Information flow and cognition affect each other: Evidence from digital learning

Kshitij Sharma[*,a], Katerina Mangaroska[a], Niels van Berkel[b], Michail Giannakos[a], Vassilis Kostakos[c]

[a] Norwegian University of Science and Technology, Trondheim, Norway
[b] Aalborg University, Aalborg, Denmark
[c] University of Melbourne, Melbourne, Australia

A B S T R A C T

In the context of learning systems, identifying causal relationships among information presented to the user, their behavior and cognitive effort required/exerted to understand and perform a task is key to building effective learning experiences, and to maintain engagement in learning processes. An unexplored question is whether our interaction with presented information affects our cognitive effort (and behaviour), or vice-versa. We investigate causal relationship between information presented and cognitive effort (and behaviour) in the context of two separate studies ($N = 40$, $N = 98$), and study the effect of instruction (active/passive task). We utilize screen-recordings and eye-tracking data to investigate the relationship among these variables. To investigate the causal relationships among the different measurements, we use Granger's causality. Further, we propose a new method to combine two time-series from multiple participants for detecting causal relationships. Our results indicate that information presentation drives user focus size (behaviour), and that cognitive load (a measure of cognitive effort exerted) drives information presentation. This relationship is also moderated by instruction type and performance-level (high/low). We draw implications for design of educational material and learning technologies.

1. Introduction

Interaction between learners and learning technologies, also referred to as Learner-Computer Interaction (LCI), is a discipline that aims to understand and support users' learning experiences (Giannakos et al., 2018). Designing technologies to enhance human learning is a complex, multi-layered challenge, which requires input from interdisciplinary fields such as Human-Computer Interaction (HCI), Software Engineering, Psychology, and the Learning Sciences. Moreover, as the systematic use of learning technologies is becoming increasingly popular in the learning sciences, various technologies (e.g., learning management systems, intelligent tutoring systems) have already been adopted for teaching and learning practices. However, most contemporary learning systems are still oblivious to users' needs and capacities, and the usage data generated by these new technologies remains mostly unused in improving end-user interactions.

Humans do not use technology solely in the context of learning (i.e., to attain new knowledge). For example, humans use mobile phones to navigate, communicate, and shop; Internet of Thing (IoT) devices to optimize their living space and everyday activities; in-car entertainment, and so on. Hence, as humans' reliance on ubiquitous devices grows, so does the need for seamless integration of these type of devices in our daily lives. In other words, humans need technology that blends into the environment, adapts to users' cognitive capacities, and works towards fulfilling users' needs. This has been an active area of research following Weiser's vision of Ubiquitous Computing (Weiser, 1991). Therefore, research on the assessment of users' cognitive load[1] has made use of sensor devices to generate findings used to improve the interaction between users and computing devices (Gjoreski et al., 2018; Haapalainen et al., 2010). Moreover, despite challenges in inferring an individual's cognitive load, there has been considerable progress in the development of context-aware systems which built upon the

---

* Corresponding author.
  E-mail address: kshitij.sharma@ntnu.no (K. Sharma).

[1] Cognitive load is a load imposed by a given task on learners' cognitive system (Paas et al., 1994). This construct has two dimensions (Paas et al., 2003); 1) a causal dimension between the task and learners' characteristics, and 2) a measurable dimension of mental effort and performance. For example, if a person is told to solve arithmetic equations while driving a car the cognitive load on the driver will be higher for the arithmetic equation "23 + 45" than that for "3 + 4".

advancements in the physiology-sensing technology (Matkovič and Pejović, 2018; Tag et al., 2017). Considering these technological developments and the importance of cognitive load in learning sciences (Sweller, 1994), our study aims to use the advances in physiological sensing to investigate the relationship between information flow[2] and cognitive load (and consequently learners' behaviour, that is focus and attention change). Our subsequent goal is to derive guidelines that researchers could follow to improve the design of learning systems and the overall learning process.

When learning, users' cognitive processing of instructional materials needs to be at a correct level; too little processing could results in below-par learning (as well as users' boredom), whereas too much processing could lead to a high cognitive load (and users' frustration), potentially inhibiting learning altogether (Mayer and Moreno, 2003). Thus, information flow is one of the critical factors to consider when designing digital learning materials (Churchill, 2014). In fact, cognitive theory of multimedia learning (Mayer and Moreno, 2003) poses twelve design principles, one of which calls for presenting information appropriately (i.e., the information flow) to enhance processing and enable learning. Although it is widely accepted that the information flow is central to the field of learner-computer interaction, the casual relationship (which causes which) between information flow and users' cognition remains inconclusive. Thus, we aim to assess users' cognitive load in two different learning contexts (i.e., a video-based learning activity and a problem-solving task) for the purpose of increasing our understanding of humans as learners and the ways they immerse in a learning activity based on the information flow. These insights, in turn, might yield new information not currently reflected in performance-based measures (Paas et al., 1994).

In the past, researchers have used gaze data to better understand and quantify information flow (Grant and Spivey, 2003). Gaze data has been extensively used to explore different actions and conditions when users interact with learning systems (Bondareva et al., 2013; Kardan and Conati, 2013). In particular, gaze has been used to understand various cognitive processes underlying learners' behavior: cognitive load during planning and editing tasks(Prieto et al., 2015a); decision making when confronted with visual objects (Martínez-Gómez and Aizawa, 2014); attention and split or change of attention during different learning activities (e.g., reading, problem-solving) (Kizilcec et al., 2014); user focus size in complementing remote collaboration (Zhang et al., 2017). These examples show that gaze has been accepted as an accurate proxy of cognitive behavior during LCI. However, one of the major gaps in these studies is the correlational nature of the analyses, and even the frequent confusion of correlation with causation when researchers claim direct causal connections between two variables (Ferguson and Clow, 2017). The confusion between causality and correlation and the difficulty in identifying the real evidence has been discussed in several fields (Bollen and Bauldry, 2011). This misuse of evidence has also been encountered in learning technology and learning analytics, for instance with measurements of attainment in numeracy or literacy to be used to provide evidence of the effectiveness of teachers, schools, or even the states (Ferguson and Clow, 2017; Klenowski, 2015). In the same vein, our contribution emphasises a shift from correlation to causality, as an important step in establishing cause and effect between variables. Before describing the main contributions of the paper, we would answer

two important questions here:

Why causality is important? Finding out the causal relation between two measurements is essential to understand the "active connection" between those measurement Spirkin (1983). A causal relation tells us about the generation and determination of processes involved, which is much more information than what is embedded in a correlation. A correlation only tells us about the mutual association of the processes. For example, it is widely known that there is a correlation between the cognitive load experienced by users and the amount of information provided Mayer and Moreno (1998). However, in absence of a causal link between the two measurements (i.e., cognitive load and information content) it becomes difficult to design real-time adaptive systems to support processes that enable efficient consumption of the information Oppewal (2010). Causal relations provide the decision makers (e.g., teachers, educational technology researchers) a stronger basis (as compared to correlations) to decide upon the necessary actions for a given desired result Oppewal (2010).

Who will benefit from knowing the causal relations? Causal relations provide the decision makers (e.g., teachers, educational technology researchers) the opportunity to take appropriate actions so that the users (e.g., students) access the information presented to them in an efficient and effective manner. The knowledge of the causal relations between the different measurements can provide content/technology design guidelines/recommendations (Sarsenbayeva et al., 2020; Sugihara et al., 2012) to the teachers and educational technology researchers. The understanding of causal relationships could help the teachers and researchers in avoiding unforeseen situations in the digital learning settings (Sugihara et al., 2012).

In this paper, we 1) describe the causal relation between information flow (as inferred from the screen output), users' cognitive load (as inferred from users' gaze), and users' behavior (focus and attention change) in technology enhanced learning (TEL); 2) identify potential differences from this relation across learning performance/gains; and 3) identify potential differences in this relation between passive (i.e., video-based learning) and active (i.e., problem-solving) tasks. To do so, we run two distinct studies (i.e., a video-based learning activity and a problem-solving learning activity) across which we collect screen output and users' gaze.

In sum, we make the following contributions:

- We present insights from two studies in which information flow, users' cognitive load, user focus size, and attention were captured during a video-based learning activity and a problem-solving task.
- We show that information flow drives users' focus and cognitive load drives the information flow.
- We identify that the causality between these variables is stronger for low performing students as compared to high performing students.
- We identify that the causality between these variables is stronger for passive learning task (i.e., video-based learning) as compared to an active learning task.
- We showcase how we can investigate the causal relationships among the different measurements in the context of digital learning, and discuss its potential in learning technology research.

2. Related work

Over the past few decades, researchers have focused on developing an understanding of how people learn (Mayer, 2008). However, a better understanding of how people learn does not automatically yield clear specifications on how to design effective instructional methods and content (Mangaroska et al., 2018; Mayer, 2008). Taking into account the advances of learning sciences, interactive technologies, and sensing technologies (Di Lascio et al., 2018; Giannakos et al., 2019a; Jeremy et al., 2017), there is a critical need for understanding how people learn in and by interacting with contemporary learning environments (Giannakos et al., 2018). Doing so might support the

---

[2] Information flow indicates the actual information content present on the screen at any given moment (Gray, 2011). Intuitively, it is translated to the amount of storage required (in terms of bits). Thus, the higher the amount of storage required to store the current content on the screen, the higher would be the information flow to the learner. For example, in a video-lecture, if the text and/or graphics cover two-thirds of the screen, the information flow would be higher than the condition in which the text and/or graphics cover only a half of the screen.

design and development of future learning systems that take full advantage of a learner's cognitive capacities.

2.1. Information processing in multimodal learning

Contemporary learning environments are highly interactive and utilize various communication channels (Moreno and Mayer, 2007). In the past, their design was often theory-driven, but at present, their design tends to follow a data-driven approach (Mangaroska and Giannakos, 2018). However, designing learner-centred environments requires convergence of techniques and methods from interdisciplinary fields such as HCI, software engineering, cognitive sciences, and technology enhanced learning (TEL) (Balacheff and Lund, 2013). Moreover, analyses of such environments need to utilize representative, objective, diverse, and accurate data which allows researchers to understand users' cognitive capacities and design relevant learning technologies (Giannakos et al., 2018). To achieve this, researchers require a holistic understanding of how users interact with the learning environment in a given context, as well as their associated learning outcomes.

State-of-the-art learning resources encompass multiple types of media to deliver learning content to the users (learners) such as, videos (Sharma et al., 2015a), digital educational games (Prensky, 2003), and narrated animations (Mayer et al., 2005). Different media types trigger different senses of the learner. For example, videos transmit the information using audio and video channels (Paivio, 1990), games might trigger motion, gestures of the body, and problem solving elements in the brain (Giannakos et al., 2019b). Hence, multimodal learning allows learners to integrate information from different stimuli into one meaningful experience (Ochoa et al., 2018). Research in multimedia learning shows that learners' understanding can be enhanced if information from more than two modes is effectively combined (Fletcher and Tobias, 2005; Mayer, 2002; Paivio, 1990). For example, learners can watch a video of an experiment in chemistry while listening to their instructor explaining the expected outcomes. This example conveys the instructors' responsibility to combine different modes, verbal (e.g., spoken words) and non-verbal (e.g., images, animations), to achieve a content presentation that fosters learning and controls the listener's cognitive load (Mayer and Moreno, 2003). Because of this, learners are required to switch between content presented using different modalities; a situation which could result in learners experiencing substantial information processing while having limited learning capacity (Mayer and Moreno, 2003).

Human information processing, in typical problem solving settings, considers two dimensions: obtaining the required amount of information, and the level of cognitive effort required to solve the problem-at-hand (Simon, 1978). A typical information-processing model consists of three stages: attention, elaboration, and behavior (Simon, 1978). According to this theory, every human processes information differently due to the innate differences in their cognitive capacities.

When designing multimodal learning activities, it is not just the modes that increase the intricacy of the learning (information processing system); the contextual set up of the learning process (passive vs active tasks) adds an additional layer of complexity (James et al., 2002). A passive task (e.g., watching a video) is primarily concerned with information internalization through some form of memorization and storage in the working memory (Vecchi and Cornoldi, 1999). An active task (e.g., debugging code) is a task in which a learner transforms, integrates, and/or manipulates the content, so the task demands are primarily concerned with processing in working memory (Vecchi and Cornoldi, 1999). Furthermore, on the one hand, in a passive task (e.g., watching a video) most of the information is received by the user and there is seldom a reciprocal of the information from the user's end (the reciprocity is achieved via the user feedback in videos) (Van Gerven et al., 2002). On the contrary, in an active task the

information is transmitted back and forth between the user and the system (James et al., 2002). Therefore, while making an attempt to comprehend the relationship between the information presented to learners, their information processing behaviour and cognitive efforts, it is important to take the context (i.e., task type) into account to obtain a holistic understanding of the interplay between these variables.

2.2. Measuring cognitive load

Information processing inherently requires a certain level of cognitive load (Mayer, 1997; Paas et al., 1994). cognitive load is a multidimensional construct representing the level of perceived effort for thinking and reasoning while performing a particular task (Paas et al., 2003). In the learning context, one of the probable causes of cognitive load might be learners' interaction with the learning technology (the content and system itself) (Mayer and Moreno, 2003). Managing cognitive load is helpful in improving learning and avoiding stress, errors, and low performance (Prabhakharan et al., 2012; Sweller, 1994). Efficiently managing cognitive load would require an effective way of measuring the cognitive load (Brunken et al., 2003).

The methods used for measuring cognitive load can be divided into four broad categories (Brunken et al., 2003): subjective direct (self-reported stress), subjective indirect (self-reported mental effort), objective direct (brain signals and dual task performance), and objective indirect (physiological). Subjective indirect methods like post-hoc self-reports of cognitive load (Kaiser et al., 2016; Paas et al., 2003), NASA task load index (NASA-TLX) (Hart and Staveland, 1988; Prieto et al., 2015a) have an innate limitation of not occurring in real-time (Prieto et al., 2017). Subjective direct methods (Jovanović et al., 2019; Van Gog et al., 2012) are considered to be favourable, however because they are done at a specific frequency and at multiple intervals during the task performance, they are susceptible to being distracting in learning contexts. Moreover, identical timing and frequency might not work for tasks with different requirements and/or complexity. Both the subjective direct and indirect methods of measuring cognitive load cannot account for rapid changes in the learner's cognitive load, as for example encountered when learning programming or reading a pop-up information during a video streaming (Palinko et al., 2010).

Objective direct measures of the cognitive load (Peitek et al., 2018; Siegmund et al., 2014) through EEG or fMRI devices, negatively affect the interaction space and ecological validity of the study (Funk et al., 2016; Kosch et al., 2018). For example, modern off-the-shelf EEG caps[3] might be discomforting for the users over a long period of interaction, while fMRI machines limit the motion and the interaction with the learning technology. Another objective direct way of measuring cognitive load is primary and secondary task (dual-task performance) techniques (Brünken et al., 2002; Verwey and Veltman, 1996). In this technique, participants are required to solve an additional task with increasing complexity along with the primary task. The cognitive load is measured by the performance on the secondary task. This is not ideal for the motivation and attention of the learner while interacting with the learning technology.

Finally, the objective indirect measures of cognitive load overcomes these aforementioned limitations. Objective indirect measures support the automated measurement of cognitive load, even when no apparent change in task performance can be detected (Brunken et al., 2003). For example, a model combining the median of electrocardiogram and heat flux has shown high accuracy at distinguishing low and high levels of cognitive load (Haapalainen et al., 2010); pupilometric data was used to measure fluctuating levels of cognitive load in drivers (Palinko et al., 2010); galvanic skin response (GSR) was found to demonstrate changes in germane cognitive load levels (Gjoreski et al., 2018; Shi et al., 2007); and even real-time automatic cognitive load measured from

---

[3] https://neurolite.ch/en/products/eeg/enobio-wireless-eeg.

speech (Yin et al., 2007).Overall, task-invoked pupillary response is a reliable and sensitive measurement of cognitive load (Granholm et al., 1996).

One of the often used physiological measures (objective indirect) of cognitive load is eye-tracking (Buettner, 2013; Klingner et al., 2008; Poole and Ball, 2006). For example, in the past Backs & Walrath (Backs and Walrathf, 1992) used a person's Number of fixations, Fixation duration mean, and Fixation rate (fixations/s) to measure cognitive load. Hyona et al. (Hyönä et al., 1995) used pupil diameter in language-related tasks (not just visual). Later, Boucsein (Boucsein, 2000) used pupillary diameter, saccadic movements, and eye-blink rate to measure cognitive load. Their findings were confirmed by Poole (Poole and Ball, 2006). In their review, they highlight how pupil diameter can be used for computing cognitive load. Furthermore, Klingner et al. (Klingner et al., 2008), through both replications and new studies, strengthen the relation between pupillary response and cognitive load in a variety of visual and non-visual (reading) tasks. More recently, Buettner (Buettner, 2013); Prieto (Prieto et al., 2017; 2015b), and Gollan (Gollan et al., 2016) combined pupil diameter mean, SD, saccade speed, and number of fixations (>500ms) to compute an overall measurement of cognitive load and demonstrated this to be a reliable and accurate objective indirect measurement of the same.

Consequently, observing learners' interactivity with and in a learning technology using eye-tracking measures, has shown to be an appropriate technique in measuring cognitive load, as such measurements can be done in real-time without interrupting the learning (problem solving) process, limiting the interaction space, or sacrificing on a study's ecological validity.

2.3. Measuring user focus size and attention change

Eye-tracking provides researchers with a powerful method to capture a users' attention and focus on the screen. Most of the methods currently used to quantify user focus size and attention changes employ eye-tracking techniques (Holmqvist et al., 2011). To measure user focus size (i.e., what proportion of the screen the user is covering in a given time window), two primary ways of quantification techniques are used: fixation and saccades (Pappas et al., 2018; Radach et al., 2008) and gaze data entropy (Olsen et al., 2018; Sharma et al., 2018a). The ratio of fixation and saccade (Pappas et al., 2018; Radach et al., 2008) can inform whether the user is focused (not the same as paying attention). Several studies in different contexts such as **reading** (Radach et al., 2008), **web-usability** (Pappas et al., 2018), **and scene perception** (Unema et al., 2005), have shown that a high fixation to saccade ratio indicates a local focus, meaning that the user is looking at a small part of the screen; while a lower value of this ratio depicts a global focus, meaning the user is looking at a wider part of the screen than the former case. Another method used to quantify user focus size is through a rectangular grid overlaid on the screen which was used to compute the proportionality of gaze-time distribution and the entropy of this proportionality vector (Sharma et al., 2018b). Previous work has shown entropy to be a reliable objective indirect measurement of user focus size in different contexts such as **debugging**(Sharma et al., 2018b), **collaborative problem solving** (Sharma et al., 2018a), **and intelligent tutors**(Olsen et al., 2018).

To measure user attention change, transitions between Areas of interest (AOIs) on the screen (Holmqvist et al., 2011) is one of the widely used methods. The AOIs can be defined in three different ways: hypothesis driven (when the researcher knows where the user needs to pay attention) (Allopenna et al., 1998; Richardson et al., 2007), grid based (overlaying a rectangular grid on the screen) (Foulsham, 2008; Goldberg, 1999), and data driven (automatic unsupervised clustering of areas of screen the users have been paying attention, e.g., heatmaps) (Blignaut, 2010; Hernandez, 2007). Hypothesis driven AOIs and the transitions between them (i.e., attention change) have been

used to compare between experts/novices (Just and Carpenter, 1980), high/low performers (Sharma et al., 2015b), high/low task performance (Allopenna et al., 1998). One of the problems of this method roots from the possibility of unexpected behaviour from the users, which might lead to overlapping fixation distributions and thus the researcher might need to alter the size of the predefined AOIs (Orquin et al., 2016). The data-driven method of creating AOIs has been used mostly for visualizations such as, heatmaps (Hernandez, 2007) and attention maps (Blignaut, 2010). Since, data-driven AOIs are mostly created in an unsupervised manner and are susceptible to individual changes, hence they create a problem in comparing different user groups (Wulff, 2007). A middle ground to hypothesis-driven and data-driven AOI construction is overlaying a grid onto the screen and measuring the attention change based on this grid (Olsen et al., 2018). The grid AOIs are fixed from pre-gaze-analysis phase; however, the researchers can modulate the size of the grid to fit their requirements (Sharma et al., 2017). Grid based AOIs were used to compare the attention change across users to distinguish experts from novices (Sharma et al., 2013), high performing students from low performing students and the different levels of task-based success (Sharif and Maletic, 2010).

3. Research questions and rationale for the studies

As detailed in the previous section, there are a number of previous studies (Foulsham, 2008; Mayer, 1997; Radach et al., 2008; Reingold et al., 2001; Sharma et al., 2015a) showing evidence of a relation between information presented to learners, their cognitive load, attention shift, and focus. Moreover, these studies established eye-tracking as a practical method to measure a user's cognitive load, attention shift, and focus. However, the results from these studies are, to the best of our knowledge, correlation-based. Thus, in this contribution, we propose a shift from correlational to causal analysis in multimedia learning studies. We study the same measurements as in the previous research, but from a different statistical lens i.e., information flow (Mayer, 1997), cognitive load (Paas et al., 2003), attention shift (Holmqvist et al., 2011) and user focus size (Sharma et al., 2018b). The main reason for choosing these particular measurements was to show how we can arrive at different implications using the same measurements as before but using a analytical shift from correlation to causation.

To present our method of how the causation among different variables could be established from the already collected data, we ground the methodology in two different studies. Theses studies were conducted in different years, preliminary hypotheses, and covered different population of learners. The learning contexts were also different: the first study was contexualized in a video based learning paradigm, while the second study was setup in a code-debugging learning by problem solving paradigm. These two studies share three basic commonalities: 1) both are based on individual learning practices (e.g., video based learning and debugging); 2) both have been investigated within the multimedia-based learning paradigm; and 3) both learning settings require deep visual information processing from learners to achieve high learning outcomes. However, these two learning contexts present a contrast in the way the information is presented to the learners and the way learners interact with the tasks. In the video-based learning study, the content is provided to the learners in a monologue with the content changing every few seconds as the teacher writes on the blackboard. The learners could manipulate the content only via the video playback controls; thus, this study depicts a passive learning activity. Contrary to this, the code-debugging activity is within the learning-by-problem-solving paradigm. The information provided to the learners (i.e., the code) is mostly textual and static, so learners are allowed to change the content on the screen in order to solve the task; thus this study depicts an active learning activity. By selecting two different learning activities (i.e., active vs passive) in two different contexts (i.e., video-based vs debugging), and with different ways of information presentation (i.e.,

static vs dynamic), we aim to demonstrate our method of causal analysis. Consequently, through this contribution, we address the following research questions:

1. What is the causal relation, if any, between the information presented to users, their behaviour (focus and attention change) and cognitive load?
2. How does learning performance relate to the causal relation between the information presentation and the user behaviour?
3. How does this causal relation change with a different type of instruction?

In terms of learning, both studies consider an objectivist view of learning. This is a class of cognitivist learning theory that considers knowledge as an independent entity irrespective of the individual learner (Hannafin, 1997; Phillips, 1998). The task for the learner is to recognize, organize and integrate the new learning objects and events with existing knowledge (Hannafin, 1997; Jonassen, 1991). Within the objectivist view of learning the emphasis is on well defined learning objects (Mergel, 1998; Phillips, 1998). One of the main strengths of such an approach is the ability to address novice learning (Phillips, 1998). In this contribution, these are learning the content from the video in the first study and learning to debug a code in the second study.

We measured the learning performance in both the studies. In the video-based learning study the performance was measured by the difference between the pretest and the posttest scores (learning gain). While in the debugging study, the performance was simply the number of bugs rectified in the code by the students. In both the studies, obtaining the higher performance required the students to use the previously acquired knowledge (in case of the video-based-learning, recent; and in the case of debugging, practice-based) to solve the given problems (rectifying the bugs or answering the questions). Table 1 indicates that the performance slightly varies across the two different studies, but with no statistically significant difference. A chi-square test between the two normalised distributions (using MinMax normalization) shows that there is no significant difference in the two performance measures ($\chi^2 = 45$, $p = 0.34$). Moreover, the mean values and their standard deviations depict that there was a healthy distribution of the cognitive performance in each of the tasks (i.e., we did not have a very difficult or very easy task). Another commonality between the two tasks is that for the user to attain high cognitive-performance score, students needed to devote the required levels of attentional and cognitive processing.

To test relationships between variables in HCI, researchers have adopted particular statistical tools. Most widely used is the frequentist null hypothesis significance testing. Similarly, researchers have embraced various quantitative standards, such as p-values and dichotomous testing procedures, which have ultimately proven to be poor at investigating bi-directional, simultaneous, and continuous relationships (Dragicevic, 2016; Dunlop and Baillie, 2009). A new practice of analysing the causal relations between observable variables is becoming popular in scientific domains such as neuroscience (Ding et al., 2006; Goebel et al., 2003), user consumption (Narayan and

Table 1
Table showing the descriptive statistics of the two learning performance measures. Dip test was used to test the bi-modality of the two distributions (with the null hypothesis of having uni-modal distribution).

Indicator	Study 1	Study 2
Mean	3.35	4.42
SD	1.87	3.13
Median	3	4
Bimodality test Statistic	0.10	0.08
Bimodality test p-value	0.0001	0.001

Smyth, 2005), stock market (Hiemstra and Jones, 1994), economics (Joerding, 1986; Thornton and Batten, 1985), and also emerges in HCI (Kirk et al., 2016; Ziabari and Treur, 2018).

4. Granger causality: Approach

Hill (1965) (Bradford Hill, 1965) provided certain criteria for any two time series to be considered having a causal relationship between them. These are regarded as the empirical conditions for causality between two time-series. In this section, first we will first explain these 'empirical' conditions for the causality between two time-series. Then, we will introduce Granger's definition of causality, and further show how and to what extent this definition satisfies conditions for causality. Finally, we will show how Granger's causality can be extended to a group of participants.

4.1. Hill'S criteria for causality

Bradford Hill (Bradford Hill, 1965) proposed the following criteria for causality between two observational time-series, stating that If a set of necessary and sufficient causal criteria could be used to distinguish causal from non-causal associations in observational studies, the job of the scientist would be eased considerably. With such criteria, all the concerns about the logic or lack thereof in causal inference could be forgotten: it would only be necessary to consult the checklist of criteria to see if a relation were causal – Hill, 1965 (page 1) (Bradford Hill, 1965). Following is the list of the conditions proposed by Hill:

- **Strength**: A relationship is more likely to be causal if the correlation coefficient is large and statistically significant.
- **Consistency**: A relationship is more likely to be causal if it can be replicated.
- **Specificity**: A relationship is more likely to be causal if there is no other likely explanation.
- **Temporality**: A relationship is more likely to be causal if the effect always occurs after the cause.
- **Gradient**: A relationship is more likely to be causal if a greater exposure to the suspected cause leads to a greater effect.
- **Plausibility**: A relationship is more likely to be causal if there is a plausible mechanism between the cause and the effect

4.2. Granger'S definition of causality

Granger causality (Granger, 1969) tests for the ability of one series to predict another one – in our case, whether information flow provides sufficient information to predict 1) user focus size, 2) cognitive load, and 3) user attention flow. Granger causality investigates bi-directional, simultaneous, and continuous relationships and has been employed to several studies in HCI (e.g., (Abdullah et al., 2015; Madan et al., 2010)). The basic definition of Granger causality has two assumptions (Granger, 1969). First, it assumes that the cause occurs prior to the effect. Second, the cause contains information about the effect that is more important than the history of the effect itself. Although Granger causality is defined for linear and stationary time-series contexts, variations for non-linear (Ancona et al., 2004; Chen et al., 2004) and non-stationary (Ding et al., 2000; Hesse et al., 2003) data exist.

The main idea behind Granger's definition of causality is that if the lag (past values) of variable one predict the current value of variable two in a better manner than the lags (past values) of the variable two itself, we can infer that variable one causes variable two. To arrive at such an inference, there is a simple method to be followed. Let us take the case of two variables X and Y. To determine whether X Granger-causes Y or the other way around, we create two models. The first model predicts the current value of Y using the past values of Y (Eq. 1),

while the second model predicts the current value of Y using the past values of X (Eq. 2). We then compare the quality of the prediction for both models; if the second model outperforms the first model, we infer that X Granger-causes Y.

To conduct the data analysis, we follow a number of statistical steps. First, we perform data treatment. We divide the dataset comprising of information flow, user focus size, cognitive load, and user attention change into 10 second windows for further analyses. We then test for stationary time series: a Ljung-Box test is used to determine whether there are significant non-zero correlation coefficients at lags 1–15. Small p-values suggest that the time series data is stationary. We also identify the optimum value for the 'lag': the number of previous data points considered for modeling the causality. The value is identified based on the Akaike information criterion (AIC) value of the model. We create models with different values of lag that has to be taken into account for the Granger causality consideration and select the model with the lowest AIC value.

Next, we test for Granger Causality (Granger, 1969) to examine the causality between the different variable pairs (information flow – user focus size; information flow – cognitive load; information flow – user attention change). As aforementioned, the basic principle of Granger causality is to compare two models to test whether x causes y. The first model predicts the value of y at time t using the previous n values of y. The second model predicts the value of y at time t using the previous n values of both x and y. The comparison of the two models can tell us whether the history of x contains more information about y than the history of y itself. If this is the case then we can say that x Granger-causes y.

$$y(t) = \sum_{j=1}^{p} \alpha_{11j} x(t-j) + \sum_{j=1}^{p} \alpha_{12j} y(t-j) + \varepsilon_1(t) \qquad (1)$$

$$y(t) = \sum_{j=1}^{p} \alpha_{22j} y(t-j) + \varepsilon_2(t) \qquad (2)$$

Where,

p = model order, maximum lag included in the model

α = coefficients matrix, contribution of each lag value to the predicted value

ε = residual, prediction error

One might argue about the choice of our method to analyze the causality between the different pairs of measurements. In this paper, we used the definition of causality provided by Granger. There are three other methods that could be used to show the causality between different variables: 1) Structured Equation Modelling (SEM, (Edwards and Bagozzi, 2000)) 2) Cross-convergent mapping (CCM, (Sugihara et al., 2012)) and 3) conducting an intervention experiment where the hypothesized 'cause' is controlled and the hypothesized 'effect' is measured (Shadish et al., 2002). SEM does not necessarily contain the information required to consider a causal relationship. Statistically speaking, testing a SEM is not a test for causality. There are certain mathematical formulation under which SEM can be used for causal inference (Steyer, 2013; Steyer et al., 2002); however, the solutions are not available commercially. Bollen and Pearl (Bollen and Pearl, 2013) provide a detailed account describing how SEM should not be used for modelling causal relations between variables.

The second method, that is, CCM is useful only in the cases where the time series is stationary (i.e., mean and variance of the variable do not change over time) and non-linear (i.e., there is no auto-correlation in the time series). Eye-tracking data is stationary (as revealed by the Ljung-Box test) but auto-correlated (where users look at current time instance vastly depends on where they were looking at previous instances). This is why CCM is not an adequate method for such data.

In the case of identifying causal relations between two variables through an experimental or pseudo-experimental setup, such setups are typically costly or require an extensive duration in order to identify the cause-effect relationship between the two variables in question (Chambliss and Schutt, 2018). Moreover, it has also been shown that for longer time series data the Granger causality outperforms other contemporary methods (Zou and Feng, 2009).

4.3. Intersection between hill and granger

Granger causality satisfies a subset of Hill's criteria, such as strength (selecting the model that is more explainable), consistency, temporality (modelling the present value of hypothesised effect based on the lags of the hypothesised cause), plausibility, and coherence (the relations can be backed by the theory and behavioural explanations). Experiment and Analogy are contextual. For example, during learning sessions, a suggested (Granger) causal relation can be tested using an intervention experiment while testing for Granger causality in analogous contexts is possible based on the temporal data collected. Finally, the 'gradient' criteria cannot always be satisfied by the Granger's definition for causality because there is no guarantee that including more lags (a longer history) from the suspected cause will increase the predictability of the present value of the suspected effect.

## 5. Our proposed method: Combining data from more than one participant

Granger's definition and analysis of causality is conducted on a pair of time-series. Consider two time-series "X" and "Y". "X" is said to Granger-cause "Y" if the past of "X" predicts the future of "Y" more efficiently than the past of "Y" itself. Once we have established that "X Granger-causes Y", we require a measure of "by how much?" This is computed as the efficiency of the model predicting the future of "Y" using past of "X". At this point, one might ask what if the past of "Y" predicts the future of "X" as well? In such case we consider the model with higher efficiency and to answer "by how much", we use the difference of the two efficiencies.

One contribution of this paper is to show how one could use the same analysis proposed by Granger to establish a causal relation between a pair of variables measured for one ecosystem. In the case of human studies, each participant could be considered as an ecosystem producing the measurements. This requires an additional level of analysis that builds upon the results of Granger causality analysis for each individual. To be able to compare groups of participants, we need a way to represent the individuals on a Cartesian space (at least 2D). This has a requirement of computing two variables from the Granger causality analysis of each individual participant. Among the options for these two variables are the effect size and the significance of the Granger causality test for each individual participant.

For the effect size of the causal models, we calculate partial η^2. The difference between the η^2 values of the two models (x causes y and y causes x) gives us the overall effect size for the causal relationship between x and y (Section 4.2). The difference in the partial η^2 values for the two models tells us about how much more the variance is explained by chosen causal model over the other mode. This could be an indication of the 'effect size' of the causal relation between two variables.

Next, the difference from the linear inter-dependency (correlation) is calculated by creating a simple linear model to measure the linear dependency (i.e., correlation) between two variables. Again, we use partial η^2 to measure the effect size of the linear relation of two variables. One of the necessary conditions for the causality between two variables is that they should be correlated. Hence, the difference between the causal and the correlational model would depict the significance of the causal relation. The difference between the partial η^2 values of a linear model and the causal model would indicate how much more the variance can be explained by the causal model as compared to the correlational model.

Fig. 1 shows how each individual could be represented as a 2D point on the 'effect size' – 'significance' Cartesian space. To identify potential
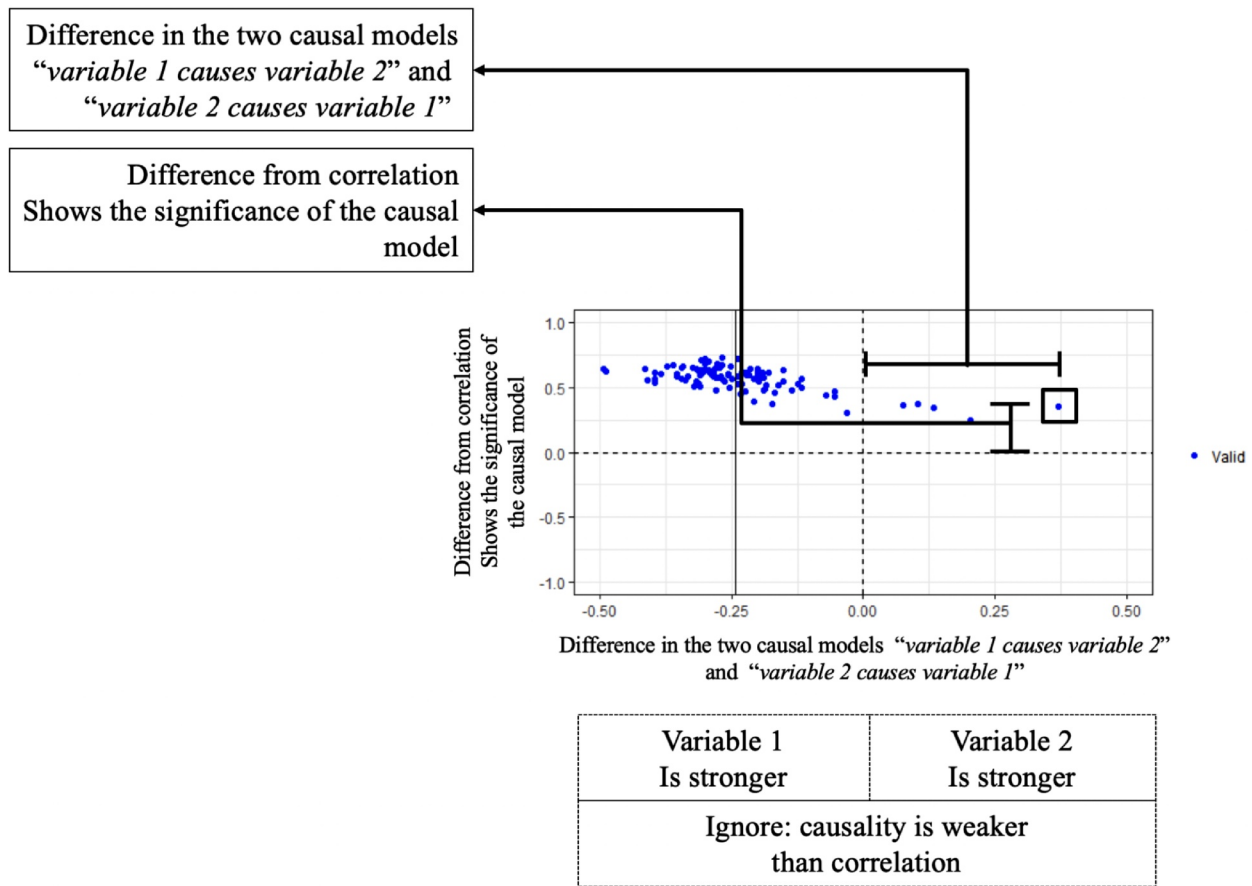
Fig. 1. A visualization of our approach to summarize causality results for multiple participants. For each participant we calculate the difference between the two causal models ('x causes y' and 'y causes x') and the respective difference between causal model and correlational model. These two values become the (x,y) coordinates of a data in our scatterplot. All retained values (blue dots) are used to calculate the x-axis mean, which represents the mean direction of interaction between variables. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

significant differences between study 1 (passive task) and study 2 (active task), and the high/low performance/learning gain, we compare the causal relationships for the two studies. Once we have the effect size of the causal models for each study, we compare them using a Wilcoxon test. We use a non-parametric test here because there is no theoretical or practical basis for assuming that effect sizes would follow a known statistical distributions (e.g., Gaussian, Poisson, Student-t). Using this method we could combine multiple pairs of time-series into one analysis, which constitutes the main contribution of the paper. This method provides a unique way to combine empirically collected time-series data from multiple users while analyzing the causal relationships between the two time-series for each users.

In particular, we apply the proposed method to respond to the following Hypotheses (H) that are stemming from the RQs described in Section 3.

- H1a: The information flow drives the learners' focus size.
- H1b: Learners' focus size drives the information flow.
- H1c: The casual relationship between learners' focus size and information flow is different between video-based learning and debugging.
- H2a: The information flow drives the learners' cognitive load.
- H2b: Learners' cognitive load drives the information flow.
- H2c: The casual relationship between learners' cognitive load and information flow is different between video-based learning and debugging.
- H3a: The information flow drives the learners' attention change.
- H3b: Learners' attention change drives the information flow.

- H3c: The casual relationship between learners' attention change and information flow is different between video-based learning and debugging.

6. Measurements

Eye position was measured in x-y coordinates of the display monitor using the SensoMotoric Instruments (SMI) RED 250 tracker. The device was mounted to the bottom of the computer monitor used by participants. It operates at a distance of 60–80cm and has a high accuracy of 0.5 degrees. The contact-free setup of the eye tracker allows for free head movement of 40cm x 20cm at a 70cm distance. We calculated the following measurements from the eye tracking data and the screen recordings: Information Flow, Cognitive Load, user focus size, and Attention Change.

6.1. Information flow

Information flow (i.e., stimulus entropy) was computed for each frame of the screen recording in Study 1 and Study 2. This indicates, in a direct manner, the amount of information transmitted to the student. To compute the stimulus entropy for each frame (Eq. 3), using a window of 10 seconds, we have used the three separate grey images (one each for red, green, and blue channels). This gives us three 2D arrays of values between 0 and 255. We then compute the Shannon entropy of these three arrays using the following formula. This is a widely used method to compute RGB image entropy in image processing applications (Gonzalez and Woods, 2007).

$$Entropy\ H(X) = -\sum p(X)\log p(X) \qquad (3)$$

Where X is the pixel intensity (between 0 and 255) and $p(X)$ is the probability of finding a pixel with intensity X.

The mean entropy[4] of the three arrays gives us the stimulus entropy (i.e., Information Flow). As indicated before, Shannon entropy is a direct measurement of the information content of the communication medium. This can be seen in the different cases shown in the Fig. 2 for the first study and in the Fig. 3 for the second study. We can see how the amount of information present on the screen changes with the content of the screen. The entropy values increase in both the case of video (Fig. 2) and IDE (Fig. 3) from top to bottom panels, as it can be also seen from their color histograms, which show the information content on the screen.

The basic intuition behind Shannon entropy (since it is key to understanding the information flow) is that an unlikely event contains more information than a likely event. In the case of screen's entropy, the blank screen corresponds to events for which all the pixel values are the same. Therefore, the blank screen will have least entropy and by extension the least information content. Contrary to the blank screen, a screen full of text and graphics will have a lot of different pixel values and having a given combination of pixel values (which creates the content on the screen) is highly unlikely. Therefore, a screen filled with content will have high entropy and thus more information content.

6.2. Cognitive load

We use the same definition of cognitive load as Pass et al (Paas et al., 2003), i.e., mental load (interaction between task and learner characteristics), mental effort (capacity allocated due to the demands of the task), and performance (learner's achievements). To measure the cognitive load of participants (Eq. 4), we calculate the four measures proposed by Buettner (Buettner, 2013) (i.e., mean pupil diameter, pupil diameter standard deviation, saccade speed, and number of fixations longer than 500ms) for every participant. This way we can estimate cognitive load as described in the existing literature (Prieto et al., 2015a; 2015b; Schultheis and Jameson, 2004; Szulewski et al., 2014). These measures are then combined into one quantity to depict the cognitive load of the students. Pupil diameter (mean and SD) and number of long fixations contribute positively to the unified measure while the saccade speed contributes negatively. Moreover, for the pupil diameter-based variables (Mean and SD), we use the first 10 seconds of each participants' data to normalize all the data for that particular participant. This is done to avoid bias caused by individual properties of the participant and the environmental properties, as they can severely impact the pupil diameter. For example, age, time of the day, and caffeine levels have been shown to affect measurements (for details see (Holmqvist et al., 2011). Based on the results of Buettner (Buettner, 2013), we compute cognitive load using the following formula. This formula was also used in Giannakos et al. (2020); Prieto et al. (2017).
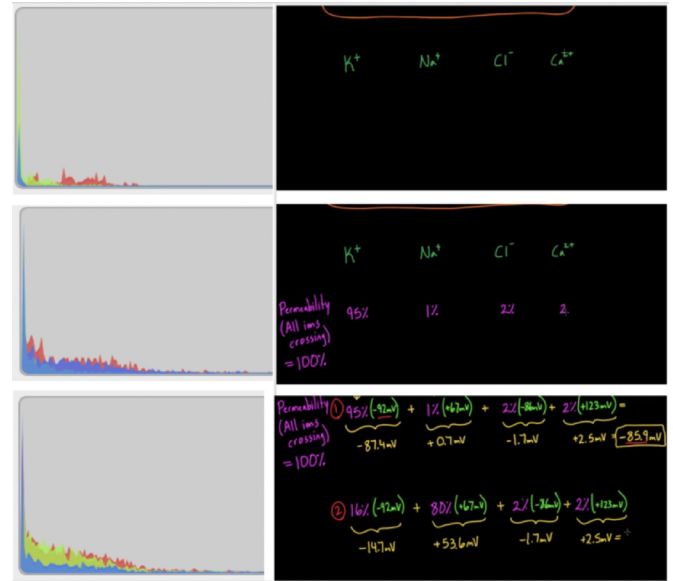

Fig. 2. Typical examples for stimulus information flow calculation (video-based-learning-study). We can visually notice that the top figure has the least amount of information and the bottom figure has the most amount of information. This can also be observed in the respective histogram representations next to each image. This figure shows that Computing the entropy of these images can give us an estimate of the information presented on the screen.

$$\left.\begin{cases} 0\ if\ PDM \le median(PDM) \\ 1\ if\ PDM > median(PDM)\} \\ \qquad + \\ \{0\ if\ PDS \le median(PDS) \\ 1\ if\ PDS > median(PDS)\} \\ \qquad + \\ \{0\ if\ NLF \le median(NLF) \\ 1\ if\ NLF > median(NLF)\} \\ \qquad + \\ \{1\ if\ SS \le median(SS) \\ 0\ if\ SS > median(SS)\} \end{cases}\right\} \qquad (4)$$

Where PDM = Pupil diameter mean, PDS = Pupil diameter SD, NLF = number of long Fixation, SS = saccade speed. This formula is obtained from the affinity that the following measures display with the cognitive load: mean pupil diameter, pupil diameter standard deviation, saccade speed, and number of fixations longer than 500ms[5]. These gaze measurements show the highest discriminating power when it comes to cognitive load measured using a dual task (a dual task is a procedure in which participants are given one task with increasing difficulty as well as another basic task, and cognitive load is measured as the rate of mistakes in the basic task (Brünken et al., 2002)).

6.3. User focus size

To measure user focus size we consider gaze entropy (Eq. 5). We overlay a 50-by-50-pixels grid on the screen, and we consider the whole experimental session in 10-second time windows. Then, we compute the proportion of time spent in each window looking at each block on

---

[4] We also analyzed the data using the sum of the entropy of the three channels, and the overall results remain consistent. The main reason for similar results using both methods could be because of the scale invariant nature of the Granger causality (using the sum instead of the average just increases the range of the measurement).

[5] In addition, we took the weighted average of the four values (using the inverted values for saccade length, since it has negative impact on the cognitive load) and re-analyzed the data. The results remain consistent as before. The main reason for similar results using both methods could be due to the fact that the linear regression in continuous case would be replaced by a quantile regression in the discrete case. In both the cases the differences between various models should not change drastically.

Fig. 3. Typical examples for stimulus information flow calculation (debugging-study). We can visually notice that the top figure has the least amount of information and the bottom figure has the most amount of information. This can also be observed in the respective histogram representations next to each image. This figure shows that Computing the entropy of these images can give us an estimate of the information presented on the screen.

the grid. The process results in a two-dimensional proportionality vector for each window. Consequently, the gaze entropy is calculated as the Shannon entropy of the proportionality vector. The decision for using the 10-second window for computing the proportionality vector is inspired from previous work (Prieto et al., 2015b; Sharma et al., 2017; 2018a).

$$user\_focus\_size = -\sum_{i=1}^{M*N} p_i \log(p_i) \qquad (5)$$

where, p_i is the proportion of time spent looking at the i_{th} block on the grid that has M rows and N columns.

This measure tells us about the focus size of the participant. A value of 0 indicates that the participant is looking at only one block on the grid during the specified window. Hence, we measure the level of uncertainty of a random variable: the objects looked at by the participants. Theoretically, the highest possible entropy value is the logarithm of the number of blocks in the grid, in our case 2.76. This is the maximal value that would indicate a uniform distribution of gaze time over the grid. Thus, a high entropy indicates that the participant was looking at a wider range of objects on the screen; in other words, the participant had a broader focus area or 'user focus size'.

Finally, it is worth mentioning that the 'user focus size' is not related to 'attention level'. The user focus size merely captures the number of objects the participant is looking at during a fixed time window. Fig. 4 shows two typical examples (Contrasting case) of entropy values. This
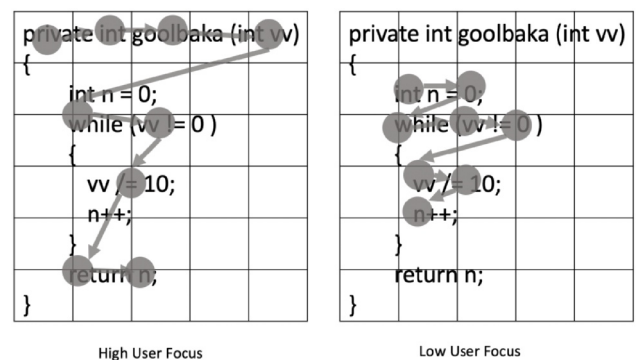

Fig. 4. Typical examples for user focus size calculation. In this example the screen contains the example code and the rectangular grid is overlaid on the screen. The circles are the fixations, the arrows are the saccades and the direction of the arrows shows the direction of the gaze movement. Both the panels represent the gaze of two learners in the same time window. In the left panel, we can observe that the learner's gaze covers more content than the learner's gaze in the right panel. Therefore the focus size of the learner on the left panel is higher than the focus size of the learner on the right panel.

measure has been used previously to quantify visual focus (Olsen et al., 2018; Sharma et al., 2017; 2018b). We can see in the left panel of Fig. 4 that the user is looking at almost everything present on the screen (i.e.,

a large number of elements on the screen) and hence the user focus size would be high in this case because the Shannon entropy of the part looked at by the user would be high. Conversely, in the right panel of Fig. 4 the user is looking at a limited number of elements on the screen, therefore in this case the user focus size would be low. We would like to empathizes at this point that we do not claim any relation (present or absent) between attention and user focus size.

6.4. User attention change

To calculate changes in users' attention we consider gaze stability. Again, we utilize the grid and proportionality vectors used to compute the user focus size. The only difference is that gaze stability is computed for pairs of consecutive 10-second windows. Gaze stability represents the similarity between the objects looked at by the participant across two consecutive windows (Sharma et al., 2018b). The inverse of the gaze stability represents 'User Attention Change'. In our analysis we compute the cosine similarity between proportionality vectors of the two windows. The value ranges from 0 to 1 (*while inverting we take precaution to not divide by zero*). A stability value of 1 indicates that the participant was looking at the same set of objects during two consecutive windows (i.e., no user attention change), whereas a stability value of 0 indicates that the participant was looking at a completely different set of objects during two consecutive windows (i.e., complete user attention change). Fig. 5 shows the contrasting cases for the User attention change computation. On the top panels we can see that in the two consecutive time frames the user is looking at different set of objects (in other words the user is gazing upon different sets of grids during two consecutive time frames). In this case the cosine similarity value for the 2D proportionality vector will be low and the user attention change will be high. Whereas, in the bottom panels of Fig. 5, there is a considerable overlap in the objects looked at (or the grids gazed upon) during the two time frames. Therefore, the cosine similarity value will be high and the user attention change will be low.
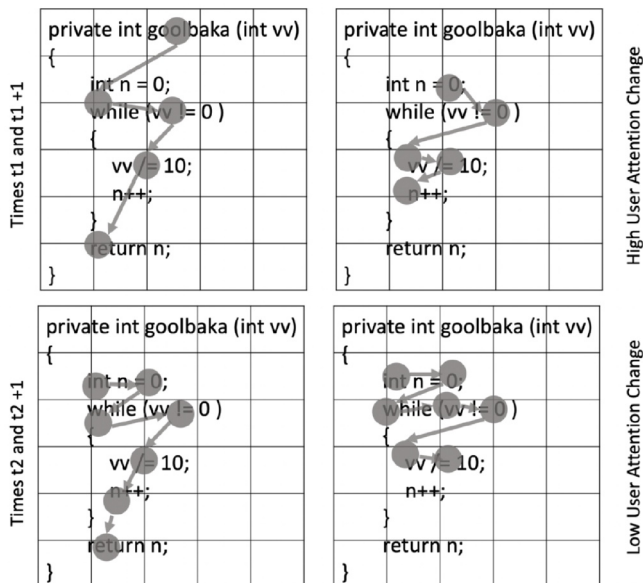

Fig. 5. Typical examples for user attention change calculation. In this example the screen contains the example code and the rectangular grid is overlaid on the screen. The circles are the fixations, the arrows are the saccades and the direction of the arrows shows the direction of the gaze movement. The top and bottom panels show gaze patterns of two learners in two consecutive time windows. We can observe that the intersection of the items looked at in two time windows is higher in the bottom panel than in the top panel. This shows that the user attention has higher change in the top panels than in the bottom panels.

7. Study 1

The first study is contextualized in a video-based learning environment (i.e, a typical MOOC). In particular, watching a video is a passive task, although the information presentation with respect to time is dynamic. This learning setting requires participants to understand and memorize information using their auditory, visual, and cognitive capacities in their working memory. However, this environment does not engage participants in higher-order thinking, such as apply, analyze, and evaluate, nor do they receive any immediate feedback. Moreover, once the participant decides to be exposed to particular information utilizing a video, they can exercise little control over that information. In other words, the participant has no control over the content of the information, or the mode through which the information is conveyed. The only control the participant has is by pausing, going forward, or going backwards in the video at a given moment.

7.1. Participants

A total of 98 students from a European university participated in a lab study (mean age = 24.68, sd = 3.09, 78 males, 20 females). Participants individually watched two digital drawing board (Khan-academy style) videos on "resting membrane potential" (see Fig. 6), a topic on which they had little to no prior knowledge. The combined length of the videos was 17 minutes and 5 seconds. While watching the videos, the participants had full control over the video player. The videos were downloaded and were shown to the students using a video player on a local machine to remove the attention deficit caused by viewing the videos on YouTube. The participants had no time constraint as to the completion time of watching the videos.

7.2. Procedure

Upon arrival at the laboratory, participants signed an informed consent form. After this, and prior to the video watching task, each participant had to pass an automatic eye-tracking calibration routine to accommodate the eye tracker's parameters for each participant's eyes to ensure accuracy in tracking the gaze. Participants' gaze during the debugging task was recorded using an SMI RED 250 eye-tracker at 250Hz. Next, participants were asked to answer a pre-test, which required answering 10 questions about the video content. After this task, participants were given up to 30 minutes to watch the video with full control over the video-playback. The participants spent an average of 20 minutes and 35 seconds (sd = 4 min. 3 sec.) with the videos. Once the participants finished watching the videos, they answered a post-test questionnaire containing 10 questions. For their participation in the experiment, participants were rewarded with an equivalent of USD 30.
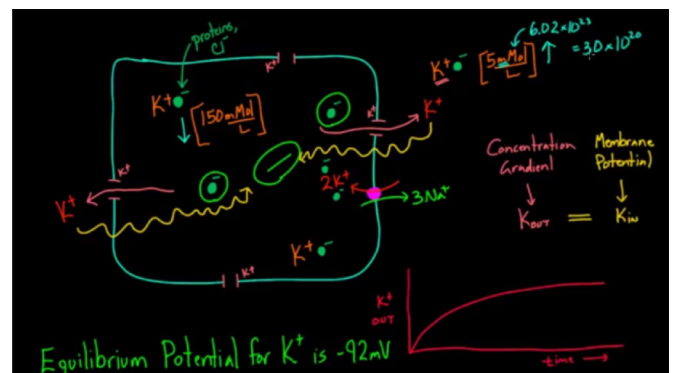

Fig. 6. Screenshot of the video presented to the users in the first study based on the video based learning paradigm.

7.3. Learning outcome – learning gain

Students answered a pre-test before watching the video content and they answered a post-test after watching the video content. The learning gain was calculated simply as the difference between the individual pre-test and post-test scores. The minimum and maximum for each test were 0 and 10, respectively. There was a floor effect for the pretest (mean = 0.87, median = 0, SD = 1.1). This is why we chose to use the simple difference between the posttest score and the pretest score as the learning gain. The mean for the learning gain was 3.35 (SD = 1.87). A bi-modality test (dip-test[6]) revealed that the learning gain distribution was bi-modal (D = 0.10, p = 0.0001). Therefore, once we obtained the score for the learning gain, we used a median split (median = 3) to distinguish between 'high' and 'low' levels of learning gains.

7.4. Analysis

To answer RQ1 in the context of video based learning, we conduct the Granger causality analysis as presented in the Sections 4.2 and 5. To answer RQ2 in the context of video based learning, we compare the students with high and low learning outcomes (learning gains in this case) using the Wilcoxon test because the effect sizes are not normally distributed. Table 2 shows the pairwise correlation among the measurements and Fig. 7 visualizes the measurement pairs.

7.5. Results

We investigated whether the information flow is controlled by the user focus size or the other way round. For this, we analysed the relation based on Granger causality between information flow and the user focus size in Study 1. The results are shown in the middle panels of Fig. 8 (left panel). From the left panel of Fig. 8 we can see the mean of the effect size, i.e., the difference between the two causal models: 1) information flow Granger-causes the user focus size, 2) User focus size Granger-causes the information flow. This mean is -0.24 (SD = 0.13), which indicates that there is a stronger support for the hypothesis that the information flow Granger-causes the user focus size (**Accepting H1a**) than the support for hypothesis that the user focus size Granger-causes the information flow (**Rejecting H1b**). Therefore, in the context of video-based learning we reject the hypothesis H1b and accept the hypothesis H1a. Meaning that in video-based learning context (common way of learning in MOOCs and other e-learning settings) the way information is presented to the learner (i.e., information flow) drives learners "reading" of the materials provided (i.e., user focus size).

Second, we investigated whether the information content of the stimulus (i.e., information flow) is controlled by the users' cognitive load or the other way round. For this, we examined the Granger-causality relation between information flow and cognitive load Study 1. The results are shown in the middle panel of Fig. 8. The solid line in the middle panel of the Fig. 8 shows the mean of the effect size i.e., the difference between the two Granger causality models: 1) information flow Grange-causes cognitive load, 2) cognitive load Granger-causes information flow. This mean is 0.25 (SD = 0.12), which shows a stronger support for the Hypothesis H2b (information flow drives the cognitive load) than the Hypothesis H2a (cognitive load drives the information flow). **Therefore, in the context of video-based learning we reject the hypothesis H2a and accept the hypothesis H2b. Meaning that video-based learning settings the load imposed by a given task (i.e., user's cognitive load) drives the way information is presented to the learning (i.e., information flow)** Examples of such a drive can be the learner pausing or stopping a video when they

---

[6] the dip test has the null-hypothesis that the given distribution is unimodal.

Table 2
Pairwise correlation for the different measurements for the video-based-learning study.

Variable 1	Variable 2	Study 1	
		Corr	p
Information Flow	**User focus Size**	0.96	0.0001
Information Flow	**Cognitive Load**	0.85	0.0001
Information Flow	**User Attention Change**	-0.22	0.02
User Focus Size	**Cognitive Load**	0.84	0.0001
User Focus Size	**User Attention Change**	-0.22	0.02
Cognitive Load	**User Attention Change**	-0.21	0.02

are cognitively overloaded or putting it on fast-forward when they are cognitively under-loaded.

Finally, we investigated whether information flow is controlled by the stability of the users' exploration (attention change) or the other way round. For this, we investigated the Granger-causality between information flow and the users' attention change in the first studies. The result is shown in the right panel of Fig. 8. The right panel of Fig. 8 displays a considerable number of violations (there are more invalid points than the valid points). This indicates that for most of the participants correlation is stronger than causation, that is, the participants for which the stimulus entropy and the gaze stability are not causally related. Therefore we cannot have proper evidence to investigate whether information flow is controlled by attention change or the other way round. In the case of video-based learning, we cannot accept or reject either of H3a or H3b.

Next we examine the nature of the above mentioned Granger-causalities (i.e., between information flow and user focus size, between information flow and cognitive load, and between information flow and user attention change) for the two levels of learning gains (i.e., high and low) for the video-based learning study (Study 1).

First, we observe that the causal relations are significantly different across the two levels of learning gains (W = 468, p < 0.0001). **In both groups (i.e., high and low learning gain) the information flow causes the user focus size. However, the Granger-causation is stronger for the students with low learning gain (mean = -0.29, sd = 0.03) than the causation for the students high learning gains (mean = -0.18, sd = 0.17). Therefore, we can support that the information presented to the learners drives the reading of low performers more compared to high performers.**

Second, we observe that the causal relationship between the information flow and the cognitive load are significantly different across the two levels of learning gains (W = 244, p < 0.0001). **For both high and low learning gain groups, the cognitive load causes the information flow. However the causation is stronger for the students with low learning gain (mean = 0.32, sd = 0.04) than the causation for the students high learning gains (mean = 0.15, sd = 0.12). Therefore, we can support that tasks load drives the information presented of low performers more compared to high performers.**

Finally, considering the nature of the causal relation between the information flow and user attention change, we observe (right panel of the Fig. 8) that there are many violations of causation (i.e, the correlation is stronger than either direction of Granger-causality). Therefore, we cannot conclude any causal relationship between information flow and user attention change for any level (high or low) of learning gains.

8. Study 2

The focus of the second study is contextualized through a problem-solving assignment (e.g., debugging code). In particular, writing code or debugging code is considered to be an active task, because the participant has the freedom to transform, integrate, and manipulate the
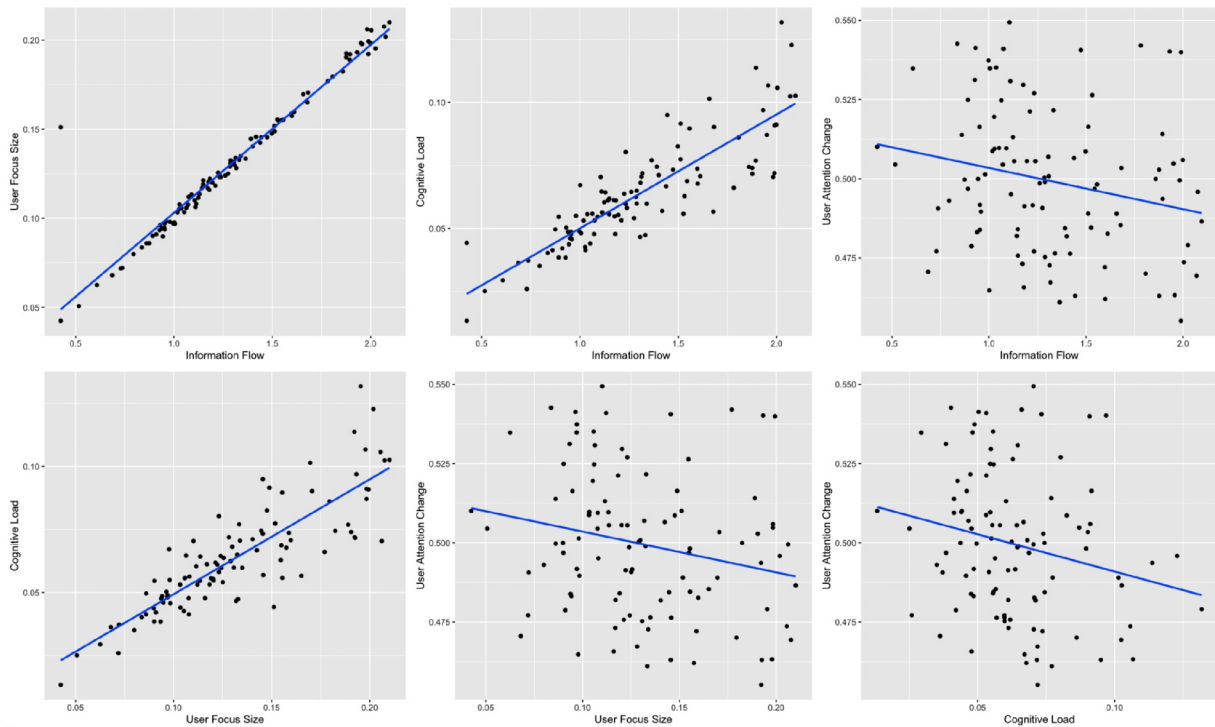
Fig. 7. Pairwise correlation plots for the different measurements for the video-based-learning study.
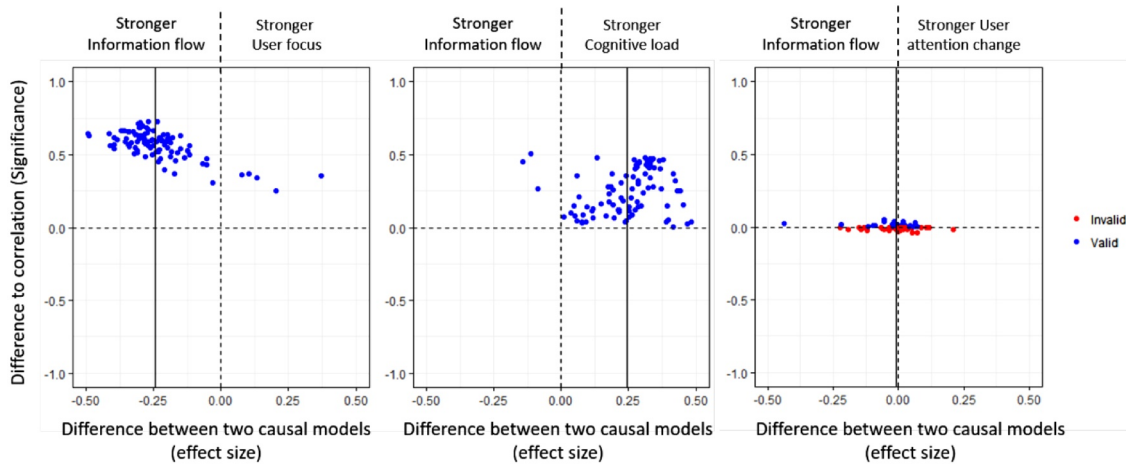

Fig. 8. Results from analyzing the relation between Information flow and the user behaviour (user focus size, Cognitive Load, Attention Change) from the video based learning study (Study 1). We can observe that in study 1, the information flow Granger causes the user focus size, the cognitive load Granger causes the information flow, while we do not observe any causal relation between information flow and user attention change.
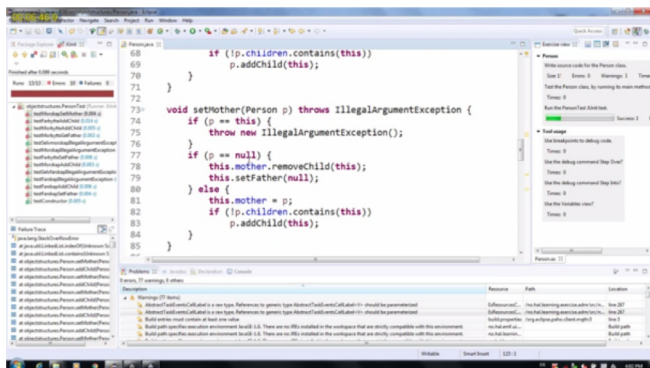

Fig. 9. Screenshot from a debugging session for the Study 2 based on the code debugging task.

content. However, the information presentation in this active task with respect to time is static.

The participants can exercise more control over the information content compared to the participants in the video watching task, and decide on their own which information they want to attend and process first.

8.1. Participants

The second study was performed at a contrived computer lab setting at a European University with 40 computer science majors (12 females and 28 males) in their third semester. The mean age of the participants was 19.5 years (sd = 1.65 years). In the previous semester, all of the participants had taken a Java course, for which they were predominantly using Eclipse as their integrated development environment (IDE). Moreover, they were also familiar with the built-in debugging

tool provided by Eclipse. The focus of this study is to examine how user-generated gaze data can be used to reinforce student reflective practices. Moreover, the study also considered whether students can practice problem-solving strategies rather than using trial and error.

8.2. Procedure

Upon arrival at the laboratory, participants signed an informed consent form. After this, and prior to the debugging task, each participant had to pass an automatic eye-tracking calibration routine to accommodate the eye tracker's parameters for each participant's eyes to ensure accuracy in tracking their gaze. Participant gaze during the debugging task was recorded using an SMI RED 250 eye-tracker at 250Hz. Next, participants were asked to perform a pre-task, which required removing 90 errors from a skeleton code within 10 minutes. After this task, participants were given 40 minutes to solve five debugging tasks presented as part of the main method of the main class of 100 lines of Java code. The code for the main debugging task contained no syntactic errors, and the participants were notified about this fact. For their participation in the experiment, participants were rewarded with the equivalent of USD 30.

8.3. Learning outcome – debugging performance

For the debugging task, there were 10 unit tests prepared by the instructor (see subsection Procedure). To limit the debugging to one of the panels of the Eclipse IDE (see Fig. 9), the researchers introduced a few bugs in otherwise complete code that would make the code fail all 10 unit tests. In order to pass all of the unit tests, the students were required to solve the debugging exercises in a particular order. Participants were given 40 minutes to complete the task. At the end of the 40 minutes, they were told to stop, and the number of unit tests passed at that point of time was taken to be the measure of 'debugging success' (i.e., performance). The mean for the debugging performance was 4.42 (SD = 3.13). A bi-modality test for the debugging performance distribution showed that the distribution is bi-modal ($D = 0.08$, $p = 0.01$). Therefore, once we calculate the debugging performance score, we used a median split (median = 4) to determine 'high' and 'low' levels of debugging success.

8.4. Analysis

To answer RQ1 in the context of code debugging, we conduct the Granger causality analysis as presented in the Sections 4.2 and 5. To answer RQ2 in the context of code debugging, we compare the students with high and low learning outcome (i.e., debugging performance) using the Wilcoxon test because the effect sizes are not normally distributed.

Table 3 shows the pairwise correlation among the measurements and Fig. 10 visualizes the measurement pairs.

Table 3
Pairwise correlation for the different measurements for the debugging study.

Variable 1	Variable 2	Study 2	
		Corr	P
Information Flow	User Focus Size	0.23	0.0001
Information Flow	Cognitive Load	0.17	0.005
Information Flow	User Attention Change	-0.02	0.73
User Focus Size	Cognitive Load	0.04	0.54
User Focus Size	User Attention Change	-0.02	0.71
Cognitive Load	User Attention Change	0.03	0.61

8.5. Results

We investigated whether the information content of the stimulus (i.e., information flow) is controlled by the information looked at by the user (i.e., user focus size) or the other way round. For this, we examined the causal relation between information flow and user focus size in Study 2. The results are shown in the left panel of Fig. 11. The solid vertical line shows the mean of the effect size, i.e., the difference in the two opposing Granger-causality models (first, information flow Granger-causes user focus size and second, user focus size Granger-causes the information flow). This mean is -0.05 (SD = 0.05). **This indicates that there is a little support for the hypothesis H1a (information flow drives the user focus size) but there is no support for hypothesis H1b (the user focus size drives the information flow). Therefore, in the context of debugging we reject the hypothesis H1b and accept the hypothesis H1a. Meaning that in the context of debugging (common way of learning in problem-solving and other algorithmic thinking settings) the way information is presented to the learner (i.e., information flow) drives learners "reading" of the materials provided (i.e.,user focus size).**

Next, We investigated whether the information flow is controlled by the cognitive load of the user or the other way round. For this, we investigated the Granger-causality between information flow and the cognitive load of the user in the debugging study. The results are shown in the middle panel of Fig. 11. The mean for the effect size is 0.02 (SD = 0.03). This is the mean of the difference in the two Granger-causality models: 1) information flow Granger-causes the cognitive load (**Hypothesis H2a**), 2) Cognitive load Granger-causes information flow (**Hypothesis H2b**). From the middle panel of the Fig. 11, we find little support for H2b (accepted) and no support for H2a (rejected). **Therefore, in the context of debugging we reject the hypothesis H2a and accept the hypothesis H2b. Meaning that in the context of debugging the load imposed by a given task (i.e., information flow) drives the way information is presented.**

Finally, we investigated whether information flow is controlled by how stable is the users' exploration (attention change) or the other way round. For this, we investigated the causal relation between information flow and the users' attention change in study 2. The results are shown in the right panel of Fig. 11, which contains a considerable number of violations, that is, the participants for which the stimulus entropy and the gaze stability are not causally related. This is because of the fact that the correlation is stronger than the causation for most of the participants (therefore more invalid points).

Next, we consider the nature of the three causal relations for the different levels of debugging success (i.e., successful and unsuccessful).

First, We observe that the Granger-causal relationship between information flow and user focus size are significantly different across the two levels of debugging success (W = 315, p < 0.0001). In both cases information flow causes user focus size. **The Granger-causation is again stronger for the unsuccessful students (mean = -0.11, sd = 0.05) than the causation for the successful students, which is almost non-existent (mean = -0.02, sd = 0.02).**

Second, we observe that the Granger-causal relationship between information flow and user cognitive load are significantly different across the two levels of debugging success (W = 75, p < 0.001). In both cases the cognitive load causes the information flow. **The Granger-causation is stronger for unsuccessful students (mean = 0.04, sd = 0.03) than the causation for the successful students, which indicates almost no causal relation (mean = 0.003, sd = 0.01).**

Finally, we consider the Granger-causality between the information flow and user attention change, for the two levels of debugging success. Since there are many successful students for whom the correlational model is better than the causal model (violations), therefore, it is difficult to conclude any concrete causal relation between information flow and user attention change for successful students in the case of
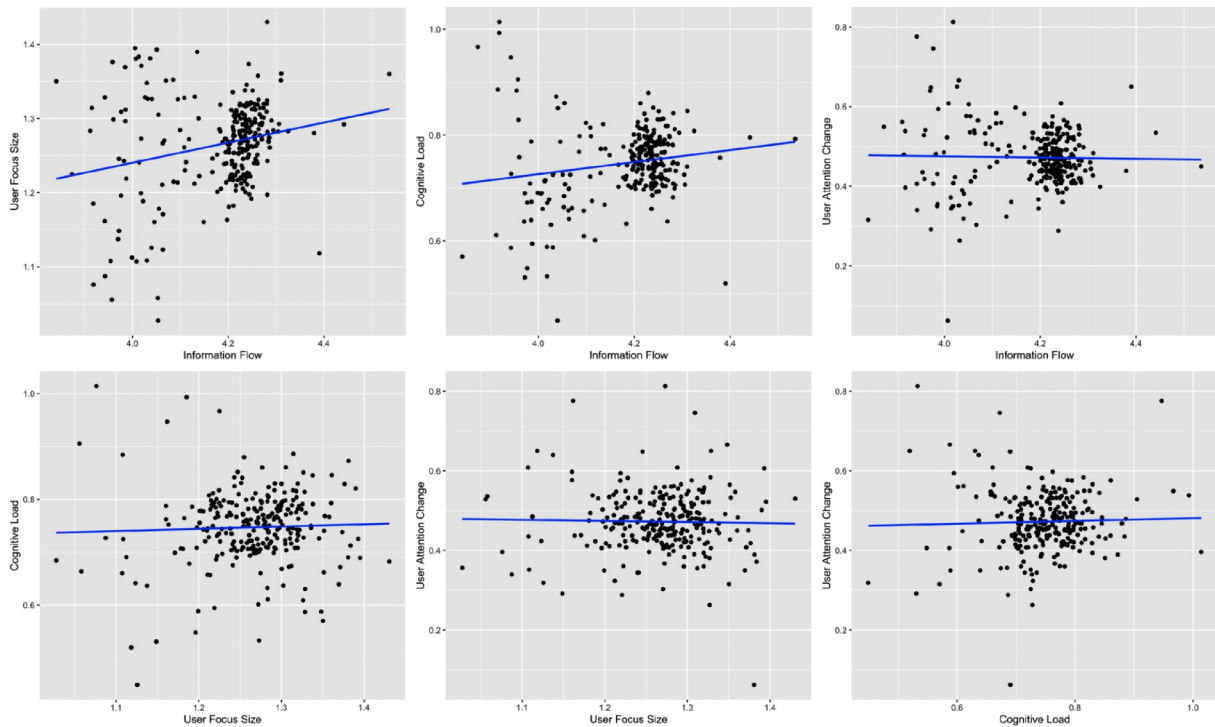
Fig. 10. Pairwise correlation plots for the different measurements for the debugging study.
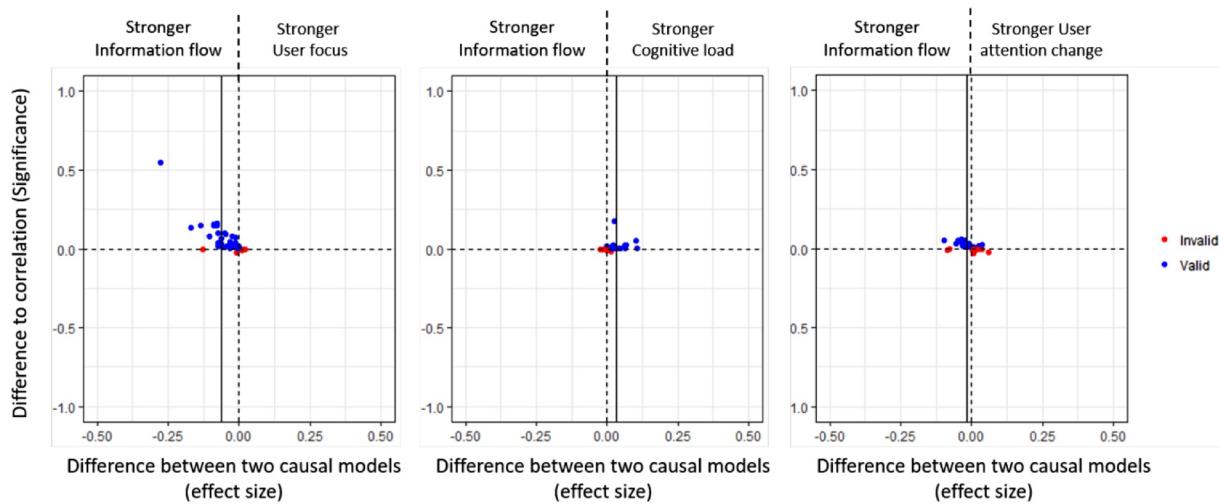

Fig. 11. Results from analyzing the relation between Information flow and user behaviour (user focus size, Cognitive Load, Attention Change) from the debugging study (study 2). We can observe that in study 2, the information flow Granger causes the user focus size, the cognitive load Granger causes the information flow, while we do not observe any causal relation between information flow and user attention change.

debugging.

9. Comparing the causal relation across the two studies

In this section, we will compare the overall results from the two studies to address the **Research Question 3**. To do so, we compare the causal relations from the two studies, i.e., video-based learning and debugging using the Wilcoxon test because the effect sizes are not normally distributed. Table 4 shows the lack of normality and homoscedasticity in the effect sizes of the causal models, which is the reason for using a non-parametric test here. The results suggest (Table 5) that *user focus size is driven by the information flow (H1a Accepted)*. The effect is stronger for the video watching activity (mean = -0.24, sd = 0.13) than for the debugging activity (mean = -0.05, sd = 0.05) (H1c

Table 4
Results from the Shapiro-Wilk's normality test for the effect sizes of the causal relations. The last two columns show the variances from the two studies.

	Shapiro-Wilk		Variance	
Causality relation	W	p-value	Study1	Study2
Information flow and user focus size	0.96	0.002	0.003	0.018
Information flow and Cognitive load	0.95	0.0001	0.0009	0.015
Information flow and User attention change	0.79	0.0001	0.001	0.006

Accepted). A Wilcoxon test shows this difference to be statistically significant ($W = 3038$, $p < 0.0001$).

Furthermore, the results suggest that *the information flow is driven by*

Table 5
Summary of overall results from Study 1 and Study 2.

Variable1 drives Variable2		Video Lecture	Debugging
Variable1	Variable2	mean (SD)	mean (SD)
Information Flow	user focus size	0.69 (0.07)	0.92 (0.07)
user focus size	Information Flow	0.93 (0.10)	0.98 (0.02)
Outcome: Information Flow drives user focus size (H1a Accepted, H1b Rejected)		**-0.24 (0.13)**	**-0.05** (0.05)
Information Flow	Cognitive Load	0.79 (0.08)	0.98 (0.03)
Cognitive Load	Information Flow	0.54 (0.12)	0.95 (0.04)
Outcome: Cognitive Load drives Information Flow (H2b Accepted, H2a Rejected)		**0.25 (0.12)**	**0.02 (0.03)**
Information Flow	Attention Change	0.96 (0.07)	0.96 (0.03)
Attention Change	Information Flow	0.97 (0.03)	0.97 (0.03)
Outcome: non conclusive (H3a H3b not sufficient evidence)		**-0.01 (0.07)**	**-0.01 (0.03)**

the cognitive load of the user (*H2b Accepted*). Again, the effect is stronger for the video watching activity (mean = 0.25, sd = 0.12) than the debugging activity (mean = 0.02, sd = 0.03) (H2c Accepted). A Wilcoxon test shows this difference to be statistically significant (W = 175, p < 0.0001).

Finally, the results *do not confirm any causal relation between user attention change and the amount of information flow*. This is similar for both the video watching activity (mean = -0.01, sd = 0.07 and the debugging activity (mean = -0.01, sd = 0.03) (H3c not sufficient evidence).

10. Discussion

In this paper, we explore the causal relations between information flow (as measured via screen recording) and learners' cognitive load (and consequently learners' behaviour, that is focus and attention change). The cognitive load and behaviour were measured using eye-tracking technology. Our results reveal two causal relations: first, between information flow and user focus size; and second, between information flow and user cognitive load (**i.e., RQ1**). Moreover, these relations have different strengths for different learning outcomes (**i.e., RQ2**) and the different instruction types; such as, active debugging versus passive video watching (**i.e., RQ3**). In this section, we provide the interpretation of the results, implications and limitations of this contribution.

10.1. Interpretation of the results

In the first causal relationship, the information flow drives user's focus (**i.e., RQ1**). In terms of information processing, we can say that the amount of information present on the screen drives the amount of information received by the user. This causality is more evident in the video watching task than in the debugging task. One possible explanation for this distinction could be the fact that in the video watching activity, the user is passively receiving the information provided by the teacher (i.e., transmitter of the information). Whereas in the debugging task there is no explicit transmitter of information. It is worth mentioning that looking at certain sections of the screen does not necessarily indicate that the user is paying attention, even though the eye-mind hypothesis (Just and Carpenter, 1980) states that 'what we see is what we process'. In other words, user focus size does not necessarily equal user attention. However, in behavioural terms, there is a possibility that user focus size and attention are correlated.

The second causal relationship states that user's cognitive load drives the information flow (**i.e., RQ1**). At first glance, this might appear contradictory, because most of the multimodal learning literature (Brunken et al., 2003; Mayer and Moreno, 2003) views "the control of information flow" as a way to manage the cognitive load, due to the user's active role in information processing. However, our causal finding suggests otherwise. One possible explanation might root from the user's interaction patterns (e.g., video navigation, code editing). In

both studies the user had complete control over the screen (i.e., full video playback control, editable program). In order to verify this complementary hypothesis we quantified video navigation and code editing patterns. In the video watching task, we calculated the proportion of users who paused and/or went backwards in the video at a given time. As shown in the previous work these actions are correlated to the perceived difficulties (Li et al., 2015) and misunderstandings (Giannakos et al., 2015) during video-based learning. In the debugging task, we calculated the proportion of users who edited the code at a given time.

For both tasks, our analysis included a Granger causality check between interaction patterns and the average cognitive load of users. On the one hand, the results showed that cognitive load drives the navigation patterns in the video watching task (pause: F[92,-5] = 4.02, p = 0.002; backward: F[92,-5] = 2.59, p = 0.03)[7]. On the other hand, for the debugging task, we observed that the average cognitive load of users drives the proportion of users editing the code (F[252,-3] = 3.71, p = 0.01). This shows that when users are experiencing a high cognitive load, they choose to reduce the information flow by pausing (not letting the information flow to increase) or going backwards in the video (actually decreasing the density of the content and thus, reducing the information flow). In the debugging task, learners were isolating a particular piece of code from the rest of the code by adding blank lines before and after that code-snippet whenever they were experiencing high cognitive load (when they could not find the solution easily or when they were not able to understand the code). Thus, reducing the computed information flow of the screen. Since cognitive load has been related to the working memory and short term memory of the users (Paas and Van Merriënboer, 1994; Sweller, 2011), this might dictate the amount of information users want on the screen; given that in both studies the users had complete control over the screen.

Considering the level of learning gain and the nature of the causal relationships between cognitive load, information flow, and user focus size (**RQ2 – study 1**), we observed that both relationships (i.e., information flow and cognitive load, and information flow and user focus size) are stronger for the group of students with low learning gains. Similarly, in study 2, the level of debugging success (i.e., learning task performance) and the nature of the causal relationships between cognitive load, information flow, and user focus size (**RQ2 – study 2**), we observed that the both relationships are stronger for the low performing students. Moreover, the direction of causality between information flow and the user's focus change, changes for the two performance groups. We observe that information flow causes the user focus size change for the low performing students and the user focus size change causes information flow for the high performing students.

An understanding of these relationships across different learning gains/performance levels could be used as the basis of providing help to

---

[7] The negative second degree of freedom in the F-test represents the optimal lag value used in the Granger test (see Section "Analysis').

students while they are interacting with the learning technology. In future research, learners" states, such as the level of expertise in the interaction, could be added for more proactive and actionable feedback strategies. This feedback could be both personalised and adaptive, as well as cognitive. For example, an opportunity for the personalized and adaptive feedback arises from the different causal relations based on the learning performance/gains of the learners. For example, the information flow for learners with low learning gain/performance is "Granger causing" the user focus size in a stronger manner than it does for learners with high learning gain/performance. For such learners (low learning gains/performance), one can start manipulating the information provided to the learner in a given time interval (e.g., slow down the video playback or point at a smaller part of the program) so that a manageable level of user focus size (for the particular learner) could be maintained throughout the learning process.

Another opportunity for cognitive feedback (such as seen in (Zeichner, 2018)), roots from the fact that the cognitive load of the students with low learning gains/performance "Grange causes" the information flow more strongly than it does for the students with the high learning gains/performance. For such learners (low learning gains/performance), additional cognitive scaffolding is required during the learning process. Such scaffolding is usually created by providing more content-oriented helps, such as reflection of a result or explaining a concept (Chen et al., 2018; Sedrakyan and Snoeck, 2016; Wu et al., 2012). Such feedback suggestions might help learners' understanding of the concepts and reduce their cognitive load, allowing them to attain a high level of learning gains/performance (Van Merriënboer and Kirschner, 2017).

A benefit of using Granger causality to analyse the relation between two variables is that this type of causality is purely based on the predictive power of the hypothesized 'cause'. Once we have established the "Granger Causality" between two variables (e.g., information flow Granger causes user focus size), we can simply use the temporal modelling methods to "forecast" the information flow and predict the user focus size without the requirement of using eye-tracking to measure it. This also allows us to prepare the learning systems, even where eye-tracking is not available, to identify the moments where the aforementioned feedback is required.

Both studies show similar causal orientations (overall and the high/low learning performance/gain), although they reveal different effect sizes (Figs. 11 and 8). The causal relation between information flow and user's gaze (i.e., focus and cognitive load) is stronger (i.e., high effect size) for the video watching task as compared to for the debugging task (**i.e., RQ3**). This could be due the differences in the stimulus and task types as explained in the beginning of the discussion section. Thus, the effect sizes are significantly larger in the video watching task as compared to the coding task. We hypothesize that this is a result of the fact that information can be more easily controlled in the video than in a code editor. Furthermore, the information content of the program is more consistent with respect to time; which could explain the lower effect size for the causal relationship between the information flow and the gaze behaviour (i.e., focus and cognitive load).

As previously indicated, the visual tasks used in the two studies differ in the information presentation (i.e., static versus dynamic). Fig. 12 shows the information flow of the two stimuli with respect to time. The information flow of the video is dynamic, while the information flow of the program is almost static. Because of the nature of information presentation, the information flow in the video is gradually increasing (as the video starts with an empty display and gets constantly filled up with textual and schematic content), whereas in the debugging task the information flow immediately displays already written code, in which only a few changes are required to accomplish the participant's goal.

The change in information flow of the video is based on the teacher's presentation flow. In this case the information flow is gradually increasing as the video starts with an empty display and gets filled up
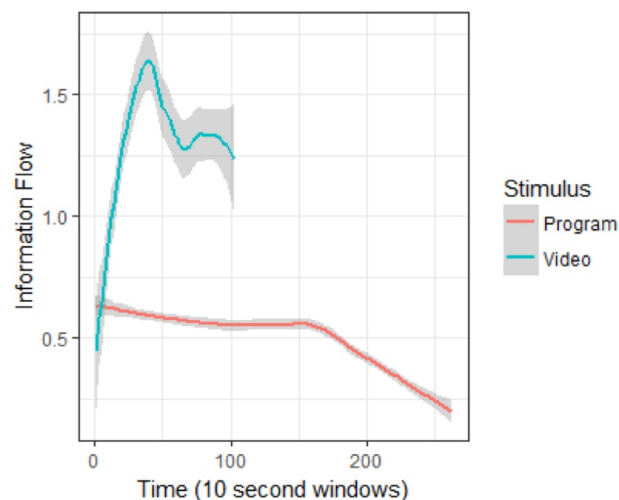

Fig. 12. Evolution of information flow over time, for the two stimuli used in the two experiments. The gray scale values show the 95% confidence interval for the different participants. We can observe that the information flow is more uniform in the program (Study 2) than that in the video (Study 1).

with the textual and schematic content of the topic. Conversely, the program is textual and displays already written code, due to the nature of the task in which only few changes need to be done by the user to accomplish the goal.

When it comes to information consumption, watching a video is passive consumption while debugging is an active task. Thus, the contextual set up in the two studies implies that different cognitive strategies are involved in completing the task. The main challenge in watching the passive video is to keep following the verbal reference of the instructor; whereas the main challenge in the active debugging task is to create a mental model of the code's functionality and rectify the mistakes. However, although debugging (active) is cognitively more challenging than watching a video (passive), the information flow is simpler (i.e., static) in the case of debugging as compared to watching a video (i.e., dynamic).

10.2. Implications

The implications for multimodal learning and instruction are two-fold. First, educators and designers should consider the limitations of user's working memory in terms of attending and processing stimuli, and design to avoid bottlenecks and high cognitive load for processing information. Failing to design activities which balance the focus on the concepts/skills to be learned with moderate levels of visually appealing and intuitive layout, might lead to less-than-optimal instruction that hinder learning (Mayer, 2005). In fact, our results supports previous research which describes that information flow drives user focus size, and not the other way around (Sweller, 1994; 2011).

Second, utilizing empirically-driven insights into the design of multimodal learning, such as the instruction type of the task (e.g., active vs passive transmission of information) and the information presentation with respect to time (e.g., static vs dynamic), could support educators to challenge their existing views of practice when designing content (based on instruction type and presentation style) and feedback (for example, the moments where the nature/strength of causal relationships changes). For example, this study presents a rather interesting finding that when users are experiencing high cognitive load, they tend to pause or go backwards during video watching (i.e., passive task) or they tend to alter the code and simplify the mental model of the code's functionality to correct the mistakes (i.e., active task). Moreover, the differences in the stimuli (e.g., video watching or program editing) and the instruction type of the task (i.e., active or passive) also show

differences in the effect size of the causality. In line with this, our study shows that by conceptualizing the relationships between various observable variables in a learning system (e.g., information flow, user focus size, attention change), one can improve the understanding of the complexity in the interaction between humans and technology, and use technology effectively and intelligently to support learning of concept or skill.

This understanding will help various stakeholders in the TEL domain, such as, teachers and researchers Sharma et al. (2018c, 2020); Sugihara et al. (2012). The teachers can create the content and the videos' timeline in a way that the student's focus size could be controlled. The teachers could also provide methods by which the students can control the information consumed by them in a better way than the current practices. Similarly, the TEL researchers could extend this work by introducing the temporal direction of causal relationship. It is yet to be investigate whether the causal relations stay the same for the whole interaction (i.e., the learning session) or it changes directions. Moreover, if the causal directions change over time, it would also be interesting for TEL researchers to investigate the salient features of the point in time where the causal direction changes (Sharma et al., 2020). Finally, the knowledge of causality would provide the TEL researchers to develop adaptive feedback tools that supports the students when the expected causal relations are not visible from the measured behaviour (Sharma et al., 2018c).

The strength and the nature of the causal relationship between the information flow and the gaze behaviour of the students differ for the two tasks presented in this paper. This indicates that the information presentation type acts as a mediator for such relationships. These findings can extend to other types of information presentation methods as well such as, textual slides, active text based debugging, or static content but varying difficulty of the problem. In the case of textual slides, if there is an option for the students to control for the certain aspects of the information flow, the causality directions should be generalizable to the video-based learning setting presented in this paper (Study 1). These controls include speed at which the information is being presented (e.g., navigation controls for a video). On the other hand further investigation would be required if the content is static without any controls for the students.

Concerning active text based debugging or writing a program from scratch, the results from the debugging study (study 2) should be generalizable to these contexts as long as the students could isolate (i.e., zoom-into) a part of the program to reduce the cognitive load and/or the focus size. Specially, for the active-script based debugging[8] the finding from the second study should extend to such scenarios. The main reason for this remains associated to the ability of the "debugger IDE" to be manipulated by the students/users, because such environments are useful in isolating the problem from the rest of the given structure (Marceau et al., 2004).

Finally, when it comes to static content with varying levels of difficulty, it requires further investigation because in such cases, the intrinsic cognitive load would change based on the individual problem. This was not the case for the either of the studies. The video explained one concept from the STEM fields (Study 1) and the unit tests were incremental (one provided the basis for the next one, Study 2). Thus, in both the studies the intrinsic cognitive load was controlled. Therefore, in cases where the intrinsic cognitive load varies in the same learning session, it would require further investigation.

10.3. Limitations

In both studies, participants were university students (both undergraduate and graduate). For a majority of digital education users (e.g., university students) this might be representative. However, it does not represent other user-groups (e.g., K-12 school students, professionals) who are also end-users of learning technologies. Moreover, the two studies were performed in highly controlled environments, producing high-quality datasets but low ecology. Therefore, the generalizability of our findings is somehow restricted by the selected tasks, because other tasks (e.g., collaborative learning, inquiry-based learning), or different representations of the same tasks (e.g., talking head in Massive open online courses' videos) might also affect the results. In particular, in our approach we applied two different tasks (i.e., video-watching and debugging) to portray an active and passive learning experience (i.e., the dependent variables). These two tasks present a good contrast in terms of the research questions asked, however, there are significant and yet subtle differences in these two tasks that might have affected the outcome (i.e., potential for confounding). Therefore, it is arguable that different tasks could have been used to portray an active and passive learning experience.

Consequently, this work considers the information flow of the stimulus as a dichotomous variable (i.e., static or dynamic). Another limitation comes from the analyses, where we solely consider eye-tracking data, although other behavioural aspects could have been computed. For example, exploring the semantics of user actions in the task (which sections of the video did the users go back to or paused, or extracting the semantics of the edited program). However, the results from the current contribution open new ventures for further investigations. For example, moving towards a holistic understanding by including other sensing modes such as, facial data, so that one can triangulate the findings. Moreover, connecting these data driven findings to theoretical bases of multimedia learning, for which more experimentation is needed.

Another limitation of this paper is that, with their primarily focus on presentation of and navigation through content, our two studies follow an objectivist view of learning. Objectivist view asserts that there is a particular body of knowledge that needs to be transmitted to a learner, and that learning is the acquisition and accumulation of a finite set of skills and facts (Tam, 2000). In fact, most of the contemporary learning systems follow a static and predefined representation of knowledge. They view knowledge as a thing that can be codified, captured, and passed along. Knowledge, however, is fluid and dynamic; and thus cannot be reduced to a merely conditional selection and sequencing of fixed and prepackaged content according to predefined rules and properties.

Finally, another limitation roots from the fact that one of the measurements, cognitive load, is derived from the pupillary response of the participants. Although, there are pre-processing steps carried out to remove the subjective and environmental bias, there might remain other noise due to head movements. The participants were not given the ophthalmological chin-rest. This might have resulted in a small deviation that might not be controlled for.

11. Conclusion

This paper presents results form two eye-tracking studies exploring the causal relationship between information flow and user behaviour. The results indicate that information flow drives users' focus, and that users' cognitive load drives information flow. The effect of the causal relations is dependent on the nature of the instruction of the learning material (i.e., active or passive). The causality is stronger for the passive transmission of information as compared to the causality for the active transmission of information. These results could inform design and feedback guidelines to achieve effective and efficient learner-computer interaction scenarios. Moreover, the results could also inform how to avoid bottlenecks and high cognitive load when users are engaged in information processing activities.

In future work, we aim to examine both theoretically and empirically how users' focus (as measured by entropy) is related to users' attention. Moreover, in the video watching task the same information is

---

[8] Example: active-script based debugging.

also coded in the audio; thus a logical extension to this work would be to include the entropy (information flow) of audio signals as to obtain the overall entropy of the video. Another possible extension of this work is to further the analyses to include levels of task-based performance and expertise. Finally, to achieve a certain level of generalizability of our results, we plan to collect data from other tasks (e.g., program comprehension, skill acquisition in games, creating knowledge maps, visual problem solving), and compare the causal relations between different behavioural measurements across these tasks.

CRediT authorship contribution statement

Kshitij Sharma: Conceptualization, Data curation, Methodology, Writing - original draft. **Katerina Mangaroska:** Conceptualization, Data curation. **Niels van Berkel:** Methodology, Writing - original draft. **Michail Giannakos:** Conceptualization, Data curation. **Vassilis Kostakos:** Conceptualization, Methodology, Writing - original draft.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abdullah, S., Murnane, E.L., Costa, J.M., Choudhury, T., 2015. Collective smile: Measuring societal happiness from geolocated images. Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing. ACM, pp. 361–374.
Allopenna, P.D., Magnuson, J.S., Tanenhaus, M.K., 1998. Tracking the time course of spoken word recognition using eye movements: evidence for continuous mapping models. J. Mem. Lang. 38 (4), 419–439.
Ancona, N., Marinazzo, D., Stramaglia, S., 2004. Radial basis function approach to nonlinear granger causality of time series. Physical Review E 70 (5), 56221.
Backs, R.W., Walrathf, L.C., 1992. Eye movement and pupillary response indices of mental workload during visual. Appl. Ergon. 23 (4), 243–254.
Balacheff, N., Lund, K., 2013. Multidisciplinarity vs. multivocality, the case of "learning analytics". Proceedings of the Third International Conference on Learning Analytics and Knowledge. ACM, New York, NY, USA, pp. 5–13. https://doi.org/10.1145/2460296.2460299.
Blignaut, P., 2010. Visual span and other parameters for the generation of heatmaps. Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications. ACM, pp. 125–128.
Bollen, K.A., Bauldry, S., 2011. Three cs in measurement models: causal indicators, composite indicators, and covariates. Psychol. Methods 16 (3), 265.
Bollen, K.A., Pearl, J., 2013. Eight Myths about Causality and Structural Equation Models. Handbook of causal analysis for social research. Springer, pp. 301–328.
Bondareva, D., Conati, C., Feyzi-Behnagh, R., Harley, J.M., Azevedo, R., Bouchet, F., 2013. Inferring learning from gaze data during interaction with an environment to support self-regulated learning. International Conference on Artificial Intelligence in Education. Springer, pp. 229–238.
Boucsein, W., 2000. Engineering psychophysiology: Issues and applications. CRC Press.
Bradford Hill, A., 1965. The environment and disease: association or causation? Proc. R. Soc. Med. 58 (5), 295–300. https://doi.org/10.1177/003591576505800503.
Brunken, R., Plass, J.L., Leutner, D., 2003. Direct measurement of cognitive load in multimedia learning. Educ. Psychol. 38 (1), 53–61.
Brünken, R., Steinbacher, S., Plass, J.L., Leutner, D., 2002. Assessment of cognitive load in multimedia learning using dual-task methodology. Exp. Psychol. 49 (2), 109.
Buettner, R., 2013. Cognitive workload of humans using artificial intelligence systems: towards objective measurement applying eye-tracking technology. Annual Conference on Artificial Intelligence. Springer, pp. 37–48.
Chambliss, D.F., Schutt, R.K., 2018. Making sense of the social world: Methods of investigation. Sage Publications, Incorporated.
Chen, X., Breslow, L., DeBoer, J., 2018. Analyzing productive learning behaviors for students using immediate corrective feedback in a blended learning environment. Computers & Education 117, 59–74.
Chen, Y., Rangarajan, G., Feng, J., Ding, M., 2004. Analyzing multiple nonlinear time series with extended granger causality. Phys. Lett. A 324 (1), 26–35.
Churchill, D., 2014. Presentation design for "conceptual model" learning objects. British Journal of Educational Technology 45 (1), 136–148.
Di Lascio, E., Gashi, S., Santini, S., 2018. Unobtrusive assessment of students' emotional engagement during lectures using electrodermal activity sensors. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2 (3), 103.
Ding, M., Bressler, S.L., Yang, W., Liang, H., 2000. Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modeling: data preprocessing, model validation, and variability assessment. Biol. Cybern. 83

(1), 35–45.
Ding, M., Chen, Y., Bressler, S.L., 2006. 17 Granger causality: basic theory and application to neuroscience. Handbook of time series analysis: recent theoretical developments and applications 437.
Dragicevic, P., 2016. Fair Statistical Communication in Hci. Modern Statistical Methods for HCI. Springer, pp. 291–330.
Dunlop, M.D., Baillie, M., 2009. Paper rejected (p> 0.05): an introduction to the debate on appropriateness of null-hypothesis testing. International Journal of Mobile Human Computer Interaction (IJMHCI) 1 (3), 86–93.
Edwards, J.R., Bagozzi, R.P., 2000. On the nature and direction of relationships between constructs and measures. Psychol. Methods 5 (2), 155.
Ferguson, R., Clow, D., 2017. Where is the evidence? a call to action for learning analytics. Proceedings of the seventh international learning analytics & knowledge conference. pp. 56–65.
Fletcher, J., Tobias, S., 2005. The multimedia principle. The Cambridge handbook of multimedia learning 117, 133.
Foulsham, T., 2008. Saliency and eye movements in the perception of natural scenes. University of Nottingham.
Funk, M., Kosch, T., Schmidt, A., 2016. Interactive worker assistance: comparing the effects of in-situ projection, head-mounted displays, tablet, and paper instructions. Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM, pp. 934–939.
Giannakos, M., Kshitij, S., Martinez-Maldonado, R., Dillenbourg, P., Rogers, Y., 2018. Learner-computer interaction. Proceedings of the 10th Nordic Conference on Human-Computer Interaction. ACM, New York, NY, USA. https://doi.org/10.1145/3240167.3240259.
Giannakos, M., Sharma, K., Pappas, I., Kostakos, V., Velloso, Velloso, E., 2019. Multimodal data as a means to understand the learning experience. Int. J. Inf. Manage.
Giannakos, M.N., Chorianopoulos, K., Chrisochoides, N., 2015. Making sense of video analytics: lessons learned from clickstream interactions, attitudes, and learning outcome in a video-assisted course. The International Review of Research in Open and Distributed Learning 16 (1).
Giannakos, M.N., Papavlasopoulou, S., Sharma, K., 2020. Monitoring children's learning through wearable eye-tracking: the case of a making-based coding activity. IEEE Pervasive Comput. 19 (1), 10–21.
Giannakos, M.N., Sharma, K., Niforatos, E., 2019. Exploring eeg signals during the different phases of game-player interaction. 2019 11th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games). IEEE, pp. 1–8.
Gjoreski, M., Luštrek, M., Pejović, V., 2018. My watch says i'm busy: Inferring cognitive load with low-cost wearables. Proceedings of the 2018 ACM International Joint Conference on Pervasive and Ubiquitous Computing and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers. ACM, pp. 1234–1240.
Goebel, R., Roebroeck, A., Kim, D.-S., Formisano, E., 2003. Investigating directed cortical interactions in time-resolved fmri data using vector autoregressive modeling and granger causality mapping. Magn. Reson. Imaging 21 (10), 1251–1261.
Goldberg, L.R., 1999. A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. Personality psychology in Europe 7, 7–28.
Gollan, B., Haslgrübler, M., Ferscha, A., 2016. Demonstrator for extracting cognitive load from pupil dilation for attention management services. Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct. ACM, pp. 1566–1571.
Gonzalez, R.C., Woods, R.E., 2007. Image processing. Digital image processing 2, 1.
Granger, C.W., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica: Journal of the Econometric Society 424–438.
Granholm, E., Asarnow, R.F., Sarkin, A.J., Dykes, K.L., 1996. Pupillary responses index cognitive resource limitations. Psychophysiology 33 (4), 457–461.
Grant, E.R., Spivey, M.J., 2003. Eye movements and problem solving guiding attention guides thought. Psychol. Sci. 14 (5), 462–466.
Gray, R.M., 2011. Entropy and information theory. Springer Science & Business Media.
Haapalainen, E., Kim, S., Forlizzi, J.F., Dey, A.K., 2010. Psycho-physiological measures for assessing cognitive load. Proceedings of the 12th ACM international conference on Ubiquitous computing. ACM, pp. 301–310.
Hannafin, M.J., 1997. The case for grounded learning systems design: what the literature suggests about effective teaching, learning, and technology. Educational Technology Research & Development 45 (3), 101–117.
Hart, S.G., Staveland, L.E., 1988. Development of Nasa-tlx (Task Load Index): Results of Empirical and Theoretical Research. Advances in psychology. 52. Elsevier, pp. 139–183.
Hernandez, T., 2007. But what does it all mean? understanding eye-tracking results (part 3). Sep 4, 2.
Hesse, W., Möller, E., Arnold, M., Schack, B., 2003. The use of time-variant eeg granger causality for inspecting directed interdependencies of neural assemblies. J. Neurosci. Methods 124 (1), 27–44.
Hiemstra, C., Jones, J.D., 1994. Testing for linear and nonlinear granger causality in the stock price-volume relation. J. Finance 49 (5), 1639–1664.
Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., Van de Weijer, J., 2011. Eye tracking: A comprehensive guide to methods and measures. OUP Oxford.
Hyönä, J., Tommola, J., Alaja, A.-M., 1995. Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. The Quarterly Journal of Experimental Psychology Section A 48 (3), 598–612.
James, K., Humphrey, G., Vilis, T., Corrie, B., Baddour, R., Goodale, M., 2002. "Active" and "passive" learning of three-dimensional object structure within an immersive virtual reality environment. Behavior Research Methods, Instruments, & Computers 34 (3), 383–390.

Jeremy, R., Wendy, M., Juhn, A., Patricia, S., 2017. Cyberlearning community report: The state of cyberlearning and the future of learning with technology.

Joerding, W., 1986. Economic growth and defense spending: granger causality. J. Dev. Econ. 21 (1), 35–40.

Jonassen, D.H., 1991. Evaluating constructivistic learning. Educational technology 31 (9), 28–33.

Jovanović, J., Gašević, D., Pardo, A., Dawson, S., Whitelock-Wainwright, A., 2019. Introducing meaning to clicks: Towards traced-measures of self-efficacy and cognitive load. Proceedings of the 9th International Conference on Learning Analytics & Knowledge. ACM, pp. 511–520.

Just, M.A., Carpenter, P.A., 1980. A theory of reading: from eye fixations to comprehension. Psychol. Rev. 87 (4), 329.

Kaiser, S., Parks, A., Leopard, P., Albright, C., Carlson, J., Goel, M., Nassehi, D., Larson, E.C., 2016. Design and learnability of vortex whistles for managing chronic lung function via smartphones. Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM, pp. 569–580.

Kardan, S., Conati, C., 2013. Comparing and combining eye gaze and interface actions for determining user learning with an interactive simulation. International Conference on User Modeling, Adaptation, and Personalization. Springer, pp. 215–227.

Kirk, R.A., Kirk, D.A., Pesheck, P., 2016. Decision making for complex ecosystems: a technique for establishing causality in dynamic systems. International Conference on Human-Computer Interaction. Springer, pp. 110–115.

Kizilcec, R.F., Papadopoulos, K., Sritanyaratana, L., 2014. Showing face in video instruction: effects on information retention, visual attention, and affect. Proceedings of the 32nd annual ACM conference on Human factors in computing systems. ACM, pp. 2095–2102.

Klenowski, V., 2015. Questioning the Validity of the Multiple Uses of Naplan Data. National Testing in Schools. Routledge, pp. 44–56.

Klingner, J., Kumar, R., Hanrahan, P., 2008. Measuring the task-evoked pupillary response with a remote eye tracker. Proceedings of the 2008 symposium on Eye tracking research & applications. ACM, pp. 69–72.

Kosch, T., Funk, M., Schmidt, A., Chuang, L.L., 2018. Identifying cognitive assistance with mobile electroencephalography: a case study with in-situ projections for manual assembly. Proceedings of the ACM on Human-Computer Interaction 2 (EICS), 11.

Li, N., Kidziński, Ł., Jermann, P., Dillenbourg, P., 2015. Mooc Video Interaction Patterns: What Do They Tell Us? Design for teaching and learning in a networked world. Springer, pp. 197–210.

Madan, A., Cebrian, M., Lazer, D., Pentland, A., 2010. Social sensing for epidemiological behavior change. Proceedings of the 12th ACM international conference on Ubiquitous computing. ACM, pp. 291–300.

Mangaroska, K., Giannakos, M.N., 2018. Learning analytics for learning design: a systematic literature review of analytics-driven design to enhance learning. IEEE Trans. Learn. Technol. 1. https://doi.org/10.1109/TLT.2018.2868673.

Mangaroska, K., Sharma, K., Giannakos, M., Trætteberg, H., Dillenbourg, P., 2018. Gaze insights into debugging behavior using learner-centred analysis. Proceedings of the 8th International Conference on Learning Analytics and Knowledge. ACM, pp. 350–359.

Marceau, G., Cooper, G.H., Krishnamurthi, S., Reiss, S.P., 2004. A dataflow language for scriptable debugging. Proceedings. 19th International Conference on Automated Software Engineering, 2004. IEEE, pp. 218–227.

Martínez-Gómez, P., Aizawa, A., 2014. Recognition of understanding level and language skill using measurements of reading behavior. Proceedings of the 19th international conference on Intelligent User Interfaces. ACM, pp. 95–104.

Matkovič, T., Pejović, V., 2018. Wi-mind: Wireless mental effort inference. Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers. ACM, pp. 1241–1249.

Mayer, R.E., 1997. Multimedia learning: are we asking the right questions? Educ. Psychol. 32 (1), 1–19.

Mayer, R.E., 2002. Multimedia Learning. Psychology of Learning and Motivation. 41. Elsevier, pp. 85–139.

Mayer, R.E., 2005. Principles for managing essential processing in multimedia learning: segmenting, pretraining, and modality principles. The Cambridge handbook of multimedia learning 169–182.

Mayer, R.E., 2008. Applying the science of learning: evidence-based principles for the design of multimedia instruction. American psychologist 63 (8), 760.

Mayer, R.E., Hegarty, M., Mayer, S., Campbell, J., 2005. When static media promote active learning: annotated illustrations versus narrated animations in multimedia instruction. Journal of Experimental Psychology: Applied 11 (4), 256.

Mayer, R.E., Moreno, R., 1998. A cognitive theory of multimedia learning: implications for design principles. J. Educ. Psychol. 91 (2), 358–368.

Mayer, R.E., Moreno, R., 2003. Nine ways to reduce cognitive load in multimedia learning. Educ. Psychol. 38 (1), 43–52.

Mergel, B., 1998. Instructional design and learning theory.

Moreno, R., Mayer, R., 2007. Interactive multimodal learning environments. Educ. Psychol. Rev. 19 (3), 309–326.

Narayan, P.K., Smyth, R., 2005. Electricity consumption, employment and real income in australia evidence from multivariate granger causality tests. Energy Policy 33 (9), 1109–1116.

Ochoa, X., Domínguez, F., Guamán, B., Maya, R., Falcones, G., Castells, J., 2018. The rap system: automatic feedback of oral presentation skills using multimodal analysis and low-cost sensors. Proceedings of the 8th international conference on learning analytics and knowledge. ACM, pp. 360–364.

Olsen, J., Sharma, K., Aleven, V., Rummel, N., 2018. Combining gaze, dialogue, and action from a collaborative intelligent tutoring system to inform student learning processes. Proceedings of the 13th International Conference of the Learning Sciences.

ISLS, pp. 689–696.

Oppewal, H., 2010. Causal research. Wiley International Encyclopedia of Marketing.

Orquin, J.L., Ashby, N.J., Clarke, A.D., 2016. Areas of interest as a signal detection problem in behavioral eye-tracking research. J. Behav. Decis. Mak. 29 (2–3), 103–115.

Paas, F., Tuovinen, J.E., Tabbers, H., Van Gerven, P.W., 2003. Cognitive load measurement as a means to advance cognitive load theory. Educ. Psychol. 38 (1), 63–71.

Paas, F.G., Van Merriënboer, J.J., 1994. Instructional control of cognitive load in the training of complex cognitive tasks. Educ. Psychol. Rev. 6 (4), 351–371.

Paas, F.G., Van Merriënboer, J.J., Adam, J.J., 1994. Measurement of cognitive load in instructional research. Percept. Mot. Skills 79 (1), 419–430.

Paivio, A., 1990. Mental representations: A dual coding approach. Oxford University Press.

Palinko, O., Kun, A.L., Shyrokov, A., Heeman, P., 2010. Estimating cognitive load using remote eye tracking in a driving simulator. Proceedings of the 2010 symposium on eye-tracking research & applications. ACM, pp. 141–144.

Pappas, I., Sharma, K., Mikalef, P., Giannakos, M., 2018. Visual Aesthetics of E-commerce Websites: An Eye-tracking Approach.

Peitek, N., Siegmund, J., Parnin, C., Apel, S., Brechmann, A., 2018. Beyond gaze: preliminary analysis of pupil dilation and blink rates in an fmri study of program comprehension. Proceedings of the Workshop on Eye Movements in Programming. ACM, pp. 4.

Phillips, R., 1998. Models of learning appropriate to educational applications of information technology. Retrieved October 4, 2004.

Poole, A., Ball, L.J., 2006. Eye Tracking in Hci and Usability Research. Encyclopedia of human computer interaction. IGI Global, pp. 211–219.

Prabhakharan, P., Molesworth, B.R., Hatfield, J., 2012. Impairment of a speed management strategy in young drivers under high cognitive workload. Accident Analysis & Prevention 47, 24–29.

Prensky, M., 2003. Digital game-based learning. Computers in Entertainment (CIE) 1 (1), 21.

Prieto, L.P., Sharma, K., Dillenbourg, P., 2015. Studying Teacher Orchestration Load in Technology-enhanced Classrooms. Design for teaching and learning in a networked world. Springer, pp. 268–281.

Prieto, L.P., Sharma, K., Kidzinski, L., Dillenbourg, P., 2017. Orchestration load indicators and patterns: in-the-wild studies using mobile eye-tracking. IEEE Trans. Learn. Technol.

Prieto, L.P., Sharma, K., Wen, Y., Dillenbourg, P., 2015. The Burden of Facilitating Collaboration: Towards Estimation of Teacher Orchestration Load Using Eye-tracking Measures. International Society of the Learning Sciences, Inc.[ISLS].

Radach, R., Huestegge, L., Reilly, R., 2008. The role of global top-down factors in local eye-movement control in reading. Psychol. Res. 72 (6), 675–688.

Reingold, E.M., Charness, N., Pomplun, M., Stampe, D.M., 2001. Visual span in expert chess players: evidence from eye movements. Psychol. Sci. 12 (1), 48–55.

Richardson, D.C., Dale, R., Kirkham, N.Z., 2007. The art of conversation is coordination common ground and the coupling of eye movements during dialogue. Psychol. Sci. 18 (5), 407–413.

Sarsenbayeva, Z., Marini, G., van Berkel, N., Luo, C., Jiang, W., Yang, K., Wadley, G., Dingler, T., Kostakos, V., Goncalves, J., 2020. Does smartphone use drive our emotions or vice versa? a causal analysis. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. pp. 1–15.

Schultheis, H., Jameson, A., 2004. Assessing cognitive load in adaptive hypermedia systems: Physiological and behavioral methods. International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems. Springer, pp. 225–234.

Sedrakyan, G., Snoeck, M., 2016. Cognitive feedback and behavioral feedforward automation perspectives for modeling and validation in a learning context. International Conference on Model-Driven Engineering and Software Development. Springer, pp. 70–92.

Shadish, W.R., Cook, T.D., Campbell, D.T., et al., 2002. Experimental and quasi-experimental designs for generalized causal inference/william r. shedish, thomas d. cook, donald t. campbell. Boston: Houghton Mifflin,.

Sharif, B., Maletic, J., 2010. An eye tracking study on camelcase and under_score identifier styles. Program Comprehension (ICPC), 2010 IEEE 18th International Conference on. IEEE.

Sharma, K., Caballero, D., Verma, H., Jermann, P., Dillenbourg, P., 2015. Looking at versus Looking through: A Dual Eye-tracking Study in Mooc Context. International Society of the Learning Sciences, Inc.[ISLS].

Sharma, K., Caballero, D., Verma, H., Jermann, P., Dillenbourg, P., 2015. Shaping learners' attention in massive open online courses. Revue internationale des technologies en pédagogie universitaire/International Journal of Technologies in Higher Education 12 (1–2), 52–61.

Sharma, K., Chavez-Demoulin, V., Dillenbourg, P., 2017. An application of extreme value theory to learning analytics: predicting collaboration outcome from eye-tracking data. Journal of Learning Analytics 4 (3), 140–164.

Sharma, K., Chavez-Demoulin, V., Dillenbourg, P., et al., 2018. Nonstationary modelling of tail dependence of two subjects' concentration. Ann. Appl. Stat. 12 (2), 1293–1311.

Sharma, K., Jermann, P., Nüssli, M.-A., Dillenbourg, P., 2013. Understanding Collaborative Program Comprehension: Interlacing Gaze and Dialogues. Proceedings of Computer Supported Collaborative Learning (CSCL 2013). Vol. 1. pp. 430–437.

Sharma, K., Mangaroska, K., Giannakos, M., Dillenbourg, P., 2018. Interlacing Gaze and Actions to Explain the Debugging Process. pp. 640–647.

Sharma, K., Olsen, J.K., Aleven, V., Rummel, N., 2018. Exploring causality within collaborative problem solving using eye-tracking. European Conference on Technology Enhanced Learning. Springer, pp. 412–426.

Sharma, K., Olsen, J.K., Aleven, V., Rummel, N., 2020. Measuring causality between collaborative and individual gaze metrics for collaborative problem-solving with

intelligent tutoring systems. Journal of Computer Assisted Learning.

Shi, Y., Ruiz, N., Taib, R., Choi, E., Chen, F., 2007. Galvanic skin response (gsr) as an index of cognitive load. CHI'07 extended abstracts on Human factors in computing systems. ACM, pp. 2651–2656.

Siegmund, J., Kästner, C., Apel, S., Parnin, C., Bethmann, A., Leich, T., Saake, G., Brechmann, A., 2014. Understanding understanding source code with functional magnetic resonance imaging. Proceedings of the 36th International Conference on Software Engineering. ACM, pp. 378–389.

Simon, H.A., 1978. Information-processing theory of human problem solving. Handbook of learning and cognitive processes 5, 271–295.

Spirkin, A.G., 1983. Dialectical materialism. Progress Publishers.

Steyer, R., 2013. Wahrscheinlichkeit und regression. Springer-Verlag.

Steyer, R., Nachtigall, C., Wüthrich-Martone, O., Kraus, K., 2002. Causal regression models iii: covariates, conditional, and unconditional average causal effects. Methods of Psychological Research Online 7 (1), 41–68.

Sugihara, G., May, R., Ye, H., Hsieh, C.-h., Deyle, E., Fogarty, M., Munch, S., 2012. Detecting causality in complex ecosystems. Science 338 (6106), 496–500.

Sweller, J., 1994. Cognitive load theory, learning difficulty, and instructional design. Learn. Instr. 4 (4), 295–312.

Sweller, J., 2011. Cognitive Load Theory. Psychology of learning and motivation. 55. Elsevier, pp. 37–76.

Szulewski, A., Fernando, S.M., Baylis, J., Howes, D., 2014. Increasing pupil size is associated with increasing cognitive processing demands: a pilot study using a mobile eye-tracking device. Open Journal of Emergency Medicine 2 (1), 8.

Tag, B., Chernyshov, G., Kunze, K., 2017. Facial temperature sensing on smart eyewear for affective computing. Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers. ACM, New York, NY, USA, pp. 209–212. https://doi.org/10.1145/3123024.3123084.

Tam, M., 2000. Constructivism, instructional design, and technology: implications for transforming distance learning. Educational Technology & Society 3 (2), 50–60.

Thornton, D.L., Batten, D.S., 1985. Lag-length selection and tests of granger causality between money and income. Journal of Money, credit and Banking 17 (2), 164–178.

Unema, P.J., Pannasch, S., Joos, M., Velichkovsky, B.M., 2005. Time course of information processing during scene perception: the relationship between saccade amplitude and fixation duration. Vis. cogn. 12 (3), 473–494.

Van Gerven, P., Paas, F., Van Merriënboer, J., Schmidt, H., 2002. Cognitive load theory and aging: effects of worked examples on training efficiency. Learn. Instr. 12 (1), 87–105.

Van Gog, T., Kirschner, F., Kester, L., Paas, F., 2012. Timing and frequency of mental effort measurement: evidence in favour of repeated measures. Appl. Cogn. Psychol. 26 (6), 833–839.

Van Merriënboer, J.J., Kirschner, P.A., 2017. Ten steps to complex learning: A systematic approach to four-component instructional design. Routledge.

Vecchi, T., Cornoldi, C., 1999. Passive storage and active manipulation in visuo-spatial working memory: further evidence from the study of age differences. European Journal of Cognitive Psychology 11 (3), 391–406.

Verwey, W.B., Veltman, H.A., 1996. Detecting short periods of elevated workload: acomparison of nine workload assessment techniques. Journal of experimental psychology: Applied 2 (3), 270.

Weiser, M., 1991. The computer for the 21 st century. Sci. Am. 265 (3), 94–105.

Wu, P.-H., Hwang, G.-J., Su, L.-H., Huang, Y.-M., 2012. A context-aware mobile learning system for supporting cognitive apprenticeships in nursing skills training. Journal of Educational Technology & Society 15 (1), 223–236.

Wulff, A., 2007. Eyes wide shut-or using eye tracking technique to test a website. International Journal of Public Information Systems 3 (1).

Yin, B., Ruiz, N., Chen, F., Khawaja, M.A., 2007. Automatic cognitive load detection from speech features. Proceedings of the 19th Australasian conference on Computer-Human Interaction: Entertaining User Interfaces. ACM, pp. 249–255.

Zeichner, O., 2018. The impact of cognitive and non-cognitive feedback on students' achievement in a distance learning environment. i-Manager's Journal of Educational Technology 14 (4), 13.

Zhang, Y., Pfeuffer, K., Chong, M.K., Alexander, J., Bulling, A., Gellersen, H., 2017. Look together: using gaze for assisting co-located collaborative search. Pers. Ubiquitous Comput. 21 (1), 173–186.

Ziabari, S.S.M., Treur, J., 2018. An adaptive cognitive temporal-causal network model of a mindfulness therapy based on music. International Conference on Intelligent Human Computer Interaction. Springer, pp. 180–193.

Zou, C., Feng, J., 2009. Granger causality vs. dynamic bayesian network inference: a comparative study. BMC Bioinformatics 10 (1), 122.