 3 (Table III).

One of the most interesting findings was that cognitive load does not participate in none of the solutions for medium/low performance. This result should not be surprising, because cognitive load is related to mental work and information processing [42], [60]. The low values of medium/low performers on TTAC (*S8, S9*) and the high values on TTAW (*S7*) indicate that those students do not focus (in terms of response time) on finding the correct answers, either because they are not well prepared [29] (further validated by their self-reported medium/low goal-expectation) or because they are not deeply processing the available information. In each case, the mental efforts are more likely to be low; even though their attention is high (*S8*), they might focus on the wrong information or not try to further process and fully understand the self-assessment items. Furthermore, not being physically or emotionally calm (high arousal–*S7*; high emotion intensity–*S9*) indicates that those students are probably experiencing stress [57], which prohibits deeper thinking and information analysis and is negatively correlated with self-efficacy [2], [11].

On the contrary, cognitive load is high in solutions *S4, S5*, and *S6* which correspond to engagement-related patterns that explain high performance, indicating that when students are mentally engaged, they are more likely to perform well. This finding indicates that mental engagement (high cognitive load) alone is not enough information to draw conclusions: emotional intensity and arousal (*S4*) or emotional intensity and attention (*S5*) or attention and response time (*S6*) can determine in what ways the learner engage with the activity to achieve high score. The critical question to address is *how to keep students mentally engaged with the learning activity*.

Another interesting finding is that emotion intensity is low in all high performance solutions it participates, whereas it is high in the one medium/low performance solution identified with this factor being present. This means that when the students can control their enthusiasm, fears, anger, overall their intense emotions, and exhibit emotional stability, it is likely that they will regulate themselves and achieve higher scores.

Previous work on the emotional aspect of engagement focused on the valence of emotions rather than on their intensity [9], [62]. Our results contribute to literature by showing how learners' feelings (from facial expressions) can moderate performance if learners have control over their intensity.

Furthermore, high arousal has been found to correlate with stress [57]. In a sense, in *S1*, stress can be seen as a factor that contributes positively to achieving high scores. It should be noted that this is the only solution (that explains high performance) that considers high arousal. Surprisingly, this solution explains the larger part of high performers (58%).

Remarkably, although students in *S5* did not exhibit high attention, they achieved high scores. Low attention had been noticed when the students read the tasks multiple times or they are not reading carefully [43]. Low attention behaviors have been found to be correlated with increased cognitive load when students are experiencing stress [91]. The fact that the students in *S5* managed to control their feelings and used their mental processing effectively to give correct answers, led them to overcome their lack of attention, and to achieve high scores.

Finally, from the per case analysis (Fig. 2) it was found that the motivation of students explained by the discovered high performance solutions, was medium/high for all cases. The goal-expectations from the self-assessment for the medium/low performing students were medium/low, as well. It should be noted that students' motivation to take the self-assessment was measured as perceived goal-expectations from the self-assessment before the adaptive procedure, but was used at a later stage of the analysis for enhancing the solutions with the necessary contextual information, to help us understand and interpret the on-task measurable engagement during the activity. However, although the relationship between motivation and performance is straightforward—further supporting previous work [2], [84]—our results concerning the relations between motivation and engagement were not clear. All students, regardless of their motivation (i.e., either highly, moderately or low motivated), exhibited different engagement behaviors: *no solution that included low values in all participating engagement-related factors was found*. Thus, we could not associate motivation with engagement patterns.

However, this finding prompts for further and more careful exploration of the role of each one of the multimodal factors and their combinations with motivation. For example, a motivated student might be emotionally calm and mentally active, but at the same time to exhibit reduced attention (*S5*), or a low motivated student may exhibit high attention and be physically calm, and not perform high (*S8*). Engaging does not necessarily means efficient learning: *S8* for medium/low performance describes students who are moderately motivated, have high attention, remain physically calm (low arousal), yet they fail to submit correct answers; they focus on the tasks, but lack the knowledge to successfully solve those tasks or the ability to further process them. Thus, the goal is not only to ensure better engagement conditions with the learning tasks (e.g., affective feedback, adaptive content), but to persuade the learners that their involvement can be self-rewarding in the long-term (e.g., gain motivation from achievement).

B. Insights and Implications from the Data Analysis Method

As seen from Table III, the solutions that explain high performance are not the exact opposites of the respective ones that explain "not high," i.e., medium/low performance. The discovery of such *asymmetric relations* prompts for designing feedback and/or services—beyond the "*one-size-fits-all*" approaches—to achieve highly fine-grained personalization and adaptation to the specific needs of smaller sub-populations within groups, that otherwise would be difficult to identify.

It is also interesting that some cases appear in more than one solutions, and all cases in *S6* appear in other solutions (*S1, S2, S3*). This exemplifies the "equifinality" of fsQCA: combinations of conditions are sufficient but not necessary to explain the outcome, as more than one combinations can lead to the same outcome. Such conditions are *insufficient but necessary* parts of causal combinations, which in turn are *unnecessary but sufficient* to explain the outcome [45]. As explained in section III.D, the variables work in synergy to holistically explain the outcome; the same learner can achieve high score e.g., through the combination of *S2* or through the combination of *S3* (those solutions describe the same sub-population, who, at the same time, has diverse characteristics). This highlights that there is no single perfect solution; there might be more than one sufficient conditions that can lead to the outcome, fitting the diverse user characteristics.

Methodologically, the principles of causal asymmetry and equifinality—inherent in fsQCA—can provide multiple complementary yet not necessarily contradictive solutions based on the fused multimodal engagement-related data. From that point on, based on researcher's experience and on available data, different designs and solutions can be employed to achieve deeper engagement both on task and on activity levels.

C. The Role of Adaptivity—Implications for Adaptive Systems

Previous results in adaptive learning contexts revealed the positive impact of adaptivity on motivation and engagement [19], [20]. Our findings further confirmed and extended those results. The adaptive self-assessment appears to be a learning context that on its own promotes and facilitates learners' engagement, regardless of their motivation and achievement, and is above and beyond learners' goal-orientation.

Furthermore, the aim of the study was to come up with generalizable results that can be used to inform the design of adaptive systems *for all*, and therefore, the data that were collected directly from the learning environment are common clickstreams (e.g., response times) easily captured by and extensively used in all adaptive learning systems.

Indeed, the results from this study demonstrated that delivering the most appropriate content to the learners can retain their engagement, regardless of their motivation or learning performance, and prevent "disengagement." As explained before, *no solution that included low values in all engagement-related factors was found*. The adaptive system delivered to students the next item based on their knowledge mastery (not on their engagement levels). Although students' involvement was different in terms of thoughts, feelings and behaviors, their engagement *as a whole* was not absent, probably because all students were involved in answering the items that better fitted their own abilities. This understanding can lead us to design decisions for adaptive systems on what kind of support (feedback) those students might need. For example, when the detected levels of arousal are high, the system could deliver a much easier task to help the learner gain calmness and give her the opportunity to answer correctly and feel more self-confident. This is important because it can contribute to improving the adaptation mechanism and to timely provide proactive (cognitive, metacognitive, or affective) feedback to prevent students from exhibiting disengagement from tasks, even though the tasks are tailored to their ability, and to encourage them to increase their attention and control their emotional and cognitive arousal. This implies that students who might have been "trapped" into a disengagement behavior that could hinder their success, could be further supported with cognitive and/or affective feedback to push them out of this loop. Similarly, when learners are in one of the S1-S6 solutions, they can remain "engaged" on their efforts to solve the tasks. To keep them in this physiological state, providing affective feedback praising the good work might work.

Actionable feedback is one of the most important issues to be dealt with in adaptive learning. By using learners' physiological data from different channels, we showed in this paper, that based on these characteristics that drive learners' engagement, we shall be able to provide this kind of feedback.

D. Generalizability and reproducibility of the results

This study utilized multimodal data from a variety of data sources that are not easy to collect for larger number of participants. Published work in the area demonstrated results typically with smaller samples or with less multimodal data channels [54], [58]. Furthermore, some of the reasons why small sample sizes are sufficient in studies that use multimodal data are: (i) the data collected are "big" in terms of the 4Vs' (Volume, Variety, Veracity, Velocity). For instance, eye-tracking data collected at a high frequency (e.g.,120 Hz in the present study) means that we have continuous measurement of the users' behavior. Collecting this kind of data results into continuously and massively gathering a few Gigabytes of data per person (Volume and Velocity). Furthermore, collecting data from multiple sources at once (i.e., eye-tracking, EEG, wristband, face videos, clickstream) satisfies Variety, whereas, previous research has utilized those data for computing arousal, emotion, memory load, working memory activation, cognitive load, attention, fatigue (Veracity); (ii) the current cost of the equipment necessary to collect those data does not allow for simultaneous use of multiple devices, but the granularity of information we can have access to, justifies their usage. Based on these reasons it is safe to say that 32 participants are indeed sufficient to arrive at generalizable conclusions.

Furthermore, it should be clarified that we did not find 9 solutions that are splitting the sample so as each participant to be represented in only one solution. The method explains each participant and not the variance in the sample. Specifically, the method allows us to identify 6 solutions that explain high performance, and 3 other solutions that explain medium/low performance (based on the principles of equifinality that multiple complex configurations of the same conditions may explain the same outcome [67], which also are not exact mirrors of each other (based on the notion of causal asymmetry, [45]). Bringing all these on the same page, about the representativeness (generalizability) of the solutions in the sample, it becomes apparent that the combination of the data considered (i.e., physiological data that are common to the population) with the analysis method (explains each participant, and as such, can cover all possible cases), provide technically sound approach in which all cases are represented by the solutions identified, and the results can be generalized.

E. Conclusions and Limitations

This study demonstrated a consolidated analysis of multimodal data collected during an adaptive self-assessment activity, utilizing fsQCA for deeper understanding engagement in this setting. What this study adds to engagement literature is that when the learning tasks facilitate one's own learning needs (motivation), it is likely that one will be deeper and more substantially involved with those tasks, yet the thorough analysis showcased that multimodal data can provide *more than one* engagement patterns to facilitate this objective.

However, as this study is among the first to employ fsQCA in learning analytics research ([29], [69], [72]), further experimentation is needed to identify complex and important configurations and reveal the full potential of the analysis. Researchers' experience with data calibration is also a limitation.

Furthermore, future studies that incorporate data from various adaptive learning activities and modalities, are within our future work plans towards making-sense of complex learning interactions and offering a holistic understanding of the potential of this data analysis technique in TEL research.

References

[1] J. A. Fredricks, P. C. Blumenfeld, and A. H. Paris, "School engagement: Potential of the concept, state of the evidence," *Rev. Educ. Res.*, vol. 74, no. 1, pp. 59–109, 2004.

[2] A. Pardo, F. Han, and R. A. Ellis, "Combining university student self-regulated learning indicators and engagement with online learning events to predict academic performance," *IEEE Trans. Learn. Technol.*, vol. 10, no. 1, pp. 82–92, 2016.

[3] I.-S. Chen, "Computer self-efficacy, learning performance, and the mediating role of learning engagement," *Comput. Hum. Behav.*, vol. 72, pp. 362–370, 2017.

[4] M. A. Lawson and H. A. Lawson, "New conceptual frameworks for student engagement research, policy, and practice," *Rev. Educ. Res.*, vol. 83, no. 3, pp. 432–479, 2013.

[5] K. Sharma, Z. Papamitsiou, and M. Giannakos, "Building pipelines for educational data using AI and multimodal analytics: A "grey-box" approach," *Br. J. Educ. Technol.*, vol. 50, no. 6, pp. 3004–3031, 2019.

[6] A. Ben-Eliyahu, D. Moore, R. Dorph, and C. D. Schunn, "Investigating the multidimensionality of engagement: Affective, behavioral, and cognitive engagement across science activities and contexts," *Contemp. Educ. Psychol.*, vol. 53, pp. 87–105, 2018.

[7] E. Joseph, "Engagement tracing: using response times to model student disengagement," in *Artificial intelligence in education: Supporting learning through intelligent and socially informed technology*. Ios Press, 2005, vol. 125, pp. 88–95.

[8] C. R. Henrie, L. R. Halverson, and C. R. Graham, "Measuring student engagement in technology-mediated learning: A review," *Comput. Educ.*, vol. 90, pp. 36–53, 2015.

[9] S. Molinillo, R. Aguilar-Illescas, R. Anaya-Sánchez, and M. Vallespín-Arán, "Exploring the impacts of interactions, social presence and emotional engagement on active collaborative learning in a social web-based environment," *Comput. Educ.*, vol. 123, pp. 41–52, 2018.

[10] B. A. Greene, "Measuring cognitive engagement with self-report scales: Reflections from over 20 years of research," *Educ. Psychol.*, vol. 50, no. 1, pp. 14–30, 2015.

[11] J. C.-Y. Sun and R. Rueda, "Situational interest, computer self-efficacy and self-regulation: Their impact on student engagement in distance education," *Br. J. Educ. Technol.*, vol. 43, no. 2, pp. 191–204, 2012.

[12] A. S. Sunar, S. White, N. A. Abdullah, and H. C. Davis, "How learners' interactions sustain engagement: a MOOC case study," *IEEE Trans. Learn. Technol.*, vol. 10, no. 4, pp. 475–487, 2016.

[13] J. Guerra, R. Hosseini, S. Somyurek, and P. Brusilovsky, "An intelligent interface for learning content: Combining an open learner model and social comparison to support self-regulated learning and engagement," in *Proc. 21st Int. Conf. Intelligent User Interfaces*, ser. IUI '16. New York, NY, USA: ACM, 2016, pp. 152—163.

[14] M. Liu, E. McKelroy, S. B. Corliss, and J. Carrigan, "Investigating the effect of an adaptive learning intervention on students' learning," *Educ. Technol. Res. Dev.*, vol. 65, no. 6, pp. 1605–1625, 2017.

[15] A. Kaplan and M. L. Maehr, "The contributions and prospects of goal orientation theory," *Educ. Psychol. Rev.*, vol. 19, no. 2, pp. 141–184, 2007.

[16] A. J. Martin, P. Ginns, and B. Papworth, "Motivation and engagement: Same or different? does it matter?" *Learn. Individ. Differ.*, vol. 55, pp. 150–162, 2017.

[17] P. R. Pintrich, "The role of goal orientation in self-regulated learning," in *Handbook of self-regulation*. Elsevier, 2000, pp. 451–502.

[18] F. Bouchet, J. M. Harley, and R. Azevedo, "Can adaptive pedagogical agents' prompting strategies improve students' learning and self-regulation?" in *Intelligent Tutoring Systems*, A. Micarelli, J. Stamper, and K. Panourgia, Eds. Springer, 2016, pp. 368–374.

[19] B. Bontchev and D. Vassileva, "Assessing engagement in an emotionally-adaptive applied game," in *Proc. 4th Int. Conf. Technological Ecosystems for Enhancing Multiculturality*, ser. TEEM '16. New York, NY, USA: ACM, 2016, pp. 747–754.

[20] N. B. A. Normadhi, L. Shuib *et al.*, "Identification of personal traits in adaptive learning environment: Systematic literature review," *Comput. Educ.*, vol. 130, pp. 168–190, 2019.

[21] S. A. Becker, M. Brown *et al.*, "Horizon report 2018 higher education edition brought to you by educause," EDU18, Tech. Rep., 2018.

[22] A. Paramythis and S. Loidl-Reisinger, "Adaptive learning environments and e-learning standards," in *2nd European Conference on e-Learning*, 2003, pp. 369–379.

[23] K. Premlatha and T. Geetha, "Learning content design and learner adaptation for adaptive e-learning environment: a survey," *Artif. Intell. Rev.*, vol. 44, no. 4, pp. 443–465, 2015.

[24] Y. Chen, X. Li, J. Liu, and Z. Ying, "Recommendation system for adaptive learning," *Appl. Psychol. Meas.*, vol. 42, no. 1, pp. 24–41, 2018.

[25] R. Bodily, J. Kay, V. Aleven *et al.*, "Open learner models and learning analytics dashboards: A systematic review," in *Proc. 8th Int. Conf. Learning Analytics & Knowledge*, ser. LAK '18. New York, NY, USA: ACM, 2018, pp. 41—50.

[26] B. Towle and M. Halm, *Designing Adaptive Learning Environments with Learning Design*. Springer Berlin Heidelberg, 2005, pp. 215–226.

[27] F. Bouchet, J. M. Harley, and R. Azevedo, "Impact of different pedagogical agents' adaptive self-regulated prompting strategies on learning with metatutor," in *Artif. Intell. Educ.*, H. C. Lane, K. Yacef *et al.*, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 815–819.

[28] K. S. Ostrow, "Enhancing student motivation and learning within adaptive tutors," in *International Conference on Educational Data Mining (EDM)*, 2015, pp. 668–670.

[29] Z. Papamitsiou, A. A. Economides, I. O. Pappas, and M. N. Giannakos, "Explaining learning performance using response-time, self-regulation and satisfaction from content: An fsqca approach," in *Proc. 8th Int. Conf. Learning Analytics and Knowledge*, ser. LAK '18. New York, NY, USA: ACM, 2018, pp. 181–190.

[30] S. D'Mello, K. Kopp, R. E. Bixler, and N. Bosch, "Attending to attention: Detecting and combating mind wandering during computerized reading," in *Proc. 2016 CHI Conf. Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '16. New York, NY, USA: ACM, 2016, pp. 1661–1669.

[31] Z. Papamitsiou and A. A. Economides, "Exploring autonomous learning capacity from a self-regulated learning perspective using learning analytics," *Br. J. Educ. Technol.*, vol. 50, no. 6, pp. 3138–3155, 2019.

[32] M. C. Duffy and R. Azevedo, "Motivation matters: Interactions between achievement goals and agent scaffolding for self-regulated learning within an intelligent tutoring system," *Comput. Hum. Behav.*, vol. 52, pp. 338–348, 2015.

[33] A. J. Elliot and K. Murayama, "On the measurement of achievement goals: Critique, illustration, and application," *J. Educ. Psychol.*, vol. 100, no. 3, pp. 613–628, 2008.

[34] H. Jiao, "Enhancing students' engagement in learning through a formative e-assessment tool that motivates students to take action on feedback," *Australasian Journal of Engineering Education*, vol. 20, no. 1, pp. 9–18, 2015.

[35] B. Ross, A.-M. Chase, D. Robbie, G. Oates, and Y. Absalom, "Adaptive quizzes to increase motivation, engagement and learning outcomes in a first year accounting unit," *International Journal of Educational Technology in Higher Education*, vol. 15, no. 1, p. 30, 2018.

[36] X. Ochoa, "Multimodal Learning Analytics," in *The Handbook of Learning Analytics*, 1st ed. Alberta, Canada: Society for Learning Analytics Research (SoLAR), 2017, pp. 129–141.

[37] B. Grawemeyer *et al.*, "Affective learning: improving engagement and enhancing learning with affect-aware feedback," *User Model. User-Adapted Interact.*, vol. 27, no. 1, pp. 119–158, 2017.

[38] D. Azcona, I.-H. Hsiao, and A. Smeaton, "An exploratory study on student engagement with adaptive notifications in programming courses," in *Lifelong Technology-Enhanced Learning*, V. Pammer-Schindler *et al.*, Eds. Cham: Springer International Publishing, 2018, pp. 644–647.

[39] H. J. Parkin, S. Hepplestone, G. Holden, B. Irwin, and L. Thorpe, "A role for technology in enhancing students' engagement with feedback," *Assessment & Evaluation in Higher Education*, vol. 37, no. 8, pp. 963–973, 2012.

[40] S. Sampayo-Vargas, C. J. Cope, Z. He, and G. J. Byrne, "The effectiveness of adaptive difficulty adjustments on students' motivation and learning in an educational computer game," *Comput. Educ.*, vol. 69, pp. 452–462, 2013.

[41] K. A. Renninger and J. E. Bachrach, "Studying triggers for interest and engagement using observational methods," *Educ. Psychol.*, vol. 50, no. 1, pp. 58–69, 2015.

[42] S. H. Fairclough, L. J. Moores, K. C. Ewing, and J. Roberts, "Measuring task engagement as an input to physiological computing," in *Proc. 3rd Int. Conf. Affective Computing and Intelligent Interaction and Workshops*. IEEE, 2009, pp. 1–9.

[43] B. W. Miller, "Using reading times and eye-movements to measure cognitive engagement," *Educ. Psychol.*, vol. 50, no. 1, pp. 31–42, 2015.

[44] C. Wang and P. Cesar, "Physiological measurement on students' engagement in a distributed learning environment." in *2nd Int. Conf. on Physiological Computing Systems*, 2015, pp. 149–156.

[45] C. C. Ragin, *Redesigning social inquiry: Fuzzy sets and beyond*. University of Chicago Press, 2008.

[46] A. G. Woodside, "Embrace• perform• model: Complexity theory, contrarian case analysis, and multiple realities," *J. Bus. Res.*, vol. 67, no. 12, pp. 2495–2503, 2014.

[47] R. Azevedo, "Defining and measuring engagement and learning in science: Conceptual, theoretical, methodological, and analytical issues," *Educ. Psychol.*, vol. 50, no. 1, pp. 84–94, 2015.

[48] L. Ding, E. Er, and M. Orey, "An exploratory study of student engagement in gamified online discussions," *Comput. Educ.*, vol. 120, pp. 213–226, 2018.

[49] R. F. Kizilcec, C. Piech, and E. Schneider, "Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses," in *Proc. 3rd Int. Conf. Learning Analytics and Knowledge*, ser. LAK '13. New York, NY, USA: ACM, 2013, pp. 170—179.

[50] M. N. Giannakos, K. Sharma, I. O. Pappas, V. Kostakos, and E. Velloso, "Multimodal data as a means to understand the learning experience," *Int. J. Inf. Manag.*, vol. 48, pp. 108–119, 2019.

[51] H. C. Lane and S. K. D'Mello, "Uses of physiological monitoring in intelligent learning environments: A review of research, evidence, and technologies," in *Mind, Brain and Technology*. Springer, 2019, pp. 67–86.

[52] R. Bixler and S. D'Mello, "Automatic gaze-based user-independent detection of mind wandering during computerized reading," *User Model. User-Adapted Interact.*, vol. 26, no. 1, pp. 33–68, 2016.

[53] J. Whitehill, Z. Serpell, Y.-C. Lin, A. Foster, and J. R. Movellan, "The faces of engagement: Automatic recognition of student engagementfrom facial expressions," *IEEE Trans. Affective Comput.*, vol. 5, no. 1, pp. 86–98, 2014.

[54] H. J. Pijeira-Díaz, H. Drachsler, S. Järvelä, and P. A. Kirschner, "Sympathetic arousal commonalities and arousal contagion during collaborative learning: How attuned are triad members?" *Comput. Hum. Behav.*, vol. 92, pp. 188–197, 2019.

[55] C. Mills *et al.*, "Put your thinking cap on: Detecting cognitive load using eeg during learning," in *Proc. 7th Int. Conf. Learning Analytics & Knowledge*. New York, NY, USA: ACM, 2017, pp. 80–89.

[56] K. R. Thorson, T. V. West, and W. B. Mendes, "Measuring physiological influence in dyads: A guide to designing, implementing, and analyzing

dyadic physiological studies." *Psychol. Methods*, vol. 23, no. 4, pp. 595–616, 2018.

[57] J. Taelman, S. Vandeput, A. Spaepen, and S. Van Huffel, "Influence of mental stress on heart rate and heart rate variability," in *4th European Conf. of the International Federation for Medical and Biological Engineering*. Springer, 2009, pp. 1366–1369.

[58] D. Di Mitri, M. Scheffel *et al.*, "Learning pulse: A machine learning approach for predicting performance in self-regulated learning using multimodal data," in *Proc. 7th Int. Conf. Learning Analytics & Knowledge*, ser. LAK '17. New York, NY, USA: ACM, 2017, pp. 188–197.

[59] D. L. Butler and P. H. Winne, "Feedback and self-regulated learning: A theoretical synthesis," *Rev. Educ. Res.*, vol. 65, no. 3, pp. 245–281, 1995.

[60] R. H. Stevens, T. Galloway, and C. Berka, "Eeg-related changes in cognitive workload, engagement and distraction as students acquire problem solving skills," in *Int. Conf. User Modeling*. Springer, 2007, pp. 187–196.

[61] H. Monkaresi, N. Bosch, R. A. Calvo, and S. K. D'Mello, "Automated detection of engagement using video-based estimation of facial expressions and heart rate," *IEEE Trans. Affective Comput.*, vol. 8, no. 1, pp. 15–28, 2016.

[62] S. K. D'Mello, N. Bosch, and H. Chen, "Multimodal-multisensor affect detection," in *The Handbook of Multimodal-Multisensor Interfaces*. Association for Computing Machinery and Morgan & Claypool, 2018, pp. 167–202.

[63] J. Grafsgaard, J. B. Wiggins, K. E. Boyer, E. N. Wiebe, and J. Lester, "Automatically recognizing facial expression: Predicting engagement and frustration," in *Educational Data Mining 2013*, 2013.

[64] A. Andrade, G. Delandshere, and J. A. Danish, "Using multimodal learning analytics to model student behavior: A systematic analysis of epistemological framing," *Journal of Learning Analytics*, vol. 3, no. 2, pp. 282–306, 2016.

[65] M. Taub and R. Azevedo, "How does prior knowledge influence eye fixations and sequences of cognitive and metacognitive srl processes during learning with an intelligent tutoring system?" *J. Artif. Intell. Educ.*, vol. 29, no. 1, pp. 1–28, 2019.

[66] S. Hutt *et al.*, ""Out of the fr-eye-ing pan": Towards gaze-based models of attention during learning with technology in the classroom," in *Proc. 25th Conf. User Modeling, Adaptation and Personalization*, ser. UMAP '17. New York, NY, USA: ACM, 2017, pp. 94–103.

[67] A. G. Woodside, "Moving beyond multiple regression analysis to algorithms: Calling for adoption of a paradigm shift from symmetric to asymmetric thinking in data analysis and crafting theory," *J. Bus. Res.*, vol. 66, no. 4, pp. 463–472, 2013.

[68] I. O. Pappas, M. N. Giannakos, and D. G. Sampson, "Fuzzy set analysis as a means to understand users of 21st-century learning systems: The case of mobile learning and reflections on learning analytics research," *Comput. Hum. Behav.*, vol. 92, pp. 646–659, 2019.

[69] I. Pappas, M. Giannakos, and D. Sampson, "Making Sense of Learning Analytics with a Configurational Approach," in *LAK 2016 Workshop on Smart Environments and Analytics in Video-Based Learning*, 2016, pp. 42–52.

[70] I. O. Pappas, M. N. Giannakos, L. Jaccheri, and D. G. Sampson, "Assessing student behavior in computer science education with an fsqca approach: The role of gains and barriers," *ACM T. Comput. Educ.*, vol. 17, no. 2, pp. 10:1–10:23, 2017.

[71] C. Plewa, J. Ho, J. Conduit, and I. O. Karpen, "Reputation in higher education: A fuzzy set analysis of resource configurations," *J. Bus. Res.*, vol. 69, no. 8, pp. 3087–3095, 2016.

[72] S. Sergis, D. G. Sampson, and M. N. Giannakos, "Supporting school leadership decision making with holistic school analytics: Bridging the qualitative-quantitative divide using fuzzy-set qualitative comparative analysis," *Comput. Hum. Behav.*, vol. 89, pp. 355–366, 2018.

[73] H. Andrade and A. Valtcheva, "Promoting learning and achievement through self-assessment," *Theory into Practice*, vol. 48, no. 1, pp. 12–19, 2009.

[74] D. J. Nicol and D. Macfarlane-Dick, "Formative assessment and self-regulated learning: A model and seven principles of good feedback practice," *Stud. High. Educ.*, vol. 31, no. 2, pp. 199–218, 2006.

[75] I. Mitchell and A. Carbone, "A typology of task characteristics and their effects on student engagement," *J. Educ. Res.*, vol. 50, no. 5-6, pp. 257–270, 2011.

[76] Z. Zainuddin, M. Shujahat, H. Haruna, and S. K. W. Chu, "The role of gamified e-quizzes on student learning and engagement: An interactive gamification solution for a formative assessment system," *Comput. Educ.*, p. 103729, 2019.

[77] O. O. Adesope, D. A. Trevisan, and N. Sundararajan, "Rethinking the use of tests: A meta-analysis of practice testing," *Rev. Educ. Res.*, vol. 87, no. 3, pp. 659–701, 2017.

[78] S. K. Carpenter, "Cue strength as a moderator of the testing effect: the benefits of elaborative retrieval." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 35, no. 6, p. 1563, 2009.

[79] H. L. Roediger III and J. D. Karpicke, "Test-enhanced learning: Taking memory tests improves long-term retention," *Psychol. Sci.*, vol. 17, no. 3, pp. 249–255, 2006.

[80] A. D. Cantor, A. N. Eslick, E. J. Marsh, R. A. Bjork, and E. L. Bjork, "Multiple-choice tests stabilize access to marginal knowledge," *Mem Cogn*, vol. 43, no. 2, pp. 193–205, Feb 2015.

[81] S. K. Carpenter, "Testing enhances the transfer of learning," *Current directions in psychological science*, vol. 21, no. 5, pp. 279–283, 2012.

[82] Z. Papamitsiou and A. A. Economides, "Towards the alignment of computer-based assessment outcome with learning goals: The LAERS architecture," in *2013 IEEE Conf. e-Learning, e-Management and e-Services*, 2013, pp. 13–17.

[83] P. Antonenko, F. Paas, R. Grabner, and T. Van Gog, "Using electroencephalography to measure cognitive load," *Educ. Psychol. Rev.*, vol. 22, no. 4, pp. 425–438, 2010.

[84] Z. K. Papamitsiou, V. Terzis, and A. A. Economides, "Temporal learning analytics for computer based testing," in *Proc. 4th Int. Conf. Learning Analytics And Knowledge*, ser. LAK '14. New York, NY, USA: Association for Computing Machinery, 2014, pp. 31–35.

[85] B. Rihoux and C. C. Ragin, *Configurational comparative methods: Qualitative comparative analysis (QCA) and related techniques*. Sage Publications, 2009, vol. 51.

[86] T. Greckhamer, V. F. Misangyi, and P. C. Fiss, "Chapter 3 the two qcas: From a small-n to a large-n set theoretic approach," in *Configurational theory and methods in organizational research*. Emerald Group Publishing, 2013, pp. 49–75.

[87] I. O. Pappas, "User experience in personalized online shopping: a fuzzy-set analysis," *European Journal of Marketing*, vol. 52, no. 7/8, pp. 1679–1703, 2018.

[88] P. C. Fiss, "Building better causal theories: A fuzzy set approach to typologies in organization research," *Acad. Manag. Ann.*, vol. 54, no. 2, pp. 393–420, 2011.

[89] J. Dul, "Identifying single necessary conditions with nca and fsqca," *J. Bus. Res.*, vol. 69, no. 4, pp. 1516–1523, 2016.

[90] F. M. Van der Kleij, R. C. Feskens, and T. J. Eggen, "Effects of feedback in a computer-based learning environment on students' learning outcomes: A meta-analysis," *Rev. Educ. Res.*, vol. 85, no. 4, pp. 475–511, 2015.

[91] S. Najmi, N. Amir, K. E. Frosio, and C. Ayers, "The effects of cognitive load on attention control in subclinical anxiety and generalised anxiety disorder," *Cogn. Emot.*, vol. 29, no. 7, pp. 1210–1223, 2015.

Zacharoula Papamitsiou is a senior researcher at the Computer Science Department of NTNU. She holds a Ph.D. in adapting the learning services for supporting learners' decision-making using learning analytics. Her research interest is on quantified-self technologies, learner-computer interaction, and autonomous learning.

Ilias O. Pappas is an Associate Professor of Information Systems at the Department of Information Systems, University of Agder, Norway, and holds a Ph.D. in business management and information systems. His teaching and research activities focus on digital transformation, social innovation and social change, as well as Internet marketing and information technology adoption.

Kshitij Sharma is a senior researcher at the Computer Science department of the Norwegian University of Science and Technology. He received his Ph.D. in Computer Science from EPFL. His research interests include eye-tracking, MOOCs, collaborative learning, multimodal learning, applied machine learning, and statistics.

Michail N. Giannakos is a Professor of interaction design and learning technologies at the Department of Computer Science of NTNU, and Head of the Learner-Computer Interaction lab. He is associate editor and/or guest editor on highly recognized journals (e.g., IEEE TLT, BJET, CHB, IEEE ToE, ACM TOCE). He is also an IEEE Senior Member and one of the outstanding academic fellows of NTNU (2017-2021).