Received: 2 January 2020 | Revised: 11 June 2020 | Accepted: 13 June 2020
DOI: 10.1111/jcal.12472

ARTICLE

Journal of Computer Assisted Learning WILEY

The impact of on-demand metacognitive help on effortful behaviour: A longitudinal study using task-related visual analytics

Zacharoula Papamitsiou[1,2] | Anastasios A. Economides[1]

[1]IPPS in Information Systems, University of Macedonia, Thessaloniki, Greece

[2]Sem Sælands vei 9, IT-Bygget, 7491, Norwegian University of Science and Technology, Trondheim, Norway

Correspondence
Zacharoula Papamitsiou, Department of Computer Science, Norwegian University of Science and Technology, Sem Sælands vei 9, 7491, Trondheim, Norway.
Email: zacharoula.papamitsiou@ntnu.no; papamitsiou.xara@gmail.com

Present address
Anastasios A. Economides, Sem Sælands vei 9, IT-Bygget, 7491, Norwegian University of Science and Technology, Trondheim, Norway

Peer Review
The peer review history for this article is available at https://publons.com/publon/10.1111/jcal.12472.

Abstract

This longitudinal study investigates the differences in learners' effortful behaviour over time due to receiving metacognitive help—in the form of on-demand task-related visual analytics. Specifically, learners' interactions (N = 67) with the tasks were tracked during four self-assessment activities, conducted at four discrete points in time, over a period of 8 weeks. The considered and coded time points were: (a) prior to providing the metacognitive help; (b) while the task-related visual analytics were available (treatment); (c) after the removal of the treatment; and (d) while the option to receive metacognitive help was available again. To measure learners' effortful behaviour across the self-assessment activities, this study utilized learners' response-times to correctly/wrongly complete the tasks and on-task effort expenditure. The panel data analysis shown that the usage of metacognitive help caused statistically significant changes in learners' effortful behaviour, mostly in the third and fourth phase. Statistically significant changes were detected also in the usage of metacognitive help. These results provide empirical evidence on the benefits of task-related visual analytics to support learners' on-task engagement, and suggest relevant cues on how metacognitive help could be designed and prompted by focusing on the "task", instead of the "self".

KEYWORDS

effortful behaviour, learning analytics, longitudinal study, metacognitive help, self-assessment, task-related visual analytics

1 | THE IMPACT OF ON-DEMAND METACOGNITIVE HELP ON EFFORTFUL BEHAVIOUR: A LONGITUDINAL STUDY USING TASK-RELATED VISUAL ANALYTICS

Supporting learners during their learning by providing feedback is an important part of the process (Richardson, Abraham, & Bond, 2012), as it can contribute to sustaining learners' self-regulation and goal attainment (Pintrich, 2004; Zimmerman, 1990). However, it might not impact learning as expected, unless the learners are willing to use it (Hattie & Timperley, 2007). It is likely that learners will use the provided support when they request it, that is, when they exhibit help-seeking behaviour. Help-seeking is "a behaviour performed by individuals who perceive themselves as needing assistance with a problem, whereby the intended outcome of this behaviour is addressing the problem faced" (Heerde & Hemphill, 2018, p. 2). Help-seeking is a self-directed, pro-active strategy: the learners concede the need for help, define help-seeking goals, estimate the benefits and costs of (not)seeking help, initiate the communication loop, select the appropriate sources, obtain and process help (Aleven, Stahl, Schworm, Fischer, & Wallace, 2003; Karabenick & Berger, 2013; Nelson-Le Gall, 1985).

However, efficient help-seeking is threatened by two unwanted behaviours: (a) the learners might avoid request help (underuse), or (b) they might excessively ask for hints that explicitly lead to the solution (overuse), without activating deeper thinking mechanisms. For instance, learners might not be motivated to learn, and thus, it is likely that they would not engage in help-seeking (Hao, Barnes, Wright, & Branch, 2016; Huet, Moták, & Sakdavong, 2016). Some learners might also feel that their learning autonomy is threatened, as they rely on instructors' or peers' assistance, and thus, they might avoid seeking help (Fletcher & Shaw, 2012; Huet et al., 2016). Other learners might continuously seek help because they do not know when they really need it, or might engage in gaming the system behaviour—that is, "attempt to succeed in an educational environment by exploiting properties of the system rather than by learning the material and trying to use that knowledge to answer correctly" (Baker, 2005, p.3)—because they are highly performance oriented (Aleven et al., 2003).

It has been argued that both overuse and underuse of help can be detrimental to learning; help-seeking overuse is associated with poor learning because it bypasses the self-explanation and self-regulation processes (Roll, Baker, Aleven, & Koedinger, 2014). The lack of such metacognitive skills has been also associated with underuse of help facilities (Vaessen, Prins, & Jeuring, 2014). The reason is that help-seeking is associated with and involves metacognitive characteristics that include the learner's knowledge of knowledge ("Do I know enough to succeed on my own?"), regulation of knowledge ("How can I obtain additional information I may need?") (Roll, Aleven, McLaren, & Koedinger, 2011), and motivation to seek it ("Why should I ask for help? Which are the benefits/costs for me?") (Karabenick, 2011). Metacognition is related to one's ability to monitor and control one's own knowing, and comprises the executive processes of reflective judgment and regulation of one's own deeper thinking; in simple terms, it is "thinking about thinking" (Flavell, 1979). Through those processes, the learners acquire their metacognitive knowledge from metacognitive monitoring, and control their learning using the metacognitive knowledge (Nelson & Nahrens, 1990).

The above characteristics strongly converge to the existence of more complicated processes underlying help-seeking: it is a self-directed strategy that involves motivational, cognitive and metacognitive mechanisms (Mäkitalo-Siegl, Kohnle, & Fischer, 2011; Ryan & Shin, 2011; Vaessen et al., 2014). Therefore, the challenge is to design and deliver appropriate help formats, aiming at motivating learners to request help at the moment they really need it, as well as at efficiently supporting their self-regulation (Daley, Hillaire, & Sutherland, 2016; Roll et al., 2011). In fact, metacognitive prompts have been suggested as instructional help formats that are designed and intended to help learners' monitoring and control of their information processing, because they can induce metacognitive and regulative strategies, such as goal setting, planning, monitoring, control and evaluation (Bannert & Mengelkamp, 2013).

To enable learners to efficiently use the available help, they need to possess sufficient knowledge about how to use it (Stone, 2000), and it should be provided regularly during the learning tasks (Zimmerman, 1990) so that learners can practice with it (Winne, 1997). Activating learner's deeper critical thinking mechanisms, reflective judgment and regulation of motivation with appropriate help formats is expected to increase their awareness of why and when they really need help, and to improve their on-task engagement and learning performance (Kautzmann & Jaques, 2019; Labuhn, Zimmerman, & Hasselhorn, 2010; Long & Aleven, 2017; Roll et al., 2006; Tsai, Lin, Hong, & Tai, 2018; Wolters, 1998). Engaging learners in metacognitive processes is not a straightforward task, unless learners are explicitly encouraged to do so through specialized instructional activities (Daley et al., 2016; Gama, 2004; Lin, 2001).

The rapid development of different forms of visual analytics has opened new perspectives on and opportunities for the design of metacognitive help (Durall & Gros, 2014; Schwendimann et al., 2017). Specifically, learning analytics dashboards are instruments intended to increase awareness of learning goals (Corrin & de Barba, 2015; Sedrakyan, Malmberg, Verbert, Järvelä, & Kirschner, 2018), to foster self-regulation (Azevedo et al., 2017; Davis, Chen, Jivet, Hauff, & Houben, 2016), and to improve decision-making (Robert Bodily et al., 2018; Yigitbasioglu & Velcu, 2012) by capitalizing on human perceptual capabilities.

The present study focuses on encouraging learners to seek help that metacognitively assists them in their efforts to successfully complete the learning tasks. Specifically, the study investigates on-demand metacognitive help in the form of task-related visual analytics. This help format is learner-oriented, that is, it is extracted from all learners' interaction data, and is expected to provide meaningful, holistic information about how all fellow peers have treated the same learning tasks, but it is not "ego"-centred. The core idea is to extract analytics from all peers' interactions trace data, to convert them into useful indices (e.g., task difficulty, effort needed) and to deliver the processed information to learners, to improve their awareness of and critical reflection on the "real" requirements of the tasks, as they are collectively refined.

The study aims at investigating the changes in learners' effortful behaviour over a period of time due to this help format. Learners exhibit effortful behaviour when they are involved and committed to solving a problem, being actively engaged and giving their best efforts to understand and complete tasks, as opposed to effortless behaviour that is, a generalized randomness in their behaviour, reflected on "cheating," "guessing" or "gaming the system" behavioural patterns that counterfeit the learning outcomes (Baker, Corbett, Koedinger, & Wagner, 2004; Sharma, Papamitsiou, & Giannakos, 2019; Sharma et al., 2020; Wise & Kong, 2005). In a way, effortful behaviour is not disconnected from "effortful control"—that is, "the ability to inhibit a dominant response to perform a subdominant response" (Rothbart & Bates, 1998, p. 137)—but it is more focused on engagement aspects of behaviour and attention regulation; learners' on-task effortful behaviour synopsizes on-task behavioural engagement. Interdisciplinary researchers in learning settings use the term “engagement” to refer to students' involvement, effort and time investment, as well as persistence in learning (Fredricks, Blumenfeld, & Paris, 2004; Lawson & Lawson, 2013). Engagement has been widely conceptualized as a multidimensional construct, involving the individual's ability to implicate in on-going learning processes, depicted through actual interaction between engagement objects and subjects (Ben-Eliyahu, Moore, Dorph, & Schunn, 2018; Fredricks et al., 2004; Lawson & Lawson, 2013). Researchers who work on the topic have classified the related factors (e.g., participation; attention; degree of interaction; response-times) into three generic dimensions of engagement, that is, thoughts (e.g., Greene, 2015; Sun & Rueda, 2012), feelings (Molinillo, Aguilar-Illescas, Anaya-Sánchez, & Vallespín-Arán, 2018; Sun & Rueda, 2012) and behaviours (Sun & Rueda, 2012; Sunar, White, Abdullah, & Davis, 2017). This study focuses on learners' behavioural patterns of effortful behaviour, as an actual interaction.

Specifically, the rationale of this study and the motivation to conduct this research is to explore the differences in the same subjects' effortful behaviour prior to using the metacognitive help, during taking this help, after removing it, and finally, after providing it again. Accordingly, the contribution of this work is two-fold:

a this longitudinal study sheds light to how the task-related metacognitive information affects learners' on-task effortful behaviour, based on their manipulations of tasks, and
b how learners' behaviour changes over time (if it does) due to the metacognitive help usage.

In that sense, this work does not focus on boosting learners' metacognition about their own help-seeking behaviour, but on bringing understanding on how visual analytics can promote the acquisition of metacognitive skills (e.g., analytics interpretation and time-management, or other analytical skills such as comparing and correlating information) in learning.

2 | RELATED WORK

Contemporary digital learning environments offer numerous options for seeking help. Typical help implementations include worked-out problems, glossaries or detailed solutions of the on-going problem (Huet, Escribe, Dupeyrat, & Sakdavong, 2011), hints on the steps required to solve a problem, asking the cognitive tutor to complete the exercise (Vaessen et al., 2014), explanations on errors, instructions for solving the problem, videos demonstrating the solutions (Huet et al., 2016), asking teachers/peers for online help, and online searching (Hao, Wright, Barnes, & Branch, 2016). Help-seeking is usually coded in terms of frequencies of requests, as patterns of sequences of choices, or as binary options between taking/not-taking the help.

The effects of seeking help on learning gains have been studied in educational practice though different approaches. For instance, it was found that higher learning gains were achieved when students paused to think and reason a hint, and elicit its implications (Shih, Koedinger, & Scheines, 2008), or when time-spent was properly allocated on help-seeking during problem solving (Arroyo & Woolf, 2005). Furthermore, it was also shown that the learning outcomes were low when learners intentionally misused help features and requested hints at a random time to obtain answers (Aleven et al., 2003; Baker et al., 2004), or frequently used executive help (Mathews, Mitrović, & Thomson, 2008).

In most cases (e.g., Huet et al., 2016; Shih et al., 2008; Vaessen et al., 2014), the help delivered to learners was concerning the knowledge/cognitive domain. Cognitive help has the potential to facilitate deeper thinking and processing mechanisms acknowledged to promote permanent learning gains. However, focusing only on cognitive aspects lacks characteristics that induce learners' self-reflection and self-awareness. Such metacognitive skills have been strongly associated with learners' awareness of when they really need help, the decision to ask for it, and the ability to evaluate the delivered help (Karabenick & Berger, 2013).

To address this issue, some studies considered delivering metacognitive information to assist learners in engaging with the learning task and regulating help-seeking (Daley et al., 2016; Roll et al., 2011). The idea was to provide learners with information about their own usage of the help-seeking facilities of the learning environment, making explicit to them their own help-seeking behaviour. For example, in intelligent tutoring system contexts, a geometry “tutored step-based problem-solving environment” (Roll et al., 2011, p. 268) modelled help-seeking by considering factors like the learners' knowledge level, previous help-seeking patterns and time spent on the problem. The system provided corrective feedback to the learners and encouraged them to change their behaviour, whenever it detected a “help-seeking error” (p. 268) according to the help-seeking model. In another example, the metacognitive information delivered to students was extracted from their own use of the online curriculum and made explicit to them their own data from the online system. This information included content knowledge (feedback with the correctness of the response), strategic use of the curriculum (exploitation of hints and other available support facilities) and engagement (in terms of difficulty ratings) (Daley et al., 2016). The goal was to support students' self-reflection and encourage them to change their help-seeking behaviour. The authors examined the degree to which the learners could translate the delivered metacognitive information, as well as the degree to which this information could trigger changes in learners' help-seeking. The results demonstrated a trend that metacognitive prompts can impact help-seeking behaviour and learning performance. However, in both approaches, the focus was on supporting the development of learners' help-seeking skills as a metacognitive, self-regulatory skill, and not on studying the effects of providing metacognitive help on learning engagement and gains.

Delivering efficient metacognitive help is not a trivial task to accomplish. Recent developments in learning analytics research have opened new opportunities and perspectives on the design and delivery of meaningful metacognitive information to help learners, in the form of visualizations dashboards (Corrin & de Barba, 2015; Durall & Gros, 2014; Martinez-Maldonado et al., 2016; Sedrakyan et al., 2018). Information visualization is an effective sense-making tool due to its ability to synthesize complex data in a way for viewers to quickly understand, compare, reflect, and ultimately decide (Heer & Agrawala, 2008).

So far, dashboards typically visualize learners' own interaction trace data, aiming at triggering learners' self-awareness and self-reflection mechanisms (Bodily & Verbert, 2017; Rodríguez-Triana, Martínez-Monés, Asensio-Pérez, & Dimitriadis, 2014). Most current visual analytics are based only on learner performance indicators (e.g., where a learner is coping/not coping well, how much time was spent, how the learner's progress compares to teacher specified and/or peer scores) that do not seem to contribute to learner's motivation and engagement (Verbert et al., 2014). Being performance-oriented, those implementations decrease learner mastery orientation (Lonn, Aguilar, & Teasley, 2015). Seminal research (Sedrakyan et al., 2018) demonstrates that to be effective, the delivered information needs to be grounded in the regulatory mechanisms underlying the learning processes. Numerous regulatory mechanisms such as setting goals, planning, time-management, monitoring progress, evaluating and reflecting have been identified in self-regulated learning (SRL) theory as core aspects that can facilitate learning processes (Zimmerman, 1990). The adoption of such regulatory mechanisms is particularly important when the learner is the main end-user of visual analytics, with a central goal to reinforce self-reflection and self-regulation (Jivet, Scheffel, Drachsler, & Specht, 2017). More precisely, learning analytics dashboards (and visual analytics in general) are intended (but not limited to) to help learners manage their time more efficiently (e.g., Tabuenca, Kalz, Drachsler, & Specht, 2015), to support them in monitoring and reflecting on the learning tasks (e.g., Nussbaumer, Hillemann, Gütl, & Albert, 2015) and understand their own learning through reflective judgement and evaluation (e.g., Authors, 2019). In a nutshell, Roll and Winne (2015) elaborated on the topic and supported that visual analytics can inform the learners about options (increase awareness) that may have an impact on the phases of their self-regulated learning as described in Winne and Hadwin's (1998) model.

However, considering learners as the main recipients (end-users) of visual analytics has received criticism: it has been argued that it might promote competition between the learners rather than chasing knowledge mastery (Jivet et al., 2017), and there is always a concern that the learners might not know how to make-sense of this information (MacNeill, Campbell, & Hawksey, 2014). Despite this unease, previous studies have shown that the learners can interpret their own performance indices, yet they reserve a scepticism on how to practically convert this information into action (Corrin & de Barba, 2015). Nonetheless, feeding this information to the learners encounters the danger that they may focus too much on their own self ("ego"), with unwanted effects on the learning (e.g., might lose motivation if the performance indices are low, or stop trying if the indices are high, just to preserve their reputation and avoid failure).

2.1 | Motivation of the research and research question

Synopsizing the above, four core facets become evident: (a) most of the previous studies delivered cognitive help with clear results regarding how it can efficiently affect the learning gains (in term of outcomes and behaviours); (b) cognitive help is not enough since help-seeking is a complex mechanism that, in order to be efficient, requires metacognitive skills (such as knowing when to seek help, monitoring and reflective judgement); (c) metacognitive prompts can support the development of learners' help-seeking skills, but it is not sufficiently clear how they can affect other learning constructs (e.g., engagement); (d) visual analytics can deliver meaningful metacognitive information to help learners engage in their learning, but their efficiency is often hindered by learners' performance orientation and focus on the "self" instead of on the "task". In addition, all previous studies followed cross-sectional research designs, that is, conducted at one particular point in time, without considering the time dimension, and rely on "static" differences rather than "change" following intervention (Setia, 2016; Turner, 2013).

However, individuals' behaviour usually changes in essential ways over time. In fact, it is difficult to imagine a theory (macro, meso, or micro) being purposely developed to explain a phenomenon at only a single point in time. Studies with repeated measurements extend beyond a single moment in time and can establish sequences of events. Collecting repeated measurements of key variables can provide a more definitive evaluation of within-person change across time; it allows the researcher to exclude the effects of individual differences, that is, to be able to detect effects that are due to the control conditions (treatment) (Howitt & Cramer, 2011). As such, to validate the effects of metacognitive help on learning outcomes, continuous processes with repeated measurements for the same sample should be employed, and changes in learners' behaviour over time—due to the use of this type of help—need to be considered. This study addresses this objective. Specifically, the study considers a task-related metacognitive help format (i.e., the visual analytics) and aims at investigating the changes in learners' on-task effortful behaviour, caused by the mediating effect of delivering on-demand this kind of information. Therefore, the core research question that guided the research is:

> **RQ: (a)** Are there any changes in learners' effortful behaviour over time, due to receiving metacognitive help in the form of task-related visual analytics? If yes, how significant are these changes? **(b)** Are there any changes in learners' usage of the metacognitive help over time? If yes, how significant are these changes?

3 | METHODOLOGY

3.1 | Study design

The study followed a crossover longitudinal research design (Ployhart & Vandenberg, 2010). Crossover longitudinal studies follow the same sample at regular intervals and make repeated observations and measurements of the same variables for the same groups of people; every subject in the sample serves as their own "control", that is, they belong both to the experimental group and the control group during the different points in time of the measurements. The experimental group receives a sequence of "treatments" (or exposures). These observations will enable researchers to track changes in independent variables (predictors - P) over time and to relate these changes to one or more treatment variables (treatments - T) that might explain why the changes in dependent variables (outcomes - O) occur. Longitudinal designs facilitate the measurement of difference or change in a variable from one period to another, that is, the description of patterns of change over time; it is hypothesized that changes in the predictors and mediators contribute to change in the outcomes, and not static levels of some variables predicting static levels in another. Measurements are taken on each variable (P, T, O) over three or more distinct time periods (three is the minimum, but more is better).

During the design of the present study, we had to address the following three core methodological issues, and take decisions accordingly:

a. Determine the optimal number of measurement occasions and their intervals to appropriately model the hypothesized form of change. Providing equally spaced repeated measurements is recommended to be avoided (Ployhart & Vandenberg, 2010). In this study, four measurements were carried out in total. The time-distance between the first and second was three weeks, between the second and third was two weeks, and the interval between the third and the fourth measurement was again three weeks. Those intervals were decided in line with the progress in the course lectures and on the basis that the students would have sufficient study-time in between of the self-assessments.
b. Maintaining the integrity of the original sample can be difficult over an extended period. In this study, from the 122 participant students at the first phase, 67 (54.9% of the initial sample) participated in all four phases. According to Hedeker, Gibbons, & Waternaux (1999); p.82 - table 2), the indicated sample size for medium effect sizes is 62, for a 4 time-points, Power = 0.8, $\alpha = 0.05$ test. Furthermore, the F-test for Repeated Measures ANOVA - within factors, indicates that the minimum total sample size required is 65, when the research design includes 1 group with 4 time-points (number of measurements), aiming at large effect sizes ($\eta_p^2 = 0.14$, f = 0.4), with Power = 0.8, $\alpha = 0.05$, and nonsphericity correction coefficient epsilon = 1. As such, the number of participants (67) is acceptable and maintains statistical power as it surpasses the recommended size. For this study, power analysis was conducted using R. Drop-out analysis was also conducted to affirm the robustness of the findings.
c. It can be difficult to show more than one treatment variables at a time. In this study, the treatment variable (T) is the task-related visual analytics usage. The dependent variable (O) is learning performance, and the independent variables (P) include the students' on-task engagement, that is, response-times to answer correct/wrong and effort expenditure.

Regarding point b, it should be noted that effect size is the most critical outcome of empirical studies because it communicates the practical significance of the results (Lakens, 2013). The common practice of simply citing universal guidelines for interpreting the magnitude of effect size coefficients has been criticized as too rigid and potentially misleading (Lipsey et al., 2012; Peng, Chen, Chiang, & Chiang, 2013), unless considering important factors such as the type of study design. Therefore, study design needs also to be taken into consideration for calculating the effect sizes. In order to ensure that this study and its reported outcomes and detected effects are valid, it is critical to carefully choose the sample size because it provides statistical support and guarantees statistical power maintenance (Guo, Logan, Glueck, & Muller, 2013).

3.2 | Research participants and experimental procedure

In this study, 122 undergraduate students (55 females [45.1%] and 67 males [54.9%], aged 19–26 years-old [M = 20.254, *SD* = 1.411, *N* = 122]) at a European University were initially enrolled in a continuous self-assessment procedure, consisting of four phases, for the Management Information Systems I course, at the University computers lab, for 60 min, each phase. The participation in the procedure was optional; it was offered to facilitate the students' self-preparation before the final exams. Due to that "option", the students could take none, one, two, or all the available self-assessment activities. Students who had not taken the first activity were allowed to participate in any of the other phases, if they wanted to. As such, 95 students participated in the second phase, whereas in the third phase, the attendee number was 86 and in the fourth it was 118. In total, 67 (54.9% of the initial sample – 29 females [43.3%] and 38 males [56.7%], aged 19–26 years-old [M = 20.182, *SD* = 1.341, *N* = 67]) students participated in all four phases. The considered students had previously used the self-assessment environment (its default version, without the help mechanism) at least one time before the present study.

Prior to their participation, all students signed an informed consent form that explained the data collection and the self-assessment processes and was giving the right to researchers to use the data for research purposes. Students were aware that their interactions would be tracked, and anonymized prior to being analyzed, and that the data would be stored for 3 years.

During the first phase of the study, all students took a fixed self-assessment test: they had to complete 15 multiple-choice tasks; each task had four possible answers, but only one was the correct. The tasks were delivered to the participants in predetermined order. The students could temporarily save their answers on the tasks, to review them, to alter their initial choices, and to save new answers; they could also skip a task and answer it (or not) later. For the second phase, three weeks after the first one, a new self-assessment activity was available (with different tasks). In this phase, in addition to the previously available actions, the students could also request task-related visual analytics for each task (treatment) (described in

Section 3.3). Prior to this phase, the students had a brief introductory presentation of the task-related visual analytics, to explain what information would be available to them, and how to use it (Lin, 2001). The instructions were available throughout the activity. In the third phase, two weeks after the second one, the treatment was removed and a new self-assessment activity (with new tasks) was available. Finally, the fourth phase, three weeks after the third one, was exactly like the second phase (with new tasks). In between the self-assessment activities, the students attended the regular course lectures. The overall process is illustrated in Figure 1.

Before the self-assessment activities, the difficulty of the tasks was determined (easy, medium, hard) by calibrating them using prior assessment results, that is, the mastery levels of previous participants and the correctness of answers. Each task's participation in the self-assessment score was according to its difficulty, varying from 0.5 points (easy) to 0.75 points (medium) to 1 point (hard), and only the correct answers were considered, that is, students received zero points for the tasks that they solved incorrectly or chose not to submit an answer.

At the end of each self-assessment activity, the test score was available to the students, along with their full-test results, including all the tasks they had completed, their responses, the correctness of the responses, the option to check the correct answers for the tasks that they had submitted wrong answers, and the task-related visual analytics in order to compare, rethink, self-reflect and self-evaluate.

It should be clarified that the decision to conduct the study in a continuous self-assessment context with multiple-choice tests 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 (Adesope, Trevisan, & Sundararajan, 2017). 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, that is, the so called "testing effect" (Carpenter, 2009; Roediger III & Karpicke, 2006). 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 (Cantor, Eslick, Marsh, Bjork, & Bjork, 2015; Little, Bjork, Bjork, & Angello, 2012), and at the same time, a growing number of studies on this topic have reported robust benefits of testing on transfer of learning (Carpenter, 2012). Furthermore, it has been argued that self-assessment leads students to a greater awareness, by training them to self-regulate their motivation and behaviour, as well as by promoting metacognition and fostering reflection on their own progress in knowledge or skills, and finally, to understanding themselves as learners (Nicol & MacFarlane-Dick, 2006).

Finally, to make sure that the effects on engagement were due to the mediation of the metacognitive help and not to the natural students' progress as they advance through the course, (a) the tasks in each self-assessment activity were totally independent from the previous activities, (b) the tasks were selected so as they cover all the spectrum of difficulty, (c) we tracked the changes in the usage of the metacognitive help, and considered the time point of the measurements as a variable in the analysis of the interactions (Sections 3.4–3.5), and (d) we did not focus on measuring the differences in performance, but only in engagement.

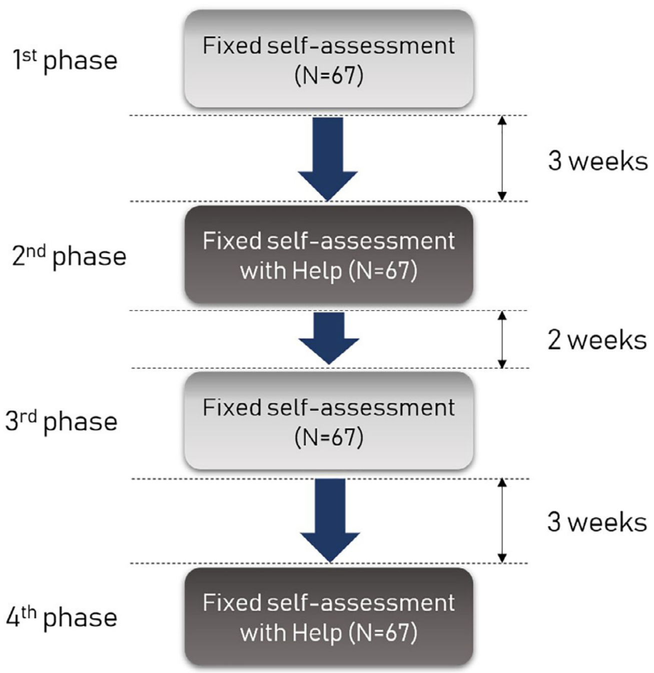

FIGURE 1 Overview of the longitudinal experimental study—Phases, duration and participants [Colour figure can be viewed at wileyonlinelibrary.com]

3.3 | The task-related visual analytics

During the design of task-related visual analytics as on-demand metacognitive help, two design models were considered: (a) the Contextualized Attention Metadata (CAM) schema (Wolpers, Najjar, Verbert, & Duval, 2007) for providing coordinated views over the data, and (b) the metacognitive computational model of help-seeking (Aleven, Mclaren, Roll, & Koedinger, 2006) for guiding the desired help-seeking behaviour (i.e., the learner should request help only when she really needs it, and receives meaningful metacognitive help). Based on these principles, the format and the content of the on-demand feedback were decided, respectively.

The task-related information that should be provided to the learner was determined so that this knowledge could activate learner's monitoring, reflection and judgment on the tasks. The ultimate goal was to help the learner meet the requirements of each task, that is, the real effort needed to deal with each task, the real difficulty, and the time required to allocate on each task. Next, regarding the presentation of this information (the content of the task-related analytics), it is delivered in three simple (easy-to-read) bar/column charts, including: (a) the number of correct vs. the number of wrong answers submitted for this task (for inferring its difficulty), (b) the average students' effort expenditure vs. their average performance (i.e., correctness of answers) for this task, and (c) the average time spent to solve this task correctly vs. the average time spent to carry out the task wrongly vs. the average time spent to solve the task. Figure 2a,b illustrate the task-related visual analytics delivered as meta-cognitive help.

To further elaborate on this decision, as explained in Section 2, on the one hand, visual analytics have been employed to mediate awareness and specific regulatory mechanisms (Bodily & Verbert, 2017; Nussbaumer et al., 2015; Sedrakyan et al., 2018; Tabuenca et al., 2015), while on the other hand, metacognitive prompts have great potential to induce regulative strategies (e.g., monitoring, controlling, reflection) during help-seeking (Bannert & Mengelkamp, 2013). Thus, the chosen analytics aim at bringing on the same page and mapping the regulatory mechanisms supported by visual analytics with those facilitated by the metacognitive prompts; the selected analytics focus on mediating regulatory mechanisms such as monitoring metacognitive information (e.g., difficulty of tasks), reflective judgement (e.g., inferring the real effort requirements), controlling the metacognitive knowledge (e.g., time-management and time allocation), and awareness of the need to seek-help (e.g., knowing for which task to request for visual analytics/on-demand usage of feedback).

In a sense, these visualizations are expected to provide fruitful and actionable insight to students about the real difficulty of the tasks, about the real effort needed to deal with the tasks, and about the average time required to allocate on the tasks. The visual analytics are displayed on-demand, as complementary metacognitive information about the tasks. Using this information properly is expected to support the learner in regulating herself: to control her effort exertion, to manage her time-allocation, and to improve her help-seeking skills (Lonn et al., 2015). To achieve that, the learner has to undertake a four-steps process: (a) to interpret the visualized information (i.e., the visualizations target at promoting analytics interpretation into meaningful information); (b) to further process and understand this information (i.e., actuation of deeper thinking mechanisms); (c) to assess this information in terms of how useful and constructive it is (i.e., triggering the critical evaluation of the available metacognitive information); and (d) to put it together with what she already knows about the task and decide on how she can use it to solve the task (i.e., initiating a self-reflection mechanism on how this information can be sufficiently exploited).

Every time the learner needs (or believes she needs) additional information about a task, that is, beyond cognitive clarifications, she has the option to ask for the above analytics. The visual analytics tool obtains the necessary temporal and performance indicators from the learning environment, and instantly generates the charts on-demand, by analysing all learners' logged interactions with that task (i.e., actual usage). For resolving "cold-start" issues, (i.e., the absence of data the first time a task is being viewed by the students) the analytics from former self-assessment procedures were employed. Those analytics were extracted during the calibration of the task pool (see sub-Section 3.2) and are updated with the arriving observations.

Providing this information per se could easily be perceived as the typical performance-oriented indicators. As explained in Jivet et al. (2017), learning analytics dashboard deliver information that typically concerns metrics about the learners themselves, that is, about the "self", "offering misguided frames of reference for comparison" (p. 82). Those metrics are very likely to put the learners in a competing mode, where "being better than others" becomes the norm in terms of what defines a successful learner, with unwanted impact on learning engagement and mastery orientation. To overcome this issue, the task-related visual analytics employed in this study, do not provide information about the "self", but instead, they deliver collective information about the "tasks". The idea behind this intervention is based on the simple shift of the focus from the "self" to the "task". Although the considered indexes have similarities with typical performance-oriented indexes computed per learner, they facilitate different goals: (a) since they are calculated from all learners' data when dealing with a specific task, the aggregated information describes the task and not the learner, (b) the accumulative information about the tasks is more action-oriented and aims to trigger deeper evaluation of the actual requirements of the tasks and guide learner's judgment and metacognitive inference, than the abstractly deduced "user-model" values, commonly delivered to learners. In a sense, those indexes do not intend to inform the learner (who requested this information) about how well all other students are performing, but rather about what one can infer about the real requirements of the task, and to engage with it in a "solution-behaviour" manner. Therefore, it is reasonable to expect that all learners will try to gain the best for themselves if they know that their interactions are accumulatively presented with those from all other learners; as such they are not competing with each other, but they rely on the "knowledge" of the others in a knowledge mastery chasing manner (they do not want to import bias in the aggregated analytics because they will be receivers of this bias, as well).

3.4 | Data collection and measures

In this study, data were collected with an online self-assessment environment (Authors, 2013). In all phases, measures commonly used in the field of learning analytics, acknowledged to satisfactorily explain students' engagement (e.g., response-times, frequencies) (Ada & Stansfield, 2017; Henrie, Halverson, & Graham, 2015; Maldonado-Mahauad, Pérez-Sanagustín, Kizilcec, Morales, & Munoz-Gama, 2018) were used. As already mentioned in the introduction, in this study, learners' on-task effortful behaviour synopsizes learners' on-task behavioural engagement. The purpose of this study is not to investigate and understand the dimensions of learners' engagement per se, but only to detect and justify changes in learners' patterns of on-task effortful behaviour when an intervention is available (i.e., the visual

FIGURE 2 (a) The task-related analytics visualizations—information about an easy task. (b) The task-related analytics visualizations—information about a hard task [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Measures used in the study

Variable	Name	Description	Phases 1, 3	Phases 2, 4
TTAC	Time to answer correctly	The response-time aggregated on submitting correct answers	✓	✓
TTAW	Time to answer wrongly	The response-time aggregated on submitting the wrong answers	✓	✓
RTE	Effort	When students exhibit solution behaviour—a measure of engagement	✓	✓
LP	Learning performance	The self-assessment score	✓	✓
TVAV	Time-spent on visual analytics viewing	The average time the student spends on viewing the analytics visualizations		✓
FVAR	Frequency of visual analytics requests	How many times the student asks for analytics visualizations		✓

analytics). Therefore, we isolated and explored only the actual interactions between the learners and the tasks and visual analytics, using factors that have been previously suggested in literature (i.e., time-spent on viewing the visual analytics, and response-time effort on the tasks – further explained in this section).

Specifically, during the two treatment phases (i.e., when the task-related visual analytics were available), analytics that indicate students' help-seeking interactions and quantify how students use the metacognitive help, were tracked, as well. Table 1 illustrates the measures captured and coded for each phase.

Briefly, time to answer correctly (TTAC) and time to answer wrongly (TTAW) indicate the respective response-time the students constantly aggregate on completing the tasks (Authors, 2014). Students' performance (LP) on each self-assessment activity was computed as: $\sum_{i=1}^{k} d_i z_i$, where zi ∈{0, 1} is the correctness of a student's solution on task i, and di is the task difficulty.

For the effort calculation, the Response Time Effort (RTE) measures the proportion of tasks which the students try to solve (solution behaviour) instead of guessing the answers (guessing behaviour) (Wise & Kong, 2005). For a student j: $RTE_j = \frac{\sum_k SB_{ij}}{k}$, where k is the number of tasks, and $SB_{ij} = \begin{pmatrix} 1, if\, RT_{ij} \geq Ti \\ 0, \text{otherwise} \end{pmatrix}$, where RTij is student's j response time to task i, and Ti is a threshold value (time) that discriminates solution from guessing. RTE is based on the hypothesis that when administered an item, unmotivated students will answer too quickly (i.e., before they have time to read and fully consider the item). Wise and Kong (2005) empirically investigated psychometric characteristics of RTE scores and found supportive evidence for score reliability and validity. For the calculation of the threshold value for each question during the calibration of the item bank, the present study followed the methodology suggested by Wise and Kong (2005). Specifically, we used the response-times from previous samples, and for each task we considered the response-time distributions and determined the span of short time spikes. The variance in spike width has been found to be strongly related to the amount of reading required by each task. We measured the total length of each task's stem and options (in characters) and established three initial task response-time thresholds as follows: if a task was shorter than 200 characters, a 5-sec threshold was used; if a task was longer than 800 characters or if the task provided some particular ancillary reading for the first time, a 12-sec threshold was used; for the remaining tasks, a 8-sec threshold was used.

It is worth mentioning that other approaches have conceptualized effort allocation by utilizing the proportion of response time allocated to an index of perceived relative strength in a scale 1 to 5, where 1 is the highest level (Hiemstra, Van Yperen, & Timmerman, 2019).

Furthermore, Time-spent on Visual Analytics Viewing (TVAV) is the time that the students spend on viewing the visual analytics and engage in order to make sense of their requirements, and Frequency of Visual Analytics Request (FVAR) is a counter that increases every time that the students make the respective request. The two metrics employed to model help-seeking behaviour were chosen in line with existing literature: (a) frequency of requests is a common measure used for coding help-seeking (e.g., Hao, Wright, et al., 2016; Huet et al., 2011); and (b) although time-spent viewing any kind of information per se does not imply understanding that information, this metric has been extensively used to model help-seeking during problem solving (e.g., Arroyo & Woolf, 2005; Roll et al., 2011), and it has also been used to model effort-allocation and to discriminate "solution behaviour" from "guessing behaviour" (Wise & Kong, 2005). The combination of time-spent on and frequency of use has also been previously used to measure the usage of tools that provide help to students, namely a dictionary, instructional goals, example questions, and help with interpreting figures and text (Clarebout & Elen, 2009). Therefore, in the present approach, time-spent on visual analytics viewing is used as a means to capture and code the learners' endeavour to interpret the displayed analytics and to make-sense from the visualized information. In other words, time-spent measures for how long the learners are "using" the metacognitive help, given that—according to the experimental procedure explained in Section 3.2—they had been previously introduced on how to read and use the analytics. It should be explicitly stated that more time-spent on viewing the visual analytics does not necessarily means more understanding of the information, but it is purposed to measure the usage of the visual analytics.

3.5 | Data analysis

Except from the methodological issues on the study design addressed in Section 3.1, prior to running the measurements, one additional analytical issue should be resolved prior to the data analysis. Time is a metric for describing change and is frequently a predictor variable in longitudinal research. As such, defining the metric and coding it, is a core issue (Ployhart & Vandenberg, 2010). In this study, the variable that codes the time dimension is coded as ordinal, implying that there is an order in the measurements, but the interval length between the measurements is not taken into account. Elaborating on this decision, the treatment applied (i.e., the visualizations) was independent of the interval between the measurements because the students did not use it continuously between the self-assessment test, but only during the tests.

3.5.1 | Drop-out analysis

When working with longitudinal data, there is often participant drop-out that can rarely be avoided. Drop-out is defined as the last phase (measurement) of a study where there is data for a particular person (Mazumdar et al., 2007). This is different from missing data on that a person could have missing data in a phase of the study, but non-missing data later on, indicating that the person did not drop-out. Drop-out could result in unbalanced dataset, and can introduce selection bias, that is, the observed dataset can be regarded as the result of a "selection" process (the person "selects" to not continue in the study). Due to the bias, different drop-out processes have different implications in terms of statistical analysis. To the extent possible, one must try to determine why the attrition occurred, and to detect if there is something systematic on the non-response at later times, and hence the possibility of bias in the results. Overall, there are three drop-out processes, that is, the Missing Not At Random—MNAR—and the Missing At Random/Missing Completely At Random—MAR/MCAR. Those processes correspond to data missingness depending or not on the study purposes/outcomes, and the introduced bias being or not "ignorable" (Little & Rubin, 2019).

In lack of other explanatory variables (e.g., motivation to participate, perceived usefulness) to associated the missingness of participation data with, we considered students' performance (scores) in each phase they participated, assuming that they opted out for reasons related to their performance. Participation and missingness of data are synopsized in Table 2.

We detected 11 patterns of (missingness of) participation. Those patterns are illustrated as columns in Figure 3. In this figure, the red color indicates that there were students with missing data in the corresponding phase, and the green color indicates that there were students with no missing data. For instance, the third pattern (from left to right) in this tabular representation indicates that there were students who did not participate in phases 1 and 3, but participated in phases 2 and 4. The number below the pattern (e.g., 67, 22) indicates how many times (i.e., for how many students) that pattern was detected (e.g., pattern10: data missing from phases 2 and 3 was detected for 22 students, pattern11: no missing data from any phase was detected for 67 students). The number above the pattern indicates in how many phases there had been missing data (e.g., pattern1 shows students who participated in two phases, that is, phases 2 and 3; pattern6 shows students who have missing data only in one phase, that is, phase 1). Each pattern of non-participation appears less times than the participation pattern (pattern11).

Furthermore, Liu (2016) elaborated on the approaches for the detection of the drop-out process, with MAR identified as the most realistic and the most frequently used assumption on the data missingness in empirical studies. In line with this assumption, this study assumes that the data is MAR. We performed the tests of the MCAR assumption (versus the MAR assumption) to test whether the missing data process is one of MCAR. Missing data were imputed using Little's test (Little, 1988) in R to assess for MCAR for multivariate data with missing values, with 10 missing-data patterns. A sig. < 0.05 is interpreted as the missing data is not MCAR (i.e., is either MAR or non-ignorable). In this study, $\chi^2 = 28.42$ and sig. $= 0.076$ with df $= 19$, verifying that the data is MCAR (failing to reject the null hypothesis). When the dropout process is MCAR, the specification of a drop-out model is not necessary and the inference is based on the observed values as if there were no drop-outs. Thus, we decided to consider only cases with non-missing values for further analysis.

3.5.2 | Panel data analysis

In preparation for analysis, the four datasets with the measurements on the on-task engagement and learning performance (i.e., TTAC,

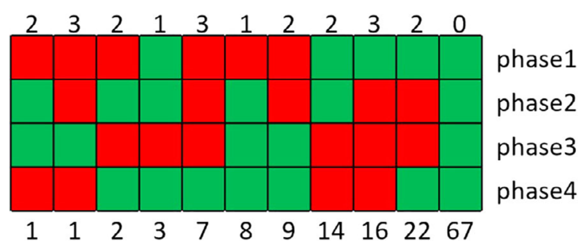

FIGURE 3 The patterns of drop-out (and participation) in each phase [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 2 Participation and data missingness in each phase

	All	Continue from phase1	Dropped—phase1	New in phase2	Continue from phase2	Dropped—phase2	New in phase3	Continue from phase3	Dropped—phase3	New in phase4
phase1	122									
phase2	95	84	38	11						
phase3	86	67	55		9	2	10			
phase4	118	92	30		10	1		9	1	7

TTAW, RTE and LP), and the usage of the metacognitive help (i.e., the task-related visual analytics, TVAV and FVAR) collected throughout the study, were merged and transformed to long format, that is, each subject had a row for each time point, and the repeated measurements were in a single column/variable. An identifier was generated as well, that is, the number of the phase each measurement was collected at, for grouping the observations: we had four groups. This variable was also used to code the time dimension.

Next, Hausman test differentiated between fixed effects and random effects model. It tests whether the coefficients estimated by the efficient random effects estimator are the same as the ones estimated by the consistent fixed-effects estimator (Hausman, 1978). The result supported the assumption of correlation between observation's errors and predictors, thus, fixed effects model was used. The initial model included the following variables: self-assessment phase identifier (group variable), the three on-task engagement variables (i.e., the two response-times variables and the effort variable), and the "interactions" between the phase identifier (group variable) and each one of the on-task engagement variables. An interaction is two or more variables multiplied together. When including interactions in the model, the individual terms should be in the model as well, because otherwise it is impossible to say if the observed effect is caused by the interactions, or by the omitted individual term. The interactions terms reveal if there is a difference in changes of outcome of interest (in this study, the learning engagement) between the groups (the self-assessment phases) due to the intervention (the usage of task-related visual analytics).

For data exploration in the pre–analysis phase we plotted RTE and LP variables against each of the time variables, with simple linear regression lines overlaid. These plots showed an extremely right-skewed distribution to the time variables and a distinctly log-normal (nonlinear) relationship between each of them and the RTE and LP responses. To remove nonlinearity from the time and effort variables, a simple log transformation was applied. Feng et al. (2014) raised their concerns that, log transformations usually do not make data conform more closely to the normal distribution, and the results of standard statistical tests performed on log-transformed data are often not relevant for the original, non-transformed data. In our study, the original data followed an approximately log-normal distribution. To verify that the transformed data follows the normal distribution, we run the Shapiro–Wilk test for all variables, as the sample size of the study is relatively small. The test returned a sig. varying from 0.165 to 0.224, that is, greater than 0.05, confirming that the transformed data is normal. Furthermore, concluding their study, Feng et al. (2014, p. 108) recommend that "if the data can be reasonably modeled by a parametric distribution such as the normal distribution, it is preferable to use the classic statistical methods because they usually provide efficient inference." In line with Feng et al., 2014, applying a simple log transformation to each time variable removed most of the skewness (Kutner, Nachtsheim, & Neter, 2004), and the log-transformed data followed the normal distribution, allowing for making inferences from it, and thus, we used these variables in all models instead of their original form.

As the assumption for MCAR data holds, missing data are non-informative, indicating that parameter estimates from linear mixed

TABLE 3 The final hierarchical linear mixed model for explaining the change in learning performance

			95% CI		
Variable	β	SE	Lower	Upper	Hedge's g
Phase (1)	2.401*	0.016	1.478	3.324	
Phase (2)	2.445*	0.010	1.792	3.089	
Phase (3)	2.413*	0.012	1.496	3.330	
Phase (4)	2.522*	0.010	1.824	3.111	
Time to answer correctly (TTAC)	0.023*	0.001	0.004	0.041	**0.27**
Time to answer wrongly (TTAW)	−0.025*	0.002	−0.038	−0.012	**0.26**
Effort (RTE)	0.190	0.013	−0.411	0.791	0.17
TTAC*Phase (1)	0.022*	0.002	0.018	0.036	**0.25**
TTAC*Phase (2)	0.024*	0.002	0.014	0.044	**0.26**
TTAC*Phase (3)	0.028*	0.001	0.017	0.039	**0.27**
TTAC*Phase (4)	0.031*	0.001	0.013	0.051	**0.29**
TTAW*Phase (1)	−0.018*	0.002	−0.022	−0.014	**0.24**
TTAW*Phase (2)	−0.013*	0.001	−0.018	−0.008	**0.25**
TTAW*Phase (3)	−0.017*	0.002	−0.021	−0.012	**0.24**
TTAW*Phase (4)	−0.014*	0.001	−0.019	−0.006	**0.26**
RTE*Phase (3)	0.086	0.008	0.063	0.096	**0.21**
RTE*Phase (4)	0.075	0.006	0.058	0.082	**0.22**

Note: Ranges for Hedge's g effect size are small > 0 .2, medium > 0.5 and large > 0.8.
Abbreviations: CI, confidence intervals; g, Hedges' g effect size; β, mean for the factor variable.
$^*p < .05$.

TABLE 4 Statistical differences between the phases of self-assessment with respect to fixed effects of the response-times variables on learning performance

Fixed effect	Between phases	*df*	F	Prob>F	η_p^2
Time to answer correctly (TTAC)	**Phase (1) vs. Phase (2)**	**66**	**5.023**	**0.014**	**0.18**
	Phase (1) vs. Phase (3)	**66**	**7.333**	**0.008**	**0.22**
	Phase (1) vs. Phase (4)	**66**	**8.664**	**0.002**	**0.23**
	Phase (2) vs. Phase (3)	66	1.024	0.346	0.07
	Phase (2) vs. Phase (4)	66	1.506	0.318	0.08
	Phase (3) vs. Phase (4)	**66**	**6.792**	**0.011**	**0.21**
Time to answer wrongly (TTAW)	**Phase (1) vs. Phase (2)**	**66**	**4.551**	**0.021**	**0.16**
	Phase (1) vs. Phase (3)	**66**	**5.386**	**0.013**	**0.19**
	Phase (1) vs. Phase (4)	**66**	**5.911**	**0.012**	**0.20**
	Phase (2) vs. Phase (3)	66	1.127	0.301	0.08
	Phase (2) vs. Phase (4)	66	2.004	0.272	0.09
	Phase (3) vs. Phase (4)	**66**	**6.635**	**0.012**	**0.21**
Effort (RTE)	**Phase (1) vs. Phase (3)**	**66**	**5.484**	**0.012**	**0.19**
	Phase (1) vs. Phase (4)	**66**	**6.212**	**0.013**	**0.20**
	Phase (2) vs. Phase (3)	66	1.117	0.296	0.05
	Phase (2) vs. Phase (4)	66	2.360	0.145	0.09
	Phase (3) vs. Phase (4)	**66**	**5.276**	**0.013**	**0.19**

Note: Bold represents statistically significant values. η_p^2: effect sizes are small > 0.01, medium > 0.06, large > 0.14.

TABLE 5 Paired t-test results for the visual analytics usage measures between the second and forth phase

				95% CI				
	Mean	*SD*	*SE* mean	Lower	Upper	t	*df*	p-value
TVAV(2)-TVAV(4)	105.466	376.610	46.010	13.603	197.328	2.292	66	.025
FVAR(2)-FVAR(4)	1.388	1.072	0.131	1.126	1.649	10.594	66	.001

Abbreviations: FVAR, frequency of visual analytics requests; TVAV, time-spent on visual analytics viewing.

models are unbiased (Liu, 2016). We fitted hierarchical linear mixed models using the REstricted Maximum Likelihood (REML) estimation, having the learning performance in the self-assessment as the outcome variable. To select the final model, we successively fitted the current model (the initial was the one with all variables), then computed the model's corrected Akaike Information Criterion (AICc) (Hurvich & Tsai, 1989) and removed the variables with the highest p-values. AICc is a correction of AIC for small samples (such as the one in this study) and is a second-order estimate of information lost by a given model. The finally selected model was the one with the smallest AICc. The analysis was performed in Stata 12.

3.5.3 | Paired *t*-test

For exploring the differences in the usage of the metacognitive help, and detect possible changes in the respective analytics factors (i.e., TVAV, FVAR) from the interactions with the task-related visual analytics between the second and the fourth phase of the study, paired sample t-test was used. Typically, a paired t-test is used when we are interested in the difference between two variables—often separated by time—for the same subject.

4 | RESULTS

4.1 | The hierarchical linear mixed model

The final model included the following variables: the self-assessment phase identifier (group variable), the two response-times and the effort variables, as well as the interactions between the phase identifier and the two response-times for all phases, and the interaction between the phase identifier and effort in the third and fourth phases. The chosen model (AICc = 2,197.7) was the best fitting model (AICc value for the initial model was 2,354.8). The hierarchical linear mixed model further revealed that both response-times and effort were statistically significant determinants of learning performance in self-assessment activities (Table 3).

Given the statistically significant effects of the interactions between phase identifier and the two response-time predictors (for all phases), and of the interactions between phase identifier and effort for the third and fourth phases, additional analyses were performed to shed light to the nature of these interactions. Specifically, a one-way repeated-measures ANOVA (i.e., within-subjects ANOVA) was used to determine whether the means of the predictors over the four phases were statistically different. The results from this analysis showed statistically significant differences in mean of the response-time variables and effort across the four different phases (F(4, 63) = 9.94, p = .007 for TTAC; F(4, 63) = 7.82, p = .011 for TTAW; F(4, 63) = 6.32, p = .022 for RTE). Table 4 shows the direct (paired) comparisons between the different phases.

As seen in this table, statistically significant differences in the effects of response-times on performance is detected between the first and second phase, between the first and third phase, between the first and fourth phase, and between the third and fourth phase, while, the difference in the effects of these variables on learning performance between the other phases are only moderate. Similarly, the differences in the effects of effort on performance are statistically significant between the first and third phase, between the first and fourth phase and between the third and fourth phase.

4.2 | Paired *t*-test results

Table 5 presents paired t-test results for the two analytics measures for students' usage of metacognitive help in the two self-assessment phases (second and fourth) when help was available.

As seen in this table, the difference in the analytics parameters that measure the use of metacognitive help was statistically significant for both measures (t(66) = 2.292, p = .025 for time-spent on viewing the help; t(66) = 10.594, p = .001 for requesting help).

5 | DISCUSSION

5.1 | Interpretation of the results

The overall results of the study, as demonstrated in Table 3, revealed a statistically significant effect of the distinct response-times to answer correctly/wrongly on learning performance, in a setting of a longitudinal study, where the subjects serve as their own control. Both response-time factors contribute the most to explaining the variance in learning performance in all stages of the study. These findings provide additional evidence to previous claims that the distinctive response-times are good determinants of learning performance in self-assessment tests (Authors, 2014, 2016, 2018; Shih et al., 2008; Wang & Hanson, 2005). However, the previously reported strong positive effect of effort on learning performance was not further confirmed (Authors, 2015, 2016; Setzer, Wise, van den Heuvel, & Ling, 2013; Silm, Must, & Täht, 2013).

Table 4 shows direct comparisons between the four phases of the study with respect to the response-times, on-task effort and their effect magnitudes. As shown in this Table, between the first measurements of analytics parameters, prior to exposing learners to the treatment (i.e., the task-related visual analytics), and the second measurement, when the metacognitive help was available, the difference in the response-times was statistically significant and the effect of the difference on explaining the difference on learning performance was statistically strong (F(1, 66) =5.023, η_p^2 =0.18 for TTAC; F(1, 66) = 4.551, η_p^2=0.16 for TTAW). Similarly, the difference in response-times explains satisfactorily the difference in performance between the first and the third phase of the repeated measurements (F(1, 66) = 7.333, η_p^2=0.22 for TTAC; F(1, 66) = 5.386, η_p^2=0.19 for TTAW), and between the first and the fourth self-assessment activity (F(1, 66) = 8.664, η_p^2=0.23 for TTAC; F(1, 66) = 5.911, η_p^2=0.20 for TTAW). On the other hand, the effects of these differences are statistically small or marginally medium between the second and third phases (F(1, 66) = 1.024, η_p^2=0.05 for TTAC; F(1, 66) = 1.127, η_p^2=0.07 for TTAW), and between the second and the fourth phases (F(1, 66) = 1.506, η_p^2 =0.06 for TTAC; F(1, 66) = 2.004, η_p^2=0.07 for TTAW). And, although on-task effort was not found to be per se a strong determinant of learning performance, the differences in on-task effort between the first and third phase (F(1, 66) = 5.484, η_p^2=0.19), between the first and fourth phase (F(1, 66) = 6.212, η_p^2=0.20) and between the third and fourth activities seem to contribute to explaining the respective differences in scores (F(1, 66) = 5.276, η_p^2=0.19).

What mediated and caused these differences was the usage of the visual analytics. Thus, what these findings imply is that the intervention employed, that is, the usage of the available metacognitive help, strongly contributes in increasing learners' on-task effortful behaviour, which in turn, results in improved performance. Previous research results shown that when students paused to think and reason a hint, and elicit its implications (Shih et al., 2008) and when time-spent was properly allocated on help-seeking (Arroyo & Woolf, 2005; Authors, 2019), the achieved learning outcomes were improved. The findings from the current study further support those results and extend them by introducing a different format of metacognitive help. Furthermore, current findings contribute to hypothesizing that students who systematically seek help are likely to be responsibly involved with the learning tasks and careful about their answers, compared to those who avoid using help.

The present findings are also in agreement with previous results that reported on the learners' ability to read and understand visual analytics as metacognitive feedback (Corrin & de Barba, 2015) and to make use of the metacognitive information (Daley et al., 2016) to improve their performance. Considering learners as the main recipients of learning analytics data might put in question how efficiently the learners could make-sense from this information (MacNeill et al., 2014). As shown in Table 5, there was a statistically significant difference in the usage of the metacognitive help between the two self-assessment activities that the visual analytics were available. Both measure of help usage (time-spent on viewing the visual analytics and frequency of request for visual analytics) appear to be increased in the fourth phase. Combining this finding with the fact that the two activities (i.e., second and fourth) were totally independent from each other in terms of knowledge content (see Section 3.2), and with the fact that learning performance in the fourth activity was also improved, one can support that the students found help and metacognitive support to regulate their responses in the task-related visual analytics. Previous studies argued that the learners can interpret their own performance indices, yet they reserve a skepticism on how to practically convert this information into action (Corrin & de Barba, 2015). Taking us a step ahead from visualizing learners' own interaction trace data that the dashboards typically do (Rodríguez-Triana et al., 2014), the core innovation of this work derives from exploiting easy-to-read task-related visual analytics to provide instrumental information about the tasks to learners, and investigates how their on-task engagement changes due to this intervention.

5.2 | Implications for research and practice

Individuals' behaviour usually changes in essential ways over time. Prior studies followed cross-sectional research designs. The core contributions of this study were threefold:

- Methodologically, it is one of the very limited in number studies in the field of learning analytics that implemented a longitudinal research design. This study showcased how the time metric for describing change can be coded to facilitate the research design. Time of measurements is frequently a predictor variable in longitudinal research. As shown from the findings, this factor was included in the final fixed effect model, and it indeed was one of the strong determinants of the change in learning performance. As such, this study provided the description of a coherent longitudinal study in the area of learning analytics and showcased how former hypotheses can be further explored and validated with respect to changes in learners' behaviour over time (*methodological implication*).
- This study provides further insight and evidence into existing body of research on the role of metacognitive help to increase learning gains. The theoretical model suggested in this study considers learners' effortful behaviour and investigates changes in learners' behaviour due to the metacognitive help. From the findings became apparent that the employed metacognitive help format caused significant changes in learners' behaviour in terms of response-times, which in turn resulted in changes in performance. Investigating the effects of changes in the usage of the on-demand task-related visual analytics on the changes in performance is necessary to be clarified, as well (*implication for research*). Designing and implementing longer longitudinal studies, with more phases of exposing the sample to the treatments (i.e., more points in time) would facilitate that objective. In addition, providing alternative forms of assistance (e.g., executive help formats, explicit hints, etc.), measuring the effects of the changes in response-times and effort, and comparing these changes to the ones estimated in this study is expected to shed light to the effect size of the employed intervention.
- Combining the findings of this study with previous results that indicated that learners interpret their own performance indices (Corrin & de Barba, 2015), further justified the role and significance of the intervention. In accordance with the literature (Lee, Drake, & Thayne, 2016), this study argues that learner data have the potential to support decision-making and enhance learning. Such a support can be transformative for students, especially the ones who are already familiar with such technologies and motivated (Lee et al., 2016). As such, this study provided a strong indication that training learners to use, read, and make-sense from learning analytics fosters their metacognition and assists them to ask for assistance at the moment they really need it (*practical implication*). Accordingly, further training the learners on how to efficiently use visual analytics is expected to build their capacity for data-driven decisions. Future work needs to collect data from other learning settings (e.g., MOOCs, problem solving), at larger scale and use different and repeated survey data collections. Cross-validating and extending our findings will allow us to generalize them and even identify activities where on-demand metacognitive help can be used to optimize its potential.

5.3 | Limitations

First and foremost, a basic limitation of this study is that it assumed that the self-assessment procedures were of similar difficulty for the learners, despite the fact that the content of the self-assessment tests differed from phase to phase. In a sense, we treated the self-assessment procedures as "black boxes". Considering the same number of tasks of similar levels of difficulty do not establish that the procedures are identical. The effects of the content itself on explaining the variance in on-task engagement and performance should be considered, as well.

Furthermore, as mentioned in Section 3.2, a drop-out of participants was observed during the different phases of the study. From the drop-out analysis (Section 3.5.1) it became apparent that the drop out was completely random. However, this result might be due to the assumption that drop-out was related to their performance in the previous phase, which might not be the case. As such, other exploratory variables need to be taken into consideration (e.g., motivation, usefulness).

One other limitation of the present study is the size of the sample which is marginal and the number of points in time selected is minimum as well. Experimentation with bigger sample sizes should be conducted and further longitudinal research following the same students over different self-assessment procedures, enhanced with metacognitive help, is needed in order to understand how responsible and effective learners become when using the task-related visual analytics.

Future work concerns another limitation: in this study, only response-times and on-task effort were considered in the hierarchical model. Other factors that have been strongly associated to the learners' performance (e.g., motivational constructs or affective states) should be explored as well, and other types of data (e.g., interviews) could also provide qualitative information that would help contextualize and interpret the quantitative data. This is within our future work plans.

Finally, we would like to clarify that in Section 3.2 we refer to the "European University" as contextual (geographical) information about the study because we want to emphasize on the fact that the findings in the European context are not necessarily generalized/applicable in other location-based contexts, for example, Australia, United States, developing countries. This kind of geographical context may implicitly provide information about the students' exposure and familiarity with learning technologies, and the extent to which specific sophisticated technologies are systematically adopted by the academic institutions and form an institutional culture (e.g., learning analytics are more widely used in Australia compared to Europe).

6 | CONCLUSIONS

The benefits of seeking help for the overall learning gains are beyond question, and the role of help-seeking in knowledge acquisition is catalytic. However, help-seeking is an inherently complex mechanism, instigated by learners' intrinsic motivational criteria and externalized as an adaptive behaviour (i.e., that evolves over time and is not stable). It is an often phenomenon that learners either underuse or overuse the available help facilities within the digital learning environments.

Former studies followed cross-sectional research designs to investigate the effects of help on learning constructs. However, in order a theory to be consolidated, multiple measurements are required. The present longitudinal study explored the changes in learners' on-task effortful behaviour and performance due to the changes in the use of metacognitive help. The results provided strong evidence that this help format contributes to increasing learners' efforts when completing learning tasks and to improving learning outcomes. Additional research is required on the role and effect of the effort factor, as well as on exploring other significant learning/learner factors that have been previously found to affect performance. The most interesting finding of this study, though, is that the learners were not "afraid" to use the visual analytics and make-sense out of it, resulting in increased response-times and better self-assessment outcome, over time.

CONFLICT OF INTEREST

The authors declare no potential conflict of interest.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Zacharoula Papamitsiou https://orcid.org/0000-0002-0982-3623

REFERENCES

Ada, M. B., & Stansfield, M. (2017). The potential of learning analytics in understanding students' engagement with their assessment feedback. *In* 2017 IEEE 17th international conference on advanced learning technologies (ICALT) (pp. 227–229). https://doi.org/10.1109/ICALT.2017.40

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research*, *87*(3), 659–701. https://doi.org/10.3102/0034654316689306

Aleven, V., Mclaren, B., Roll, I., & Koedinger, K. (2006). Toward meta-cognitive tutoring: A model of help seeking with a cognitive tutor. *International Journal of Artificial Intelligence in Education*, *16*(2), 101–128.

Aleven, V., Stahl, E., Schworm, S., Fischer, F., & Wallace, R. (2003). Help seeking and help design in interactive learning environments. *Review of Educational Research*, *73*(3), 277–320. https://doi.org/10.3102/00346543073003277

Arroyo, I., & Woolf, B. P. (2005). Inferring learning and attitudes from a Bayesian network of log file data. In *2005 conference on artificial intelligence in education: Supporting learning through intelligent and socially informed technology* (pp. 33–40). Amsterdam, The Netherlands: IOS Press.

Azevedo, R., Taub, M., Mudrick, N. V., Millar, G. C., Bradbury, A. E., & Price, M. J. (2017). Using data visualizations to foster emotion regulation during self-regulated learning with advanced learning technologies. In J. Buder & F. W. Hesse (Eds.), *Informational environments: Effects of use, effective designs* (pp. 225–247). Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-319-64274-1_10

Baker, R. S. (2005). *Designing intelligent tutors that adapt to when students game the system* (Doctoral dissertation), Carnegie Mellon University, Pittsburgh, PA. Published as CMU Technical Report CMU-HCII-05-104.

Baker, R. S., Corbett, A. T., Koedinger, K. R., & Wagner, A. Z. (2004). Off-task behavior in the cognitive tutor classroom. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 383–390). New York, NY: ACM. https://doi.org/10.1145/985692.985741

Bannert, M., & Mengelkamp, C. (2013). Scaffolding hypermedia learning through metacognitive prompts. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 171–186). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4419-5546-3_12

Ben-Eliyahu, A., Moore, D., Dorph, R., & Schunn, C. D. (2018). Investigating the multidimensionality of engagement: Affective, behavioral, and cognitive engagement across science activities and contexts. *Contemporary Educational Psychology*, *53*, 87–105. https://doi.org/10.1016/j.cedpsych.2018.01.002

Bodily, R., & Verbert, K. (2017). Review of research on student-facing learning analytics dashboards and educational recommender systems. *IEEE Transactions on Learning Technologies*, *10*(4), 405–418. https://doi.org/10.1109/TLT.2017.2740172

Bodily, R., Kay, J., Aleven, V., Jivet, I., Davis, D., Xhakaj, F., & Verbert, K. (2018). Open learner models and learning analytics dashboards: A systematic review. In *Proceedings of the 8th international conference on learning analytics and knowledge* (pp. 41–50). New York, NY: ACM. https://doi.org/10.1145/3170358.3170409

Cantor, A. D., Eslick, A. N., Marsh, E. J., Bjork, R. A., & Bjork, E. L. (2015). Multiple-choice tests stabilize access to marginal knowledge. *Memory & Cognition*, *43*(2), 193–205.

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(6), 1563.

Carpenter, S. K. (2012). Testing enhances the transfer of learning. *Current Directions in Psychological Science*, *21*(5), 279–283.

Clarebout, G., & Elen, J. (2009). The complexity of tool use in computer-based learning environments. *Instructional Science*, *37*(5), 475–486. https://doi.org/10.1007/s11251-008-9068-3

Corrin, L., & de Barba, P. (2015). How do students interpret feedback delivered via dashboards? In *5th international conference on learning analytics and knowledge* (pp. 430–431). New York, NY: ACM. https://doi.org/10.1145/2723576.2723662

Daley, S. G., Hillaire, G., & Sutherland, L. M. (2016). Beyond performance data: Improving student help seeking by collecting and displaying influential data in an online middle-school science curriculum. *British Journal of Educational Technology*, *47*(1), 121–134. https://doi.org/10.1111/bjet.12221

Davis, D., Chen, G., Jivet, I., Hauff, C., & Houben, G.-J. (2016). Encouraging metacognition & self-regulation in MOOCs through increased learner feedback. In S. Bull, B. M. Ginon, J. Kay, M. D. Kickmeier-Rust, & M. D. Johnson (Eds.), *LAL 2016 - learning analytics for learners* (pp. 17–22). CEUR.

Durall, E., & Gros, B. (2014). Learning analytics as a metacognitive tool. In *6th international conference on computer supported education* (pp. 380–384). https://doi.org/10.5220/0004933203800384

Feng, C., Wang, H., Lu, N., Chen, T., He, H., Lu, Y., & Tu, X. M. (2014). Log-transformation and its implications for data analysis. *Shanghai Archives of Psychiatry*, *26*(2), 105–109. https://doi.org/10.3969/j.issn.1002-0829.2014.02.009

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Fletcher, A., & Shaw, G. (2012). How does student-directed assessment affect learning? Using assessment as a learning process. *International Journal of Multiple Research Approaches*, *6*(3), 245–263. https://doi.org/10.5172/mra.2012.6.3.245

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*(1), 59–109. https://doi.org/10.3102/00346543074001059

Gama, C. (2004). Metacognition in interactive learning environments: The reflection assistant model. In J. C. Lester, R. M. Vicari, & F. Paraguaçu (Eds.), *Intelligent Tutoring Systems* (pp. 668–677). Berlin, Heidelberg: Springer Berlin Heidelberg.

Greene, B. A. (2015). Measuring cognitive engagement with self-report scales: Reflections from over 20 years of research. *Educational Psychologist*, *50*(1), 14–30. https://doi.org/10.1080/00461520.2014.989230

Guo, Y., Logan, H. L., Glueck, D. H., & Muller, K. E. (2013). Selecting a sample size for studies with repeated measures. *BMC Medical Research Methodology*, *13*, 100. https://doi.org/10.1186/1471-2288-13-100

Hao, Q., Barnes, B., Wright, E., & Branch, R. M. (2016). The influence of achievement goals on online help seeking of computer science students. *British Journal of Educational Technology*, *48*(6), 1273–1283. https://doi.org/10.1111/bjet.12499

Hao, Q., Wright, E., Barnes, B., & Branch, R. M. (2016). What are the most important prediction of computer science students' online help-seeking behaviors? *Computers in Human Behavior*, *62*, 467–474. https://doi.org/10.1016/j.chb.2016.04.016

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, *46*(6), 1251–1271.

Hedeker, D., Gibbons, R. D., & Waternaux, C. (1999). Sample size estimation for longitudinal designs with attrition: Comparing time-related contrasts between two groups. *Journal of Educational and Behavioral Statistics*, *24*(1), 70–93. https://doi.org/10.2307/1165262

Heer, J., & Agrawala, M. (2008). Design considerations for collaborative visual analytics. *Information Visualization*, *7*(1), 49–62. https://doi.org/10.1145/1391107.1391112

Heerde, J. A., & Hemphill, S. A. (2018). Examination of associations between informal help-seeking behavior, social support, and adolescent psychosocial outcomes: A meta-analysis. *Developmental Review*, *47*, 44–62. https://doi.org/10.1016/J.DR.2017.10.001

Henrie, C. R., Halverson, L. R., & Graham, C. R. (2015). Measuring student engagement in technology-mediated learning: A review. *Computers & Education*, *90*, 36–53. https://doi.org/10.1016/j.compedu.2015.09.005

Hiemstra, D., Van Yperen, N. W., & Timmerman, M. E. (2019). Students' effort allocation to their perceived strengths and weaknesses: The moderating effect of instructional strategy. *Learning and Instruction*, *60*, 180–190.

Howitt, D., & Cramer, D. (2011). *Introduction to research methods in psychology* (3rd ed.). Harlow, England Prentice Hall. Retrieved from. http://openlibrary.org/books/OL25074414M

Huet, N., Escribe, C., Dupeyrat, C., & Sakdavong, J.-C. (2011). The influence of achievement goals and perceptions of online help on its actual use in an interactive learning environment. *Computers in Human Behavior*, *27*(1), 413–420. https://doi.org/10.1016/J.CHB.2010.09.003

Huet, N., Moták, L., & Sakdavong, J. C. (2016). Motivation to seek help and help efficiency in students who failed in an initial task. *Computers in Human Behavior*, *63*, 584–593. https://doi.org/10.1016/J.CHB.2016.05.059

Hurvich, C. M., & Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, *76*(2), 297–307. https://doi.org/10.1093/biomet/76.2.297

Jivet, I., Scheffel, M., Drachsler, H., & Specht, M. (2017). Awareness is not enough: Pitfalls of learning analytics dashboards in the educational practice. In É. Lavoué, H. Drachsler, K. Verbert, J. Broisin, & M. Pérez-Sanagustín (Eds.), *Data driven approaches in digital education* (pp. 82–96). Cham, Switzerland: Springer International Publishing.

Karabenick, S. A. (2011). Classroom and technology-supported help seeking: The need for converging research paradigms. *Learning and Instruction*, *21*(2), 290–296. https://doi.org/10.1016/J.LEARNINSTRUC.2010.07.007

Karabenick, S. A., & Berger, J.-L. (2013). Help seeking as a self-regulated learning strategy. In B. J. Zimmerman (Ed.), *Applications of self-regulated learning across diverse disciplines: A tribute to* (pp. 237–261). Charlotte, NC: IAP Information Age Publishing.

Kautzmann, T. R., & Jaques, P. A. (2019). Effects of adaptive training on metacognitive knowledge monitoring ability in computer-based learning. *Computers & Education*, *129*, 92–105. https://doi.org/10.1016/j.compedu.2018.10.017

Kutner, M, Nachtsheim, C, & Neter, J (2004). *Applied linear regression models*, (4th ed.), New York: McGraw-Hill/Irwin.

Labuhn, A. S., Zimmerman, B. J., & Hasselhorn, M. (2010). Enhancing students' self-regulation and mathematics performance: The influence of feedback and self-evaluative standards. *Metacognition and Learning*, *5*(2), 173–194. https://doi.org/10.1007/s11409-010-9056-2

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*, 863. https://doi.org/10.3389/fpsyg.2013.00863

Lawson, M. A., & Lawson, H. A. (2013). New conceptual frameworks for student engagement research, policy, and practice. *Review of Educational Research*, *83*(3), 432–479. https://doi.org/10.3102/0034654313480891

Lee, V. R., Drake, J. R., & Thayne, J. L. (2016). Appropriating quantified self technologies to support elementary statistical teaching and learning. *IEEE Transactions on Learning Technologies*, *9*(4), 354–365. https://doi.org/10.1109/TLT.2016.2597142

Lin, X. (2001). Designing metacognitive activities. *Educational Technology Research and Development*, *49*(2), 23–40. https://doi.org/10.1007/BF02504926

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., … Busick, M. D. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms*. National Center for Special Education Research.

Little, J. L., Bjork, E. L., Bjork, R. A., & Angello, G. (2012). Multiple-choice tests exonerated, at least of some charges: Fostering test-induced learning and avoiding test-induced forgetting. *Psychological Science*, *23*(11), 1337–1344.

Little, R. J., & Rubin, D. B. (2019). *Statistical analysis with missing data* (Vol. 793). John Wiley & Sons.

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*(404), 1198–1202.

Liu, X. (2016). Chapter 14 - methods for handling missing data. In X. Liu (Ed.), *Methods and applications of longitudinal data analysis* (pp. 441–473). Oxford: Academic Press. https:// doi.org/https://doi.org/10.1016/B978-0-12-801342-7.00014-9

Long, Y., & Aleven, V. (2017). Enhancing learning outcomes through self-regulated learning support with an open learner model. *User Modeling and User-Adapted Interaction*, *27*(1), 55–88. https://doi.org/10.1007/s11257-016-9186-6

Lonn, S., Aguilar, S. J., & Teasley, S. D. (2015). Investigating student motivation in the context of a learning analytics intervention during a summer bridge program. *Computers in Human Behavior*, *47*, 90–97. https://doi.org/10.1016/j.chb.2014.07.013

MacNeill, S., Campbell, L. M., & Hawksey, M. (2014). Analytics for education. *Journal of Interactive Media in Education*, 1–12.

Mäkitalo-Siegl, K., Kohnle, C., & Fischer, F. (2011). Computer-supported collaborative inquiry learning and classroom scripts: Effects on help-seeking processes and learning outcomes. *Learning and Instruction*, *21*(2), 257–266. https://doi.org/10.1016/j.learninstruc.2010.07.001

Maldonado-Mahauad, J., Pérez-Sanagustín, M., Kizilcec, R. F., Morales, N., & Munoz-Gama, J. (2018). Mining theory-based patterns from big data: Identifying self-regulated learning strategies in massive open online courses. *Computers in Human Behavior*, *80*, 179–196. https://doi.org/10.1016/j.chb.2017.11.011

Martinez-Maldonado, R., Schneider, B., Charleer, S., Shum, S. B., Klerkx, J., & Duval, E. (2016). Interactive surfaces and learning analytics: Data, orchestration aspects, pedagogical uses and challenges. In *6th International Conference on Learning Analytics & Knowledge* (pp. 124–133). New York, NY: ACM. https://doi.org/10.1145/2883851.2883873

Mathews, M., Mitrović, T., & Thomson, D. (2008). Analysing high-level help-seeking behaviour in ITSs. In W. Nejdl, J. Kay, P. Pu, & E. Herder (Eds.), *Adaptive hypermedia and adaptive web-based systems* (pp. 312–315). Berlin, Heidelberg: Springer Berlin Heidelberg.

Mazumdar, S., Tang, G., Houck, P. R., Dew, M. A., Begley, A. E., Scott, J., … Reynolds, C. F., 3rd. (2007). Statistical analysis of longitudinal psychiatric data with dropouts. *Journal of Psychiatric Research*, *41*(12), 1032–1041. https://doi.org/10.1016/j.jpsychires.2006.09.007

Molinillo, S., Aguilar-Illescas, R., Anaya-Sánchez, R., & Vallespín-Arán, M. (2018). Exploring the impacts of interactions, social presence and emotional engagement on active collaborative learning in a social web-based environment. *Computers & Education*, *123*, 41–52. https://doi.org/10.1016/j.compedu.2018.04.012

Nelson, T. O., & Nahrens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 125–173). New York: Academic Press.

Nelson-Le Gall, S. (1985). Help-seeking behavior in learning. *Review of Research in Education*, *12*, 55–90.

Nicol, D. J., & MacFarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, *31*(2), 199–218.

Nussbaumer, A., Hillemann, E.-C., Gütl, C., & Albert, D. (2015). A competence-based service for supporting self-regulated learning in virtual environments. *Journal of Learning Analytics*, *2*(1), 101–133.

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

Papamitsiou, Z., & Economides, A. A. (2014). Temporal learning analytics for adaptive assessment. *Journal of Learning Analytics*, *1*(3), 165–168. https://doi.org/10.18608/jla.2014.13.13.

Papamitsiou, Z., & Economides, A. A. (2015). Temporal learning analytics visualizations for increasing awareness during assessment. *International Journal of Educational Technology in Higher Education*, *12*(3), 129–147. https://doi.org/10.7238/rusc.v12i3.2519.

Papamitsiou, Z., Karapistoli, E., & Economides, A. A. (2016). Applying classification techniques on temporal trace data for shaping student behavior models. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (LAK '16)* (pp. 299–303), New York, NY: ACM. https://doi.org/10.1145/2883851.2883926.

Papamitsiou, Z., Economides, A. A., Pappas, I. O., & Giannakos, M. N. (2018). Explaining learning performance using response-time, self-regulation and satisfaction from content: an fsQCA approach. In Proceedings of the 8th International Conference on Learning Analytics and Knowledge (LAK '18) (pp. 181–190), New York, NY: ACM. https://doi.org/10.1145/3170358.3170397.

Papamitsiou, Z., Economides, A. A., & Giannakos, M. N. (2019). Fostering Learners' Performance with On-demand Metacognitive Feedback. In: Scheffel M., Broisin J., Pammer-Schindler V., Ioannou A., Schneider J. (eds), *Transforming Learning with Meaningful Technologies. EC-TEL 2019*. *Lecture Notes in Computer Science*, (Vol. 11722. pp. 423–435). Cham: Springer. https://doi.org/10.1007/978-3-030-29736-7_32.

Peng, C.-Y. J., Chen, L.-T., Chiang, H.-M., & Chiang, Y.-C. (2013). The impact of APA and AERA guidelines on effect size reporting. *Educational Psychology Review*, *25*(2), 157–209. https://doi.org/10.1007/s10648-013-9218-2

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review*, *16*(4), 385–407. https://doi.org/10.1007/s10648-004-0006-x

Ployhart, R. E., & Vandenberg, R. J. (2010). Longitudinal research: The theory, design, and analysis of change. *Journal of Management*, *36*(1), 94–120. https://doi.org/10.1177/0149206309352110

Richardson, M., Abraham, C, & Bond, R. (2012). Psychological correlates of university students' academic performance: a systematic review and meta-analysis. *Psychological Bulletin*, *138*(2), 353–387.

Rodríguez-Triana, M. J., Martínez-Monés, A., Asensio-Pérez, J. I., & Dimitriadis, Y. (2014). Scripting and monitoring meet each other: Aligning learning analytics and learning design to support teachers in orchestrating CSCL situations. *British Journal of Educational Technology*, *46*(2), 330–343. https://doi.org/10.1111/bjet.12198

Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249–255.

Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2011). Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system. *Learning and Instruction*, *21*(2), 267–280. https://doi.org/10.1016/j.learninstruc.2010.07.004

Roll, I., Aleven, V., McLaren, B. M., Ryu, E., Baker, R. S. J., & Koedinger, K. R. (2006). The help tutor: Does metacognitive feedback improve students' help-seeking actions, skills and learning? In M. Ikeda, K. D. Ashley, & T.-W. Chan (Eds.), *Intelligent Tutoring Systems* (pp. 360–369). Berlin, Heidelberg: Springer Berlin Heidelberg.

Roll, I., Baker, R. S. J. d., Aleven, V., & Koedinger, K. R. (2014). On the benefits of seeking (and avoiding) help in online problem-solving environments. *Journal of the Learning Sciences*, *23*(4), 537–560. https://doi.org/10.1080/10508406.2014.883977

Roll, I., & Winne, P. H. (2015). Understanding, evaluating, and supporting self-regulated learning using learning analytics. *Journal of Learning Analytics*, *2*(1), 7–12.

Rothbart, M. K., & Bates, J. E. (1998). Temperament. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology: Vol. 3. Social, emotional, personality development* (pp. 105–176). New York: Wiley.

Ryan, A. M., & Shin, H. (2011). Help-seeking tendencies during early adolescence: An examination of motivational correlates and consequences for achievement. *Learning and Instruction*, *21*(2), 247–256. https://doi.org/10.1016/J.LEARNINSTRUC.2010.07.003

Schwendimann, B. A., Rodríguez-Triana, M. J., Vozniuk, A., Prieto, L. P., Boroujeni, M. S., Holzer, A., … Dillenbourg, P. (2017). Perceiving learning at a glance: A systematic literature review of learning dashboard research. *IEEE Transactions on Learning Technologies*, *10*(1), 30–41. https://doi.org/10.1109/TLT.2016.2599522

Sedrakyan, G., Malmberg, J., Verbert, K., Järvelä, S., & Kirschner, P. A. (2018). Linking learning behavior analytics and learning science concepts: Designing a learning analytics dashboard for feedback to support learning regulation. *Computers in Human Behavior*, *107*, 105512. https://doi.org/10.1016/j.chb.2018.05.004

Setia, M. S. (2016). Methodology series module 3: Cross-sectional studies. *Indian Journal of Dermatology*, *61*(3), 261–264. https://doi.org/10.4103/0019-5154.182410

Setzer, J. C., Wise, S. L., van den Heuvel, J. R., & Ling, G. (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

Sharma, K., Papamitsiou, Z., & Giannakos, M. (2019). Building pipelines for educational data using AI and multimodal analytics: A "grey-box" approach. *British Journal of Educational Technology*, *50*, 3031. https://doi.org/10.1111/bjet.12854

Sharma, K, Papamitsiou, Z, Olsen, J. K, & Giannakos, M (2020). Predicting learners' effortful behaviour in adaptive assessment using multimodal data. https://doi.org/10.1145/3375462.3375498.

Shih, B., Koedinger, K. R., & Scheines, R. (2008). A response time model for bottom-out hints as worked examples. In R. de Baker, T. Barnes, & J. Beck (Eds.), *Proceedings of the 1st international conference on educational data mining* (pp. 117–126).

Silm, G., Must, O., & Täht, K. (2013). Test-taking effort as a predictor of performance in low-stakes tests. *Trames Journal of the Humanities and Social Sciences*, *17*(4), 433–448.

Stone, N. J. (2000). Exploring the relationship between calibration and self-regulated learning. *Educational Psychology Review*, *12*(4), 437–475. https://doi.org/10.1023/A:1009084430926

Sun, J. C.-Y., & Rueda, R. (2012). Situational interest, computer self-efficacy and self-regulation: Their impact on student engagement in distance education. *British Journal of Educational Technology*, *43*(2), 191–204. https://doi.org/10.1111/j.1467-8535.2010.01157.x

Sunar, A. S., White, S., Abdullah, N. A., & Davis, H. C. (2017). How learners' interactions sustain engagement: A MOOC case study. *IEEE Transactions on Learning Technologies*, *10*(4), 475–487. https://doi.org/10.1109/TLT.2016.2633268

Tabuenca, B., Kalz, M., Drachsler, H., & Specht, M. (2015). Time will tell: The role of mobile learning analytics in self-regulated learning. *Computers and Education*, *89*, 53–74. https://doi.org/10.1016/j.compedu.2015.08.004

Tsai, Y., Lin, C., Hong, J., & Tai, K. (2018). The effects of metacognition on online learning interest and continuance to learn with MOOCs. *Computers & Education*, *121*, 18–29. https://doi.org/10.1016/j.compedu.2018.02.011

Turner, J. R. (2013). Cross-sectional study. In M. D. Gellman & J. R. Turner (Eds.), *Encyclopedia of Behavioral Medicine (p. 522)*. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4419-1005-9_1010

Vaessen, B. E., Prins, F. J., & Jeuring, J. (2014). University students' achievement goals and help-seeking strategies in an intelligent tutoring system. *Computers & Education*, *72*, 196–208. https://doi.org/10.1016/J.COMPEDU.2013.11.001

Verbert, K., Govaerts, S., Duval, E., Santos, J. L., Assche, F., Parra, G., & Klerkx, J. (2014). Learning dashboards: An overview and future research opportunities. *Personal Ubiquitous Comput.*, *18*(6), 1499–1514. https://doi.org/10.1007/s00779-013-0751-2

Wang, T., & Hanson, B. A. (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

Winne, P. H. (1997). Experimenting to bootstrap self-regulated learning. *Journal of Educational Psychology*, *89*(3), 397–410. https://doi.org/10.1037/0022-0663.89.3.397

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In *Metacognition in educational theory and practice* (pp. 277–304). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Wise, S. L., & Kong, X. (2005). Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education.*, *18*, 163–183. https://doi.org/10.1207/s15324818ame1802_2

Wolpers, M., Najjar, J., Verbert, K., & Duval, E. (2007). Tracking actual usage: The attention metadata approach. *Journal of Educational Technology & Society*, *10*(3), 106–121.

Wolters, C. A. (1998). Self-regulated learning and college students' regulation of motivation. *Journal of Educational Psychology*, *90*(2), 224–235. https://doi.org/10.1037/0022-0663.90.2.224

Yigitbasioglu, O. M., & Velcu, O. (2012). A review of dashboards in performance management: Implications for design and research. *International Journal of Accounting Information Systems*, *13*(1), 41–59. https://doi.org/10.1016/j.accinf.2011.08.002

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist*, *25*(1), 3–17. https://doi.org/10.1207/s15326985ep2501_2

How to cite this article: Papamitsiou Z, Economides AA. The impact of on-demand metacognitive help on effortful behaviour: A longitudinal study using task-related visual analytics. *J Comput Assist Learn*. 2020;1–18. https://doi.org/10.1111/jcal.12472