Explaining Learning Performance Using Response-Time, Self-Regulation and Satisfaction from Content: An fsQCA Approach

Zacharoula Papamitsiou
University of Macedonia, Information Systems Program
156 Egnatia Avenue, Thessaloniki, 54621
Greece
papamits@uom.edu.gr

Anastasios A. Economides
University of Macedonia, Information Systems Program
156 Egnatia Avenue, Thessaloniki, 54621
Greece
economid@uom.gr

Ilias O. Pappas
Norwegian University of Science and Technology,
Department of Computer Science
Sem Sælands vei 7-9, Trondheim, 7491
Norway
ilpappas@ntnu.no

Michail N. Giannakos
Norwegian University of Science and Technology,
Department of Computer Science
Sem Sælands vei 7-9, Trondheim, 7491
Norway
michailg@ntnu.no

ABSTRACT

This study focuses on compiling students' response-time allocated to answer correctly or wrongly, their self-regulation, as well as their satisfaction from content, in order to explain high or medium/low learning performance. To this end, it proposes a conceptual model in conjunction with research propositions. For the evaluation of the approach, an empirical study with 452 students was conducted. The fuzzy set qualitative comparative analysis (fsQCA) revealed five configurations driven by the admitted factors that explain students' high performance, as well as five additional patterns, interpreting students' medium/low performance. These findings advance our understanding of the relations between actual usage and latent behavioral factors, as well as their combined effect on students' test score. Limitations and potential implications of these findings are also discussed.

CCS CONCEPTS

• **Applied computing~E-learning**

KEYWORDS

Fuzzy set qualitative comparative analysis, configurations, response-time, self-regulation, satisfaction from content

ACM Reference Format:

Z. Papamitsiou, A. A. Economides, I. O. Pappas, and M. N. Giannakos. 2018. Explaining learning performance using response-time, self-regulation and satisfaction from content: an fsQCA approach. In *LAK'18: International Conference on Learning Analytics and Knowledge, March 7–9, 2018, Sydney, NSW, Australia. ACM, New York, NY, USA*, 10 pages. https://doi.org/10.1145/3170358.3170397

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
LAK '18, March 7–9, 2018, Sydney, NSW, Australia
© 2018 Association for Computing Machinery.
ACM ISBN 978-1-4503-6400-3/18/03…$15.00
https://doi.org/10.1145/3170358.3170397

1 INTRODUCTION

Online assessment tests are a typical and popular format for the evaluation of knowledge acquisition [1,19]. In general, testing procedures are treated by the teachers worldwide as "diagnostic tools" to gradually mark their students' progress on the course and measure the learning gain, i.e., the learning performance[5]. It is common practice to use tests to measure academic performance, since they setup a preamble of students' overall achievements on a specific course; grades are required at the end of the course and they are critical to the students' academic success. Thus, tests could be regarded as a "mean" to early distinguish students who are likely to achieve high or dropout.

The existing methods (i.e., Classical Test Theory and Computerized Adaptive Testing) have provided well-established testing formats. However, assessment tests have received comprehensive criticism; chasing grades may distract students from deeper learning [51], yet good grades do not necessarily reflect mastery [8] and put academic honesty in question, since they are conducive to cheating [1]. Gaining in-depth insight of students' interactions and seeking for explanation of their actions in testing contexts is a demand to further interpret the test result, and the overall learning gain and performance.

Towards understanding students' behavior during assessment tests, prior studies have contributed by holistically exploring students' response-time, i.e., by analyzing the amounts of time the students allocate on test items [24]. It was claimed that response-time should be treated as a fixed predictor [48]. It was also suggested that considering additional students' attributes – beyond response-time – might provide more concise prediction of their score [53]. The investigation of symmetric dependencies between goal expectations, correctness of answers, response-time and performance, provided encouraging findings [31]. Self-regulatory strategies and motivation [2,3,15,20], content validity and satisfaction from assessment items [9,14,25,36], demographic backgrounds [22] and personality traits [29] have been identified as factors that affect learning performance as well.

However, it is quite unclear which combinations of the previously identified factors better explain the obtained assessment outcome. The goal of this study is to build on complexity theory and to seek for specific patterns and configurations that foretell and explain students' performance and achievement level on assessment tests. Thus, the following research question guided this study:

***RQ**: What configurations of response-time, self-regulation and students' satisfaction from the test's content lead to high or medium/low learning performance?*

Determining these configurations is expected to advance our knowledge on why students act the way they do in assessment tests and how their scores reflect both their learning and actions, resulting to a more concise interpretation of performance.

To address the research question, it is crucial to use an analysis technique that can explore and identify important interrelationships amongst variables. In this study, we employ fuzzy set qualitative comparative analysis (fsQCA) [37], a striking alternative to traditional variance-based approaches [52]. When fsQCA is applied together with complexity theory, researchers have the opportunity to gain deeper and richer perspectives on their data [13,17,35,52]. In the technology enhanced learning context, fsQCA suits for explaining complex combinations and interdependencies between various forms of learning analytics and performance data, and can lead to interpreting the observed learning outcome [34,35,40]. Here, the detected asymmetries between the identified factors expand the results from previous studies [31] that elaborated on symmetric relationships and contribute to better explaining performance.

The reminder of the paper is organized as follows: next section briefly reviews the most relevant work and develops the propositions of this study. Section 3 outlines the research methodology and section 4 demonstrates the results. Finally, section 5 elaborates on the findings, limitations and possible implications, and summarizes conclusions drawn.

2 RELATED WORK, CONCEPTUAL MODEL, AND RESEARCH PROPOSITIONS

Explaining the students' learning performance and achievements is a timeless research topic. Over the past decades, several studies have reported results for addressing this objective, with respect to students' response-time [16,31], self-regulatory factors [15,36], as well as non-cognitive students' perceptions [7,42]. This section briefly reviews relevant literature for identifying the core factors to be further explored regarding their capacity to be combined for reasoning students' performance; i.e., all measures to be used in this study are carefully extracted from prior related work. We outline the underlying conceptual model along with the research propositions regarding these factors.

2.1 Response-time and behavioral factors for explaining students' performance

Scholars from the fields of Psychometrics and Intelligent Tutoring Systems have extensively explored time-related factors and investigated their appropriateness for explaining students' behavioral aspects during assessment [23]. For example, response-time were associated to lack of test-taking motivation and guessing behavior, coded in time-driven students' test-taking effort [6,41,50]. Results have shown that students tend to be more engaged in the beginning of a session, while guessing behaviors are more likely to occur when students are less engaged [2]. No direct relation of response-time to test score was found, though [6,16]. The result from exploring the efficiency of using students' previous response-time for directly predicting the correctness of their next actions and test score was statistically insignificant [53]. However, exploring indirect dependencies between correctness of answers, goal expectations, response-time and performance provided promising results [31].

From a different perspective, response-time was associated to pacing [18]. Although the studies that associate efficient use of available time with performance are limited, time-management seems to reduce the need to game test completion strategies [4]. Usually, lack of time-management can create a bad cycle for performance in the test, whereas, good time-management implies gaining control over the items and eventually, less stress. Thus, more items will probably be answered correctly, which is likely will be reflected on the test score. Indeed, high achieving students often exhibit strong time-management skills [26].

Time-management has been regarded as a self-regulation strategy. In general, self-regulatory strategies are acknowledged as significant predictors of students' overall performance and academic success [3]. In fact, high achieving students tend to demonstrated more self-regulatory skills [15,20]. Results from exploring the adoption of self-regulation for prediction of performance [20,44] indicate that reviewing responses and efficiently using the available time are the strategies that (during testing) affect test score more, whereas goal expectations are more predictive of performance prior to taking the test [20].

Moreover, students' perceived comprehensibility of the test items was explored regarding its effect on test result: since the test items are designed to measure knowledge acquisition, their validity and clarity are critical for students' response strategies [9,14,43]. If students understand the items and are satisfied from them, they are more likely to be successful [36]. Clarity of content was proposed as a determinant of student satisfaction [21,49]. This subjective perception of how well a learning item meets the student's expectations for learning and supports success [25] may help refine our insight on the test result.

However, response-time, self-regulation, or satisfaction exclusively can be unreliable to explain performance. Nonetheless, the existence of limited results throughout the literature, as mentioned above, suggests that more research is necessary on the interpretation of learning outcome, along with new methods that will offer fresh insight into the existing literature. The present study takes a different methodological approach by implementing configurational analysis.

2.2 fsQCA in technology enhanced learning and learning analytics

In Technology Enhanced Learning (TEL), a typical approach to knowledge construction is to evaluate innovative technologies empirically with user studies. In most of the cases, such evaluation is driven by front-loaded research questions/ hypotheses and concludes into the acceptance or rejection of the front-loaded assumptions. To do so, many studies in TEL analyse data using variance-based approaches such as analysis of variance (ANOVA) or multiple regression analysis (MRA). Using these methods, it is possible to examine net effects between variables, and conceptually each statistical test offers a single solution/model to explain the observed outcome. The key advantage of such methods is their ability to test hypotheses, whereby researchers use the method to test an assumption and obtain a single concrete assessment (e.g. mean, p-value, etc.).

Although learners with different characteristics may form different sub-groups within a sample, these sub-groups have the potential to be analysed in-depth and receive targeted and learner-centric design recommendations. For example, if in a sample we have performance data for participants, one possible configuration would be "learners with high response time, who spent a lot of time each day with system A and had taken similar courses in the past". Intuitively, when we consider the different configurations in our sample independently, it is likely that different configurations may lead to the same or to different outcomes. The inability of quantitative methods to account for such "relativist flexibility" in analysis has been a weakness that we often account for through qualitative data such as interviews.

Conversely, fsQCA has been designed to embrace the notion of configurations, and the fact that different ways of slicing the data may tell different stories. The main benefit of the technique is that it can identify multiple unique configurations that explain a large part of the sample. While variance-based approaches also explain parts of a sample, it is often the case that their models have a relatively low R^2 value, meaning that the model explains or predicts only a portion of the sample [52]. In such cases, it can be beneficial to use fsQCA to identify multiple configurations that jointly explain a much larger portion of the sample. As such, fsQCA explains certain parts of the sample that otherwise would had been considered as outliers. This is an important methodological difference: fsQCA can help us identify how to design learning technologies for all [34,35], unlike variance-based methods that test competing models to identify the fittest.

2.3 Conceptual Model and Research Propositions

Apparently, much of the research into response-time has endeavored to identify item factors and human factors that determine and affect the students' response-time strategies in testing conditions [16,18,31,53]. It was claimed that response-time should be treated as fixed predictor, because time-limit may affect students' performance [48]. Nonetheless, response-time only reflect the time-spent on individual items and do not tell the full story about how students complete a test.

To this end, self-regulation is expected to provide additional evidence regarding students' behavioral strategies. However, although reports of self-regulation in prior research have been found to be predictive of learning performance [54], very little is known about the role of self-regulation in test performance, and thus, estimating whether students' test outcomes are influenced by the use of a self-regulatory strategy is still an open issue[20]. Among the key self-regulatory processes expected to affect test performance are goal expectations, self-monitoring (reviewing responses), and time-management [26,44].

Moreover, students' satisfaction from content has been acknowledged for reflecting the consistency between expected gain and the actual experience [25]; more research is required in how perceived clarity of content influences performance.

This study posits that there is a synergy among self-regulation (goal expectations, time-management), response-time and satisfaction from content (perceived clarity of content) in predicting students' learning performance. Indeed, there is not one unique, optimal, configuration of such values. Instead, multiple and equally effective configurations of causal conditions exist, which may include different combinations of self-regulation, response-time and clarity of content. Depending on how they combine they may or may not explain students' high or medium/low performance. High performance refers to the presence of a condition, and 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 performance, that is medium/low performance. This approach allows the identification of asymmetrical relations among the examined factors and the outcome.

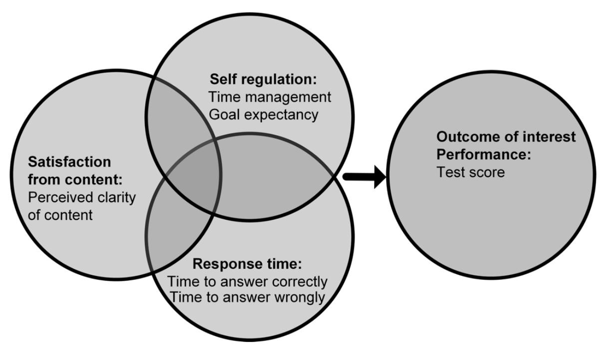

Figure 1: Venn diagram of the conceptual model.

To conceptualize these relationships, we propose a theoretical model (Figure 1) illustrating three constructs, their intersections, and the outcome of interest. On the left we present self-regulation (i.e., time management, goal expectancy), response-time (i.e., time to answer correctly and time to answer wrongly), and satisfaction from content (i.e., students' perception about the clarity of content). On the right, the outcome of interest is presented, that is students' performance on assessment tests. The overlapped areas represent possible combinations among factors, that is, areas that one factor may exist together with the other factors. Furthermore, to identify such patterns of factors in complex systems dedicated to educational measurement and assessment of learning (e.g., Computerized Adaptive Testing systems implementing complex Item Response Theory – IRT – solutions for large-scale high-stakes tests [47]), formulating hypotheses, common in variance based methods that are framed as correlational expressions, does not allow for a holistic approach that will lead to the identification of multiple solutions.

Indeed, in configuration theory approaches, research propositions are formulated as causal recipes to capture the different combinations among factors, and theoretically specify which should be present or absent from the causal recipe [10,28].

The principle of *equifinality* is inherent in both complexity theory and configuration theory, based on which a result may be equally explained by alternative sets of causal conditions [10,13]. In a complex system, relations among factors (i.e., causes) are also complex and depending on how they combine, both high and low conditions of a certain factor may explain high scores of an outcome. These conditions may be combined in sufficient configurations to explain the outcome [13,52]. For example, students may respond quickly and correctly because of a lucky guess, or because of high self-confidence due to high self-preparation to take the assessment test [31,41]. However, self-confident students who believe that they can perform well, tend to be more careful when they answer the test items: these students are more likely to persist longer in their efforts to accomplish the tasks successfully (higher response-time) than less self-reliant students [45]. Similarly, students may respond slowly and correctly because of lower perception regarding the clarity of items' content or because of carefulness in time-management and self-regulation. Studies have shown that questions' content and students' test performance are indirectly associated with each other, mediated by response-time [6]. Moreover, students who lack self-regulatory skills, tend to be more engaged in the beginning of a session and to exhibit guessing behavior at its end, not-trying to understand the test items and exhibiting low time-management [2,12]. As configurations can include different combinations of the examined constructs, they lead to the following proposition:

Proposition 1: *No single configuration of self-regulation, response-time, and satisfaction from content is sufficient for explaining high performance; instead, multiple, equally effective configurations exist.*

Further, configuration theory proposes the principle of causal asymmetry, which means that, for an outcome to occur, the presence and absence of a causal condition depends on how this condition combines with the other conditions [13,35]. A predictor variable may have an asymmetric relation with the outcome, which means that even if one variable is insufficient for the outcome to occur, it is still able to serve as a necessary condition for the outcome variable [13,52]. For example, using students' previous response-time for prediction of correctness of their next actions and, consequently, their test performance provided only statistically insignificant results [53]. However, when the response-time were associated to self-regulatory strategies, e.g., goal expectations, the result regarding the prediction of test score was significantly improved [31]. In addition, high goal-expectations exclusively do not imply high score; unless the students use the available time efficiently, they might achieve a low test score although they have been well-prepared. On the contrary, high time-management could lead to answering correctly those items that seem more clear and understandable, beyond the students' prior self-preparation to take the test. Hence, we form the following propositions:

Proposition 2: *Single conditions of self-regulation, response-time, and satisfaction from content can have opposite effects on performance, depending on how they combine with other conditions to form a solution.*

Proposition 3: *Configurations of self-regulation, response-time, and satisfaction from content for high performance are no mirror opposites of configurations for its negation (i.e., medium/low performance).*

3 RESEARCH METHODOLOGY

3.1 Research participants and data collection

Data were collected with LAERS (see sub-section 3.2) at a European University during a progress assessment test with 452 undergraduate students (211 males [46.7%] and 241 females [53.3%], aged 20-28 years old (M=21.18, SD=1.47, N=452)). The students attended the testing procedure for the Microeconomics II course (related to monopolistic competition, oligopoly, competitive strategy, and general equilibrium theory) at the University computer lab, for 75 min. each group, on April 2017.

For the assessment needs, 60 multiple choice items were used in total, distributed in 5 equivalent tests of 15 items each (some of the items were shared in more than two assessment tests). Each item had two to four possible answers, but only one was the correct, and corresponded to one of the first five levels of the factual, conceptual and procedural domains of the knowledge dimension according to the revised Bloom's taxonomy. We chose to evaluate students' knowledge by using items of these five levels only, due to the available time of the assessment test.

Before taking the assessment and right after the completion of the procedure, each participant had to answer to a pre-test and a post-test questionnaire that measure each student's goal expectancy and time-management, and their perceived clarity of the items' content respectively. The participation to the procedure was mandatory. All participants signed an informed consent form prior to their participation, explaining to them the procedure and giving the right to researchers to use the data collected for research purposes. Students were aware that their answers were being tracked, but not their time-spent, because we wanted them to act spontaneously.

3.2 The Learning Analytics & Educational Recommender System (LAERS)

Data were collected using the Learning Analytics and Educational Recommender System (LAERS) [30], a web-based implementation of a layered architecture for testing systems. The default version of LAERS consists of (a) a testing interface, (b) a tracker that logs the students' interaction data, and (c) a database storing information about the students and the items.

The testing interface (Figure 2) displays the test items delivered to students separately and one-by-one. The interface delivers the items to the students in predetermined order, and it allows them to temporarily save their answers, to review them, to alter their initial answer choices, and to save new answers. Students can also skip an item (because they are not sure about the answer, or because they think it is too difficult), and answer it (or not) later. They finalize and submit their answers only once, whenever they are ready to do so, within the test duration.

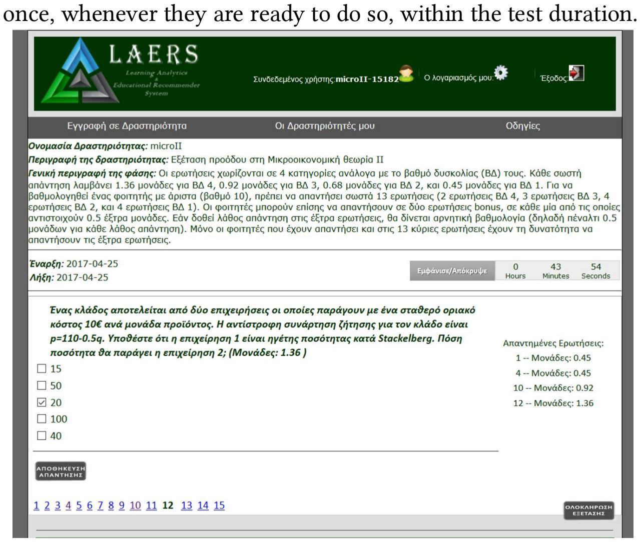

Figure 2: The testing interface in LAERS.

The tracker aggregates in log files the students' response-time, breaking it into the time on correctly and time on wrongly answered items. It also computes how many times the students review each item, how many times they change the answers and the respective time intervals. The system also calculates the test score (TS) for each student according to the correctness (0/1) of the student's answer on item *i*, and to the difficulty of the item.

Finally, a pre-test and a post-test questionnaire are embedded into LAERS to measure students' goal expectancy, time-management, and their perceived clarity of the items.

3.3 Measures

As stated in section 2.3, the identified set of factors to be included in the conceptual model consists of actual variables (i.e., response-time) and latent behavioral factors and perceptions (i.e., self-regulation and satisfaction from content). The targeted outcome of interest is the students' test score, calculated for each student as: $TS = \sum_{i=1}^{N} d_i z_i$, where $z_i \in (0,1)$ is the correctness of the student's answer on item *i*, and d_i is the difficulty of the item.

More precisely, in this study, response-time are by definition the respective time-spent the students constantly aggregate on answering the assessment questions, and 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 test items and submitting correct and wrong answers respectively.

Furthermore, for grading self-regulation, two latent factors were measured via pre-test questionnaires, i.e., goal expectancy and time-management. Goal expectancy (GE) [46] reflects the students' dispositions regarding their achievement expectations from the assessment and has two dimensions: (a) students' perception of preparation for the assessment and (b) their desirable level of success. Moreover, time-management (TM) [26] reflects students' perception of their own planning abilities and it is associated with students' exercising conscious control over the amount of time spent on items during assessment.

For measuring students' satisfaction from content, the latent factor for perceived clarity of items (CONT) [46] was measured via post-test questionnaire. CONT stores information related to whether the students considered the items to be clear, understandable and relative to the course's content, allowing students to evaluate the quality of the items and to self-reflect on their understandings of this content.

All items from the questionnaires are measured in a 7 point Likert-like scale (1 = not at all to 7 = very much, Table 1).

Table 1. Constructs and items from the questionnaires

Construct	Items	Description
Time Management (TM)	TM1	I spend more time than I want trying to find things.
	TM2	I use goal setting to determine my most important activities.
	TM3	I put off tasks that are difficult or I don't like.
Goal Expectancy (GE)	GE1	Courses' preparation was sufficient for the test.
	GE2	My personal preparation for the test was sufficient.
	GE3	My performance expectations for the test.
Clarity of Content (CONT)	CONT1	Questions were clear and understandable.
	CONT2	Questions were relative with the syllabus.
	CONT3	Questions were suitable for measuring my understanding of the course's concepts.

Table 2 summarizes all factors included in the conceptual model along with a short description, their type and value range.

3.4 FsQCA

3.4.1 Data calibration. FsQCA analysis was performed based on Pappas et al. [35]. When performing fsQCA, the researcher starts by defining the outcome of interest and the independent measures. Next all measures must be recoded into fuzzy sets, that is receiving values from 0 to 1. This process is called data calibration, and defines the extent to which cases are members of a certain group (or set) [37]. Every case of a dataset has a distinct place as determined by its fuzzy-set membership. A value of 1 means that a case is a full member of a set, and a value of 0 means that a case is fully non-member of the set. A value of 0.5 is exactly in the middle, thus a case is both a member and a non-member of the set, creating the intermediate set membership.

Table 2. List of factors considered in the conceptual model

Factor	Description	Type	Value
Pre-test (Self-regulation)			
Time management **(TM)**	perception of exercising conscious control over the amount of time spent	Latent – measured via questionnaire	1-7
Goal expectancy **(GE)**	perception of preparation and motivation to succeed	Latent – measured via questionnaire	1-7
During test (response-time)			
Total time to answer correctly **(TTAC)**	response-time aggregated on submitting correct answers	Simple – computed from actual data	≥0 (msec)
Total time to answer wrongly **(TTAW)**	response-time aggregated on submitting wrong answers	Simple – computed from actual data	≥0 (msec)
Post-test (satisfaction from content)			
Perceived clarity of the test items **(CONT)**	perception about the clearness of test items	Latent – measured via questionnaire	1-7
Test Score **(TS)**	The test result	Computed	1-10

Data calibration can be done either directly or indirectly. Direct calibration means that three qualitative thresholds need to be chosen, which define the level of membership in the fuzzy set for every case. On the other hand, indirect calibration means that the measurements need to be rescaled following qualitative assessments. Either method can be followed, as it depends on one's substantive knowledge of the data and the underlying theory [37,39]. Data calibration is critical, because disparities in the calibration may lead to disparities in the outcome, thus cases in the dataset should be transformed into membership scores following a well-documented and qualitatively justified manner. The direct method of setting three values, corresponding to full-set membership, full-set non-membership, and intermediate-set membership is recommended [37].

Next, the question is how to choose the three thresholds. The simplest way is to choose the values of 1, 0.5, and 0. For instance, in a 7-point Likert scale, the values 7, 4, and 1 would be calibrated into 1, 0.5, and 0 respectively, with the rest (6, 5, 3, 2) following accordingly. For 7-point Likert scales multiple studies suggest that the values of 6, 4, and 2 should be used as thresholds [17,28,35]. Also, measures can be calibrated by using percentiles. In this case, the following percentiles can be set as the full-set membership, intermediate-set membership, and full-set non-membership, 80%, 50%, and 20%, respectively[35]. However, as it is up to the researcher to choose the three thresholds, these values can be changed accordingly. In this study, since data are skewed to the right, data calibration is done based on percentiles. Calibration based on the survey scale might lead to less meaningful results, producing a single solution with all the conditions identified as necessary [32,35].

Once running the analysis, fsQCA creates a truth table of 2^k rows, where k represents the number of outcome predictors (i.e., independent variables) and each row represents every possible combination. For instance, a truth table between five variables (i.e., conditions) would provide thirty-two possible logical combinations. For each combination, fsQCA computes the minimum membership value (i.e., the degree to which a case supports the specific combination). The threshold of 0.5 is used to identify the combinations that are acceptably supported by the cases. Thus, all combinations that are not supported by at least one case with membership larger than the threshold of 0.5 are automatically removed from further analysis.

Next, the truth table must be sorted based on frequency and consistency [37]. Frequency refers to the number of observations for each possible combination, and consistency refers to "the degree to which cases correspond to the set-theoretic relationships expressed in a solution" [13]. Since fsQCA computes all logical combinations, many combinations will have a frequency of zero. It is important to set a frequency cut-off point which will ensure that a minimum number of empirical observations is obtained for the assessment of subset relationships. A higher frequency threshold means that every combination will refer to more cases in the sample, but it will reduce the percentage (i.e., coverage) of the sample explained by the solutions. On the other hand, a small frequency threshold will increase the coverage of the sample, although each combination will refer to fewer cases in the sample. For small and medium-sized samples, a cut-off point of 1 is appropriate, but for larger samples (e.g., 150 or more cases), the cut-off point should be set higher [37]. The researcher can decide if a larger cut-off point should be set for very large datasets. Low frequency combinations are removed from further analysis and the truth table must be sorted based on "raw consistency."

A consistency threshold needs to be set, with the minimum recommended value being 0.75 [39]. A good indication for choosing this threshold is to identify big changes in the consistency of each combination. For instance, a combination may have a consistency of 0.82 and the next may have 0.79. Although both values are above the recommended threshold of 0.75, this is an indication of what the consistency threshold should be. In any case, it is up to the researcher to choose the exact threshold. A low consistency threshold may produce more necessary conditions, reducing type II errors (i.e., false negatives), but increasing type I errors (i.e., false positives), and vice versa [11]. The last step is to insert the value of 1 or 0 in the column with the outcome variable, depending on the consistency threshold that has been chosen. Combinations with consistency higher than the threshold will get the value of 1, otherwise, 0.

3.4.2 Obtain the solution sets. Following the sorting of the truth table, fsQCA computes the following three sets of solutions: complex, parsimonious, and intermediate; "solution" is a combination of conditions that is supported by a high number of cases, and the rule "the combination leads to the outcome" is consistent. The *complex* solution presents all possible combinations of conditions when traditional logical operations are applied. The number of complex solutions can be large, including configurations with several terms, their interpretation is difficult and often impractical [27]. Thus, they are simplified automatically into *parsimonious* and *intermediate* solutions.

The *parsimonious* solution is a simplified version of the complex solution and presents the most important conditions that cannot be left out from any solution. These are called "core conditions" [13] and are identified automatically by fsQCA. Finally, the *intermediate* solution is computed when performing counterfactual analysis on the complex and parsimonious solutions [35,37]. FsQCA uses simplifying assumptions to compute the parsimonious and intermediate solutions, and if needed the researcher may employ more assumptions, regarding the connection between each causal condition and the outcome, based on theoretical or substantive knowledge [13,38]. The intermediate solution is a part of the complex solution and includes the parsimonious solution. Conditions that are part of the intermediate solution but of the parsimonious solution are called "peripheral conditions" [13]. A more detailed description of the steps in counterfactual analysis is provided by [27].

3.4.3 Interpretation of the solutions. FsQCA computes the complex and parsimonious solutions regardless of any simplifying assumptions employed by the researcher, while the intermediate solution depends directly on these assumptions. For better interpreting the results, combining the parsimonious and intermediate solutions is recommended. A table that will include both core and peripheral conditions should be created[13,35]. To do this, the researcher should identify the conditions of the parsimonious solution in the intermediate solution. This leads to a combined solution, which will include all core and peripheral conditions, thus helping in the interpretation of the findings. Further, to improve the visualization of the results, the presence of a condition is presented with a black circle (●), the absence with a crossed-out circle (⊗), and the "do not care" condition with a blank space. The distinction between core and peripheral is done by using large and small circles, respectively. The overall solution consistency and the overall solution coverage are presented. Consistency measures the degree to which a subset relationship has been approximated, and overall coverage describes the extent to which the outcome is be explained by the different configurations, and is comparable with the R-square reported on regression-based methods [39,52].

4 RESULTS

The results for high performance and medium/low performance are shown in Tables 3 and 4 respectively. Both tables also present consistency values for the overall solution and for each solution separately. All values are higher than the recommended threshold (> 0.75) [37]. An overall solution coverage of .77 and .79 suggests that the five solutions account for a substantial proportion of high performance and medium/low performance respectively. FsQCA also estimates the empirical relevance of every solution, by calculating raw and unique coverage. The raw coverage describes the amount of the outcome explained by a specific alternative solution, while the unique coverage describes the amount of the outcome that is exclusively explained by a specific alternative solution. Solutions for high performance explain a large amount of users' performance, ranging from 22% to 40% of cases associated with the outcome (Table 3). Similarly, solutions for medium/low performance explain a vast amount of the absence of performance, ranging from 22% to 58% of cases associated with the outcome (Table 4).

For high performance (Table 3), solutions A1-A5 present combinations for which the different factors may be present or absent depending on how they combine with each other.

- *Solution A1*: Students achieved high performance when they did not spend a lot of time to answer the questions, either the answers were correct or wrong, but they did good time management. This solution explains the behaviour of 22% of the high performing students.
- *Solutions A2, A3 and A4*: These solutions show 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 as it shows that students who give all their focus only in finding the correct answer will achieve high performance. However, it is interesting to note that this happens when students were neither sufficiently prepared for the progress assessment nor they believe that they can manage their time properly (solution A2). Also, this happens when students have good time management skills, but they had not understood the questions very well (solution A3). In fact, according to the results, solution A2 explains 28% of the high performing students, whereas solution A3 explains a bigger sub-population of the high performing students, since row coverage is 36%. Nonetheless, spending enough time, to find the correct answer will lead to high performance even if the questions are unclear or not so relative to the syllabus. On the other hand, if the students understand well the questions, then they can achieve high test score, even when spending a lot of time both on the questions answered wrongly and correctly (Solution A4). This solution explains 26% of the high performing cases.
- *Solution* A5: Finally, the students can achieve high performance regardless of how much time they spend to answer the questions: they have set high goal expectations, they have high time management skills, and they have a good understanding of the questions. This solution explains the larger part of the cases of high performing students (40%).

Next, Table 4 presents the solutions for not achieving a high performance, that is achieving medium/low performance. The findings show that the solutions that explain medium/low performance are not perfect opposites of the solutions that explain high performance. Specifically:

- *Solutions B1 and B2*: Students that do not have high goal expectancy will have a low or medium performance, when they do not spend a lot of time to answer the questions (either correctly or wrongly) (Solution B1), or when they have low time management skills, leading them to not using efficiently the available time (Solution B2). These solutions explain 34% and 58% of the cases of medium/low performing students.
- *Solution B3*: Students that have spent a lot of time to answer the questions wrongly, and they are not well prepared, they

will have a medium or low performance, even if they have understood the questions. This behaviour is observed in 33% of the cases of medium/low performing students.

- *Solution B4*: Students that give correct answers fast, but spend a lot of time to questions that they do not know the answer, will have a medium or low performance even if they have high goal expectancy and believe that they can use their time properly. This solution explains 22% of the cases for medium/low performing students.
- *Solution B5*: Students who perceive the questions as clear and relative to the syllabus, and spent a lot of time in answering them (both correctly and wrongly), will not achieve a high performance unless they have a good time management. This finding highlights the importance of time management in high performances. However, this behaviour for medium/low performing students is not very common (17% of the cases).

Table 3. Configurations for high performance

Solutions for high performance						
Configuration		A1	A2	A3	A4	A5
Response-time	Time to answer correctly	⊗	●	●	●	
	Time to answer wrongly	⊗			•	
Self-regulation	Goal Expectancy		⊗			●
	Time Management	●	⊗	•		•
Satisfaction	Clarity of Content			⊗	•	●
Consistency		.82	.86	.90	.88	.88
Raw Coverage		.22	.28	.36	.26	.40
Unique Coverage		.04	.08	.01	.01	.01
Overall Solution Consistency		.81				
Overall Solution Coverage		.77				

Note: Black circles (●) indicate the presence of a condition, and circles with "x" (⊗) indicate its absence. All circles indicate core conditions. Blank spaces indicate don't care conditions.

The results provide support for all three propositions. In detail, multiple configurations lead to high performance, verifying equifinality (Proposition 1). Also, the results provide configurations that explain performance in which conditions may be either present or absent, depending on how the combine with each other, verifying the existence of causal asymmetry (Proposition 2). Finally, the findings support proposition 3, i.e., configurations that explain high performance are not the exact opposites of those explaining medium/low performance.

5 DISCUSSION AND CONCLUSIONS

Understanding the factors that affect assessment outcome, as well as their interrelationships, could contribute to the sufficient explanation and interpretation of the obtained high or medium/low performance. Previous research identified critical factors that affect the performance (e.g., response-time, self-regulatory strategies, non-cognitive perceptions of satisfaction from the content), but failed to reveal asymmetric relationships between these factors, mostly due to the variance-based analysis methods they employed for exploring the data [20,31,36]. This study focuses on compiling students' response-time allocated to answer correctly or wrongly, their self-regulation, as well as their satisfaction from the content, targeting at explaining high or medium/low performance achieved on the assessment test. For this purpose, fuzzy set qualitative comparative analysis (fsQCA) was applied for exploring multiple configurations of causal conditions which may include different combinations of goal-expectations, time-management, response-time and clarity of content. Data were collected during a progress assessment test with 452 undergraduate students from a European University. The results provided several interesting findings.

Table 4. Configurations for medium/low performance

Solutions for medium/low performance						
Configuration		B1	B2	B3	B4	B5
Response-time	Time to answer correctly	⊗	⊗		⊗	•
	Time to answer wrongly	⊗		●	●	●
Self-regulation	Goal Expectancy	⊗	⊗	⊗	•	
	Time Management		⊗		•	⊗
Satisfaction	Clarity of Content			●		●
Consistency		.82	.85	.85	.89	.85
Raw Coverage		.34	.58	.33	.22	.17
Unique Coverage		.03	.16	.02	.07	.01
Overall Solution Consistency		.81				
Overall Solution Coverage		.79				

Note: Black circles (●) indicate the presence of a condition, and circles with "x" (⊗) indicate its absence. All circles indicate core conditions. Blank spaces indicate don't care conditions.

Firstly, as seen from Table 3, highly performing students who believe that they have good time-management skills, ended up spending little time in giving answers, and not spending their time in a meaningless way (solution A1). This finding is in agreement with [26], who supported that high achieving students often exhibit strong time-management skills. Moreover, according to solutions A2, A3, and A4, students who aggregated non-neglectable response-time for correct answers, although they were neither sufficiently prepared for the progress assessment nor they managed their time efficiently (solution A2), however, they finally achieved a high score in the test. This finding contradicts with [31] which identified that poorly-prepared students (i.e., scoring low in goal-expectations) achieve low scores, and indicates that regardless of preparation and time-management, the students may still get high grades if they engage more on answering the questions. This contradiction, however, might be due to the fact that in [31] the authors investigated only symmetric solutions. Thus, the current approach sheds more light into the interrelationships between preparation, response-time and test-score. This also happens when students did a good time management but they had not understood all questions very well (solution A3). This means that probably, these students managed their time efficiently in order to answer correctly on those items that we clear to them, and although they struggled to understand the rest of the items, they finally delivered more correct answers. Nonetheless, spending enough time, in an appropriate manner, to find the correct solution will lead to high performance even if the questions are unclear or not so relative to the syllabus.

On the other hand, if the students understand well the questions they can achieve high performance even when spending a lot of time for all questions (both the ones answered wrongly and correctly) (Solution A4). This is an expected finding, since perceiving the test content as comprehensible does not necessarily mean that it is trivial or easy to answer, and as such, wrong answers are likely to occur, but they are not the dominant ones, leading to an overall high performance.

These two findings are interesting and innovative in terms of the rather limited literature on the issue of the effects of time-management along with content comprehensibility on the assessment test outcome. To the best of our knowledge, this is the first study to explore this combinational/conditional interrelationship, and extends previous work [25], which focuses solely on the effect of content on performance. It should be noted that spending a lot of time to answer correctly is a present as a core factor in these solutions, highlighting its importance in achieving high performance. This is in full agreement with [31] and provides additional evidence regarding the role of aggregated response-time to interpreting performance.

Finally, solution A5 verifies that well-prepared students are quite likely to perform well and get high scores, regardless of how much time they need to answer any of questions [15].

Regarding the results for medium/low performance, solution B1 and B2 claim that poorly prepared students answer the questions relatively quick are not expected to have a high performance, which seems intuitive and further complies with the literature [12]. However, comparing these findings with solution A1 from Table 3, highlights how important is time management in achieving high performance.

Moreover, solution B3 describes students that had a good understanding of the questions, but because they were not well prepared they did not know the answers, thus making them to use a lot of time on answering wrongly, leading to medium or low performance. There are two interesting clues about this finding: (a) it is in contrast to previous results that comprehensibility of content is directly reflected on performance [49], and (b) it is surprising from a slightly different point of view: as seen from table 4, this solution explains 58% of the cases of medium/low performing students. The surprising thing is that, these students admitted that the test questions were clear, comprehensible and related to the course's content, yet they were neither prepared to answer them, nor they tried to guess the answers. It would be really valuable to explore a measure of guessing regarding this sub-group of medium/low performing students in order to identify/evaluate their guessing intentions.

Finally, one of the most important implications of this paper is related to how learning analytics researchers and practitioners can utilize the fsQCA method to make sense of diverse analytics and take design decisions for various user groups [33,40]. Future studies should combine fsQCA with variance-based techniques to gain a deeper insight into the learning analytics, and combine both methods towards extending current theories and practices as well as developing new ones. As this study is among the first to employ fsQCA in learning analytics context [34,35], further innovative research is needed to identify complex and important configurations that reveal the full potential of this analysis. Future studies should incorporate data from various learning activities and modalities, making-sense of complex learning interactions and offering a holistic understanding of the potential of this data analysis technique in technology enhanced learning and analytics.

ACKNOWLEDGMENTS

This work received funding from the Research Council of Norway under the project FUTURE LEARNING (255129/H20), and the European Union's Horizon 2020 research and innovation programme, under the Marie Skłodowska-Curie grant agreement No. 751550.

REFERENCES

[1] I J M Arnold. 2016. Cheating at online formative tests: Does it pay off? *Internet High Educ.* 29: 98–106. https://doi.org/10.1016/j.iheduc.2016.02.001
[2] J E Beck. 2005. Engagement Tracing: Using response times to model student disengagement. *In Proceedings of the 2005 Conference on Artificial Intelligence in Education: Supporting Learning Through Intelligent and Socially Informed Technology*, 88–95.
[3] J Broadbent and W L Poon. 2015. Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *Internet High Educ.* 27: 1–13. https://doi.org/10.1016/j.iheduc.2015.04.007
[4] J Burrus, T Jackson, S Holtzman, R D Roberts, and T Mandigo. 2013. Examining the efficacy of a time management intervention for high school studetns.. *ETS Research Report Series* 2013, 2: i--35. https://doi.org/10.1002/j.2333-8504.2013.tb02332.x
[5] Di Challis. 2005. Committing to quality learning through adaptive online assessment. *Assessment & Evaluation in Higher Education* 30, 5: 519–527. https://doi.org/10.1080/02602930500187030
[6] S-R Chang, B S Plake, G A Kramer, and S-M Lien. 2011. Development and application of detection indices for measuring guessing behaviors and test-taking effort in computerized adaptive testing. *Educational and Psychological Measurement* 71, 3: 437–459. https://doi.org/10.1177/0013164410385110
[7] Y P Chua. 2012. Effects of computer-based testing on test performance and testing motivation. *Computers in Human Behavior* 28, 5: 1580–1586. https://doi.org/10.1016/j.chb.2012.03.020
[8] A Davis. 1999. *The Limits of Educational Assessment.* Hoboken, NJ: Wiley-Blackwell.
[9] W J Doll and G Torkzadeh. 1988. The Measurement of End-user Computing Satisfaction. *MIS Q.* 12, 2: 259–274. https://doi.org/10.2307/248851
[10] D H Doty, W H Glick, and G P Huber. 1993. Fit, equifinality, and organizational effectiveness: A test of two configurational theories. *Academy of Management journal* 36, 6: 1196–1250.
[11] J Dul. 2016. Identifying single necessary conditions with NCA and fsQCA. *Journal of Business Research* 69, 4: 1516–1523.

[12] B Eilam and I Aharon. 2003. Students' planning in the process of self-regulated learning. *Contemporary Educational Psychology* 28, 3: 304–334. https://doi.org/10.1016/S0361-476X(02)00042-5

[13] P C Fiss. 2011. Building better causal theories: A fuzzy set approach to typologies in organization research. *Academy of Management Journal* 54, 2: 393–420.

[14] A R Fitzpatrick. 1983. The Meaning of Content Validity. *Applied Psychological Measurement* 7, 1: 3–13. https://doi.org/10.1177/014662168300700102

[15] C B Hodges and C Kim. 2010. Email, self-regulation, self-efficacy, and achievement in a college online mathematics course. *Journal of Educational Computing Research* 43, 2: 207–223. https://doi.org/10.2190/EC.43.2.d

[16] L F Hornke. 2000. Item response times in computerized adaptive testing. *Psicológica* 21, 1–2: 175–189.

[17] M N Giannakos, I O Pappas, P Mikalef and P A Pavlou. 2017. Value co-creation and trust in social commerce: An fsQCA approach. In *Proceedings of the 25th European Conference on Information Systems (ECIS)*, 2153–2168.

[18] N Kahraman, M M Cuddy, and B E Clauser. 2013. Modeling pacing behavior and test speededness using latent growth curve models. *Applied Psychological Measurement* 37, 5: 343–360. https://doi.org/10.1177/0146621613477236

[19] N Kim, M. J Smith, and K Maeng. 2008. Assessment in online distance education: a comparison of three online programs at a university. *Online Journal of Distance Learning Administration* 11, 1.

[20] A Kitsantas. 2002. Test Preparation and Performance: A Self-Regulatory Analysis. *The Journal of Experimental Education* 70, 2: 101–113. https://doi.org/10.1080/00220970209599501

[21] M Kurucay and F A Inan. 2017. Examining the effects of learner-learner interactions on satisfaction and learning in an online undergraduate course. *Comput Educ* 115, C: 20–37. https://doi.org/10.1016/j.compedu.2017.06.010

[22] Y-H Lee and S J Haberman. 2016. Investigating test-taking behaviors using timing and process data. *International Journal of Testing* 16, 3: 240–267. https://doi.org/10.1080/15305058.2015.1085385

[23] C H Lee Y.-H. 2011. A review of recent response-time analyses in educational testing. *Psychological Test and Assessment Modeling* 53, 3: 359-379.

[24] W J van der Linden. 2009. A bivariate lognormal response-time model for the detection of collusion between test takers. *Journal of Educational and Behavioral Statistics* 34, 3: 378–394. https://doi.org/10.3102/1076998609332107

[25] C C Lo. 2010. How student satisfaction factors affect perceived learning. *Journal of Scholarship of Teaching and Learning* 10, 1: 47–54.

[26] S C Dipboye, R L Phillips, A P Macan, Therese H. 1990. College students' time management: Correlations with academic performance and stress. *Journal of Educational Psychology* 82, 4: 760-768.

[27] J M Mendel and M M Korjani. 2012. Charles Ragin's fuzzy set qualitative comparative analysis (fsQCA) used for linguistic summarizations. *Information Sciences* 202: 1–23.

[28] A Ordanini, A Parasuraman, and G Rubera. 2014. When the recipe is more important than the ingredients a Qualitative Comparative Analysis (QCA) of service innovation configurations. *Journal of Service Research* 17, 2: 134–149. https://doi.org/10.1177/1094670513513337

[29] Z Papamitsiou and A A Economides. 2017. Exhibiting achievement behavior during computer-based testing: What temporal trace data and personality traits tell us? *Computers in Human Behavior* 75: 423–438. https://doi.org/10.1016/j.chb.2017.05.036

[30] Z Papamitsiou and A A. Economides. 2013. Towards the alignment of computer-based assessment outcome with learning goals: The LAERS architecture. In *2013 IEEE Conference on e-Learning, e-Management and e-Services, IC3e 2013*, 13–17. https://doi.org/10.1109/IC3e.2013.6735958

[31] Z K. Papamitsiou, V Terzis, and A A. Economides. 2014. Temporal learning analytics for computer based testing. In *Proceedins of the Fourth International Conference on Learning Analytics And Knowledge - LAK '14*, 31–35. https://doi.org/10.1145/2567574.2567609

[32] I O Pappas, S Papavlasopoulou, M N Giannakos, and D G Sampson. 2017. An Exploratory Study on the Influence of Cognitive and Affective Characteristics in Programming-Based Making Activities. In *2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT)*, 507–511. https://doi.org/10.1109/ICALT.2017.78

[33] I O Pappas, M N Giannakos, L Jaccheri, and D G Sampson. 2017. Assessing Student Behavior in Computer Science Education with an fsQCA Approach: The Role of Gains and Barriers. *ACM Transactions on Computing Education (TOCE)* 17, 2: Article No. 10.

[34] I O Pappas, M N Giannakos, and D G Sampson. 2016. Making sense of learning analytics with a configurational approach. In *Proceedings of the workshop on Smart Environments and Analytics in Video-Based Learning (SE@ VBL)*, 42–52.

[35] I O Pappas, M N Giannakos, and D G Sampson. 2017. 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. *Computers in Human Behavior.* https://doi.org/10.1016/j.chb.2017.10.010

[36] M Puzziferro. 2008. Online technologies self-efficacy and self-regulated learning aspredictors of final grade and satisfaction in college-level online courses. *American Journal of Distance Education* 22, 2: 72–89. https://doi.org/10.1080/08923640802039024

[37] C C Ragin. 2008. *Redesigning social inquiry: Fuzzy sets and beyond.* Wiley Online Library.

[38] C C Ragin, K A Drass, and S Davey. 2006. Fuzzy-set/qualitative comparative analysis 2.0. *Tucson, Arizona: Department of Sociology, University of Arizona.*

[39] B Rihoux and C C Ragin. 2009. *Configurational comparative methods: Qualitative comparative analysis (QCA) and related techniques.* Sage Publications, Thousand Oaks, CA.

[40] S Sergis, D G Sampson, and M Giannakos. 2017. Enhancing student digital skills: Adopting an ecosystemic school analytics approach. In *2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT)*, 21–25. https://doi.org/10.1109/ICALT.2017.87

[41] J C Setzer, S L Wise, J R van den Heuvel, and G Ling. 2013. An Investigation of examinee test-taking effort on a large-scale assessment. *Applied Measurement in Education* 26, 1: 34–49. https://doi.org/10.1080/08957347.2013.739453

[42] D Y Shee and Y-S Wang. 2008. Multi-criteria evaluation of the web-based e-learning system: A methodology based on learner satisfaction and its applications. *Comput. Educ.* 50, 3: 894–905. https://doi.org/10.1016/j.compedu.2006.09.005

[43] P-C Sun, R J Tsai, G Finger, Y-Y Chen, and D Yeh. 2008. What drives a successful e-Learning? An empirical investigation of the critical factors influencing learner satisfaction. *Comput. Educ.* 50, 4: 1183–1202. https://doi.org/10.1016/j.compedu.2006.11.007

[44] D L Sundre and A Kitsantas. 2004. An exploration of the psychology of the examinee: Can examinee self-regulation and test-taking motivation predict consequential and non-consequential test performance? *Contemporary Educational Psychology* 29, 1: 6–26. https://doi.org/10.1016/S0361-476X(02)00063-2

[45] F Tabak, N Nguyen, T Basuray, and W Darrow. 2009. Exploring the impact of personality on performance: How time-on-task moderates the mediation by self-efficacy. *Personality and Individual Differences* 47, 8: 823–828. https://doi.org/10.1016/j.paid.2009.06.027

[46] V Terzis and A A Economides. 2011. The acceptance and use of computer based assessment. *Comput. Educ.* 56, 4: 1032–1044. https://doi.org/10.1016/j.compedu.2010.11.017

[47] H Wainer. 2000. *Computerized adaptive testing: A Primer (2nd Edition).* Mahwah, NJ: ELawrence Erlbaum Associates.

[48] T Wang and B A Hanson. 2005. Development and calibration of an item response model that incorporates response time. *Applied Psychological Measurement* 29, 5: 323–339. https://doi.org/10.1177/0146621605275984

[49] Y-S Wang. 2003. Assessment of learner satisfaction with asynchronous electronic learning systems. *Information & Management* 41, 1: 75–86. https://doi.org/10.1016/S0378-7206(03)00028-4

[50] S L Wise and X Kong. 2005. Response Time Effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education* 18, 2: 163–183. https://doi.org/10.1207/s15324818ame1802_2

[51] T Wolsey. 2008. Efficacy of instructor feedback on written work in an online program. *International Journal on E-Learning* 7, 2: 311–329.

[52] A G Woodside. 2013. 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. *Journal of Business Research* 66, 4: 463–472.

[53] X Xiong. Z Pardos & N Heffernan. 2011. An analysis of response time data for improving student performance prediction. In In KDD 2011 Workshop: Knowledge Discovery in Educational Data, Held as part of 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining.

[54] B J Zimmerman and M Martinez Pons. 1986. Development of a structured interview for assessing student use of self-regulated learning strategies. *American Educational Research Journal* 23, 4: 614–628. https://doi.org/10.3102/00028312023004614