Received: 14 August 2019 | Revised: 23 March 2020 | Accepted: 25 May 2020
DOI: 10.1111/jcal.12467

ARTICLE

Journal of Computer Assisted Learning WILEY

Measuring causality between collaborative and individual gaze metrics for collaborative problem-solving with intelligent tutoring systems

Kshitij Sharma[1] | Jennifer K. Olsen[2] | Vincent Aleven[3] | Nikol Rummel[4]

[1]Department of Computer Science, Norwegian University of Science and Technology, Trondheim, Norway

[2]Computer Human Interaction in Learning and Instruction, Department of Computer Science, Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland

[3]Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania

[4]Institute of Educational Research, Ruhr-Universität Bochum, Bochum, Germany

Correspondence
Kshitij Sharma, Department of Computer Science, Norwegian University of Science and Technology, Sem Sælandsvei 9, Gløshaugen, 7491 Trondheim, Norway.
Email: kshitij.sharma@ntnu.no

Peer Review
The peer review history for this article is available at https://publons.com/publon/10.1111/jcal.12467.

Abstract

When students are working collaboratively and communicating verbally in a technology-enhanced environment, the system cannot track what collaboration is happening outside of the technology, making it difficult to fully assess the collaboration of the students and adapt accordingly. In this article, we propose using gaze measures as a proxy for cognitive processes to achieve collaboration awareness. Specifically, we use Granger causality to analyse the causal relationships between collaborative and individual gaze measures from students working on a fractions intelligent tutoring system and the influence that the students' dialogue, prior knowledge, or success has on these relationships. We found that collaborative gaze patterns drive the individual focus in the pairs with high posttest scores and when they are engaged in problem-solving dialogues but the opposite with low performing students. Our work adds to the literature by extending the correlational relationships between individual and collaborative gaze measures to causal relationships and suggests indicators that can be used within an adaptive system.

KEYWORDS
collaboration, collaborative learning, CSCL, dual eye-tracking, Granger causality, ITS

1 | INTRODUCTION

As learning technology and data collection advances, researchers have been able to produce more accurate models of students' current states and understandings to provide better learning support. For example, within intelligent tutoring systems (ITSs), student knowledge is modelled, and based upon this assessment, cognitive support is provided in terms of individualized problem selection, hints and error feedback (VanLehn, 2006). Recently there has been a move to expand ITS support to students working collaboratively (Baghaei, Mitrovic, & Irwin, 2007; Harsley, Di Eugenio, Green, Fossati, & Acharya, 2016; Olsen, Rummel, & Aleven, 2016; Rodríguez & Boyer, 2015; Walker, Rummel, & Koedinger, 2014), so students can benefit from both the cognitive support provided by the system as well as the exchange of ideas and explanations within their group.

However, for the collaboration to be beneficial, like with any skill, students must learn how to collaborate (Rummel & Spada, 2005), making it important for learning technologies to provide social support in addition to the cognitive learning support (Weinberger, Ertl, Fischer, & Mandl, 2005) that students need to be successful. When students are collaborating, they do not all start with the same expectations about collaboration, and adaptive collaborative learning support (ACLS) may be needed to provide students with the correct support at the correct time (Magnisalis, Demetriadis, & Karakostas, 2011; Walker, Rummel, & Koedinger, 2011). ACLS can be used to adapt to the collaborative learning environment to provide appropriate support for the students by assessing student interactions, comparing them to a set of productive interactions and providing interventions that will guide students closer to a productive interaction (Soller, Martínez, Jermann, & Muehlenbrock, 2005; Walker et al., 2011).

In the classroom, it remains difficult to provide this adaptive social support in real-time due to difficulties in assessing the current state of students' collaborations. When students are in the classroom, they are often collaborating face-to-face, even when using personal devices, and communicating verbally. These verbal communications are difficult to assess in real-time and leave the system with an incomplete picture of the collaboration due to both the lack of data and possible long pauses in system interaction as the students discuss a problem. We are interested in supporting the collaborative interactions that occur as students work on a collaborative technology where all the interactions may not be captured through the system, while avoiding the different problem of fully understanding students' open-ended, ongoing verbal communications.

In this article, we propose using eye-tracking to assess students' collaboration behaviours by investigating the causal relationships between different process variables to find indicators that can be tracked and measured in real-time within a collaborative setting. Specifically, we investigate the causal relationships between students' individual and collaborative cognitive processes, using gaze patterns (i.e. focus and similarity) as a proxy, for primary school students working on a collaborative fractions ITS and examine how their dialogue plays a role in this relationship. For this analysis, we used time series data from the students working on the tutor. The points in time where the causal relationship of the gaze patterns changes may provide indications of a change in the students' collaboration and places where the system can intervene.

In this article, we are interested in finding the causal relation between individual and collaborative cognitive processes as exhibited through gaze patterns. The direction of causality, that is, whether individual gaze causes collaborative gaze patterns or the vice versa, is hypothesized to change based on multiple factors. These factors include: prior knowledge, performance and dialogues. The causal direction would inform us whether the individual efforts are guiding the collaboration or the other way around. We hypothesize that this nature of causality is not consistent throughout the collaboration; and it will change direction according to what stage of problem solving the students are in.

Specifically, we address the following research questions in this article:

1. What is the direction of causality between the collaborative and individual gaze patterns?
2. How do dialogue, prior knowledge and success relate to this causality?

From our results, we make a contribution to the work on ACLS by proposing a new metric that can be used to assess student collaboration in real-time even when students are communicating face-to-face. Additionally, we contribute to the theoretical knowledge of gaze in educational technology by extending our understanding of the relations between different gaze measures and how these relations change as the students' collaborative relationship changes. Moreover, we analyse the causal relationship between the different cognitive processes with gaze-based variables (individual and collaborative) as a proxy in a temporal manner with the dialogues as a covariate in the analysis. The results give us an opportunity to pinpoint the moments in the collaboration to provide proactive, actionable feedback.

2 | RELATED WORK

Within the field of educational technology, ITSs have a long track record of successfully supporting individual student learning (Kulik & Fletcher, 2016; Ma, Adesope, Nesbit, & Liu, 2014), particularly within the domain of mathematics (Ritter, Anderson, Koedinger, & Corbett, 2007). ITSs support student learning through both *outer loop adaptation*, where the problems students work on are adapted to their knowledge growth (VanLehn, 2006), and *inner loop adaptation* through step-by-step guidance for students within problems, both through the use of immediate feedback on steps and on demand hints. Through the inner loop adaptation, students know immediately when an error occurs and they can decide to request help from the system to figure out how to do any problem-solving step correctly. By providing students with this level of support, ITSs have been shown to be nearly as effective as human tutoring (VanLehn, 2011).

However, learning in the classroom does not only consist of individual learning moments, and interactions between students have been shown to be beneficial to the learning process (Chi & Wylie, 2014). To support these student interactions, more recently, ITSs have been extended to support students working collaboratively. These collaborative ITSs have been developed with both goals in mind of bringing more cognitive support to a collaborative task (Baghaei et al., 2007; Harsley et al., 2016) and bringing more collaboration to a typically individual task (Olsen et al., 2016; Rummel, Mullins, & Spada, 2012; Walker et al., 2014). By combining student collaboration with the cognitive support provided in the ITS, students may be able to more effectively construct knowledge to both avoid and overcome errors when they occur and effectively use the support provided through hints. Although ITSs have a strong history of modelling student learning to provide individualized cognitive support, much of the collaborative support is still fixed within these systems (Harsley et al., 2016; Rummel et al., 2012) with a limited number of ITSs exploiting the data collected to provide adaptive social support (Dowell, Cade, Tausczik, Pennebaker, & Graesser, 2014; Walker et al., 2014). In this section, we briefly review how dialogues and gaze data have been used to aid our understanding of collaborative learning processes and outcomes. Moreover, we also review in what ways certain causal relations could be established and which domains have used Granger's definition (the one used in this article) of causality.

2.1 | Role of students' dialogue in education

When students are collaborating, their dialogue is very important, as it is the main way that they share and build upon each other's ideas (Chi

& Wylie, 2014). Because of the role that the dialogue plays within the collaboration, it is often the target of assessment for understanding and providing ACLS to students. When analysing students' speech and chats, the assessment can take place with surface level features to a more in depth analysis of the dialogue content. Many previous ACLS systems have used shallow indicators from dialogue to support student collaborations such as the number of student utterances (Dowell et al., 2014; Rosatelli & Self, 2004), used sentence openers (Baker & Lund, 1997; McManus & Aiken, 2016), or tracked particular sequences of dialogue actions (e.g. use of a question mark or dialogue talk moves, Adamson & Rosé, 2012). More recently, advances in machine learning have allowed for the content of chats to be analysed to provide a model of the student collaboration (Adamson, Dyke, Jang, & Rosé, 2014; Bergner, Walker, & Ogan, 2017; Walker et al., 2014). When students are in face-to-face collaborations, they often are communicating through speech and surface speech indicators such as talk time, turn taking, and voice inflextion can be used to assess the collaboration (Martinez, Wallace, Kay, & Yacef, 2011; Viswanathan & VanLehn, 2017, 2018).

With educational technologies, in addition to the student dialogue, log data is often collected of the actions that the students have taken with the system. Together with the dialogue data, some systems have used both of these data streams to model the student collaboration (Martinez-Maldonado, Kay, & Yacef, 2013; McLaren, Scheuer, & Mikšátko, 2010; Viswanathan & VanLehn, 2017; 2018; Walker, Walker, Rummel, & Koedinger, 2010). By including features of the learning environment in the assessment of the collaboration, often the intervention can be more impactful. However, it is not always possible to assess the speech data of students, especially in a loud classroom. Instead of combining dialogue and log data information, some research has modelled the student collaboration using only the actions that the students took within the system (Diziol, Walker, Rummel, & Koedinger, 2010; Evans, Wobbrock, & Davis, 2016; Rodríguez & Boyer, 2015). Yet, these logged actions are limited and may not be as useful in understanding what is happening outside of the system if there are long pauses between system interactions when students may be having discussions.

2.2 | Eye-tracking in educational technology research

Eye-tracking can be used to make this link between the information provided in the learning technology and the group discussions students have while working with them (Cherubini, Nüssli, & Dillenbourg, 2008; Jermann & Nüssli, 2012; Sharma, Jermann, Nüssli, & Dillenbourg, 2013). Eye-tracking may be a promising method to use to assess student collaboration as research has shown that gaze is tied to communication (Meyer, Sleiderink, & Levelt, 1998). Previous research has shown a link between speech and gaze when people are working together on a task. There is a coupling of the collaborators' gaze around a reference (Richardson, Dale, & Kirkham, 2007), meaning that the collaborators' gaze may fixate at approximately the same point in time on the object referenced in the dialogue. For example, just before mentioning it and just after hearing about it people tend to look at the referenced object. The gaze has a closer coupling when each of the collaborators has the same initial information and when collaborators can visually share important objects that they are referencing in speech (Jermann & Nüssli, 2012; Richardson et al., 2007), suggesting that concrete references may have more of an impact on gaze compared to abstract references. However, these explicit references encompass a small amount of dialogue in a collaborative task. Therefore, in this contribution, we analyse the gaze data corresponding to other dialogue types as well during the collaborative sessions.

Over the past few years, eye-tracking has become a key source of process data in educational research covering a wide range of educational ecosystems. Eye-tracking has not only been used to understand the learning processes in various contexts (Prieto, Sharma, & Dillenbourg, 2015; Raca & Dillenbourg, 2013; Sharma, Caballero, Verma, Jermann, & Dillenbourg, 2015), but it also has been used to provide students with appropriate, real-time and adaptive feedback on their learning processes (D'Angelo & Begel, 2017; Sharma, Alavi, Jermann, & Dillenbourg, 2016). Specifically, with ITSs, eye-tracking has previously been used to better understand student processes during the learning intervention. The use of eye-tracking as an analysis tool in ITSs has spanned investigating both affective and cognitive states of students (Jaques, Conati, Harley, & Azevedo, 2014; Rau, Michaelis, & Fay, 2015). Within affective states, eye-tracking can be used to gauge student boredom, curiosity and attention or mind-wandering (Bixler & D'Mello, 2016; Feng, D'Mello, & Graesser, 2013) that can influence the student learning (Jaques et al., 2014). By identifying these states, interventions can be put in place. For example, D'Mello, Olney, Williams, and Hays (2012) designed a gaze-reactive ITS that detects boredom and disengagement to direct students to pay attention to the tutor using dialogue moves. In addition to tracking the affective state of the student, gaze has also been related to the cognitive state of the student. Rau et al. (2015) found that the gaze patterns of students were correlated with the types of self-explanations that students provided. However, the majority of this research has been conducted on students working individually and does not extend the analysis of eye-tracking to students working collaboratively (Belenky, Ringenberg, Olsen, Aleven, & Rummel, 2014). When students work collaboratively, they can influence each other's thought processes that can be expressed through both speech and gaze patterns.

2.3 | Eye-tracking and collaborative learning

In terms of collaborative learning scenarios, eye-tracking has most often been used with collaborating partners' dialogues. Research has shown that there is a time lag between looking at an object and referring to the same object (eye-voice span, Griffin & Bock, 2000) and a time lag between a speaker's reference and a listener's gaze on the referred object (voice-eye span, Allopenna, Magnuson, & Tanenhaus, 1998). Additionally, in terms of dual gaze, there is a lag in the eye–eye (speaker's eye listener's eye) span (i.e. the time difference between the moment a speaker looks at an object and the moment the listener looks at the same object, Richardson et al., 2007). Most of the dual eye-tracking studies have shown that the amount of time that the collaborating partners spend while looking at the same objects at the same time (cross-recurrence) is predictive of several collaborative constructs (e.g. collaboration quality (Jermann & Nüssli, 2012); misunderstandings (Cherubini, Nüssli, & Dillenbourg., 2008); learning gains (Sangin, Molinari, Nüssli, & Dillenbourg, 2011).

2.4 | Establishing causal relationships

In this article, we go beyond correlational links to explore where there may be causal links between gaze measures and how they change during different forms of dialogue. The key idea is to use the 'cause' to 'forecast' the effect to prepare adaptations in the ITS to student needs. There are three methods that could be used to show the causality between different variables: (a) Granger causality (Granger, 1969), (b) Structured Equation Modelling (SEM; Edwards & Bagozzi, 2000), and (c) conducting an intervention experiment where the hypothesized 'cause' is controlled and the hypothesized 'effect' is measured (Shadish, Cook, & Campbell, 2002). All of these methods allow for establishing the causal (hypothesized or explored) relationship between the variables of interest through different mathematical formulations. In educational research, the causality is measured mostly between the variables, such as educational levels and income (Dănăcică, Belaşcu, & Ilie, 2010; Islam, Wadud, & Islam, 2007; Kumar Narayan & Smyth, 2006), motivation and regulation (Murcia, Coll, & Garzón, 2009), goal orientation and perceived competence (Goudas, Biddle, & Fox, 1994) and a given intervention and the learning behaviour (Rau, Scheines, Aleven, & Rummel, 2013). In this contribution, we propose a shift to analyse the process of collaboration, which might help the learning design of proactive feedback tools.

2.5 | Granger causality

In this article, we used the definition of causality provided by Granger (1969), which states that a variable 'X' Granger causes another variable 'Y', if the past values of 'X' contain more information to predict the present value of 'Y' than the past values of 'Y' itself. Within the Section 3.3, we provide further detail on how Granger Causality meets the criteria for causality. Granger causality has been used in a multitude of domains to understand the relationship between observable variables, such as Neuroscience (Ding, Chen, & Bressler, 2006; Goebel, Roebroeck, Kim, & Formisano, 2003), user-consumption (Narayan & Smyth, 2005), stock-market (Hiemstra & Jones, 1994) and economics (Joerding, 1986; Thornton & Batten, 1985). We chose to use Granger causality due to the nature of the relationship between our variables. With the gaze variables, there is not a way to easily experimentally control and manipulate students' gaze in order to identify the cause-effect relationship between the two variables in question (Chambliss & Schutt, 2018). In this case, we must rely on causality methods that use existing data. In the case of SEM, the model relies on the causal assumptions of the researcher (Bollen & Pearl, 2013) based on experiment setup and prior knowledge/studies. Without theory behind the assumptions, or if there are two competing theories, the model cannot be credibly specified. However, in the case of Granger causality, such restrictions are not imposed and instead the data is used to explore the causal relationships within time series data. This method extends previous results (Olsen, Sharma, Aleven, & Rummel, 2018; Sharma et al., 2015), which show correlations between individual and collaborative gaze patterns but do not give any indication about the causal direction. This prompts the use of Granger's definition of causality to answer our research questions.

2.6 | Hill's criteria and Granger's definition

In this section, we first explain the general definition, listed as a set of criteria, for causal relations between two variables (Hill, 1965). Then we describe the extent to which Granger's definition of causality satisfies the causality criteria.

2.7 | Hill's criteria for causality

Strength: A relationship is more likely to be causal if the correlation coefficient is large and statistically significant (Hill, 1965).

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.

Coherence: A relationship is more likely to be causal if it is compatible with related facts and theories.

Experiment: A relationship is more likely to be causal if it can be verified experimentally.

Analogy: A relationship is more likely to be causal if there are proven relationships between similar causes and effects.

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 hypothesized effect based on the lags of the hypothesized cause), plausibility and coherence (the relations can be backed by the theory and behavioural explanations). Experiment and Analogy are contextual. For example, in collaborative learning, 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 definition for causality since 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.

3 | METHODS

3.1 | Experimental design and procedure

The data used for this analysis came from an experimental study (for a more detailed description, see Olsen, Belenky, Aleven, & Rummel, 2014) where the goal was to investigate the differences between individual and collaborative performance when working on conceptually or procedurally oriented tutor problems. For our analysis, we included the 14 fourth and 14 fifth grade dyads that worked collaboratively—excluding the students that worked individually—on either conceptually or procedurally oriented problem sets.

For the experiment, each teacher paired the students participating in the study based on students who would work well together and did not have drastically different mathematics abilities. The dyads were engaged in a problem-solving activity using a networked collaborative ITS, which allowed them to synchronously work in a shared problem space where they could see each other's actions while sitting at their own computers.

The morning before working with the tutor and the morning after working with the tutor, students were given 25 min to complete a pretest or posttest individually on the computer to assess their learning. The pretest and posttest used the same software as the tutoring system but did not provide the students with any feedback on their answers. During the experiment, each dyad worked with the tutor for 45 min in a pullout study design at their school. The students sat across the room from each other at a fixed SMI Red 250 Hz infrared eye-tracker facing away from one another. The students were able to communicate verbally (no video) through a Skype connection. The researchers collected dual eye-tracking data, dialogue data, and tutor log data from the sessions in addition to the pretest and posttest measures.

3.2 | Intelligent tutoring system

During the study, the dyads engaged with an ITS oriented towards supporting the acquisition of knowledge about fraction equivalence. Within each problem, the tutor provided standard ITS support, such as prompts for steps (i.e. revealing steps sequentially), on demand hints, and step-level feedback (i.e. correct or incorrect feedback) that allows the problem to adapt to the student's problem-solving strategy (VanLehn, 2006).

Within the design of the tutors, all steps, hints and feedback were shown to both students on their screens. Because new information was revealed throughout the problem instead of all at the beginning, these reveals could help to guide the students' gaze. When a new step was shown in the problem, the previous information and steps remained on the screen increasing the amount of information with which the student could engage. Although there was only one step for the students to work on at a time, this step could consist of multiple boxes to fill in or multiple buttons with which the students could interact.

To support students collaborating, the ITS support mentioned above was combined with embedded collaboration scripts that guided the students through the actions they could take within the tutor. The embedded collaboration scripts included three theoretically proven types of collaboration support: roles, cognitive group awareness, and individual accountability (see Figure 1a). First, for many steps, the students were assigned roles (King, 1999). On steps with roles, one student was responsible for entering the answer and the other was responsible for asking questions of their partner and providing help with the answer. The tutor indicated the current role for the students through the use of icons on the screen. Both students had a different icon in the same place for each step forming a symmetry in what they saw on their screen. A second way the collaboration was supported was by providing students with information their partner did not have that they were responsible for sharing for the problem to be completed causing individual accountability (Slavin, 1996). Like with the icons, when one student had text on their screen, the other student had text in the same place informing them that their partner had something to share. The final feature was cognitive group awareness, where knowledge that each student has in the group is made known to the group (Janssen & Bodemer, 2013). On steps where this feature was implemented, each student was given an opportunity to answer a question individually before the students were shown each other's answers and asked to provide a consensus answer.

3.3 | Data analysis

We used a Granger causality (Granger, 1969) test to examine the causality between our gaze measures. The basic definition of Granger causality has two assumptions (Granger, 1969). The first assumption is that cause occurs before effect and that the cause has information about the effect that is more important than the history of the effect. In terms of the nature of the concerned time series, Granger causality is defined for linear and stationary time-series contexts, but variations for non-linear (Ancona, Marinazzo, & Stramaglia, 2004; Chen, Rangarajan, Feng, & Ding, 2004; Freiwald et al., 1999) and non-stationary (Ding, Bressler, Yang, & Liang, 2000; Hesse, Möller, Arnold, & Schack, 2003) contexts exist.

We start by creating two models: one that predicts the effect using the lags of the effect and one that predicts the effect using the lags of both the cause and the effect. The basic principle of Granger causality is to compare these two models to test if x *causes* y. The first model predicts the value of y at time t using the previous p-values of y. The second model predicts the value of y at time t using the

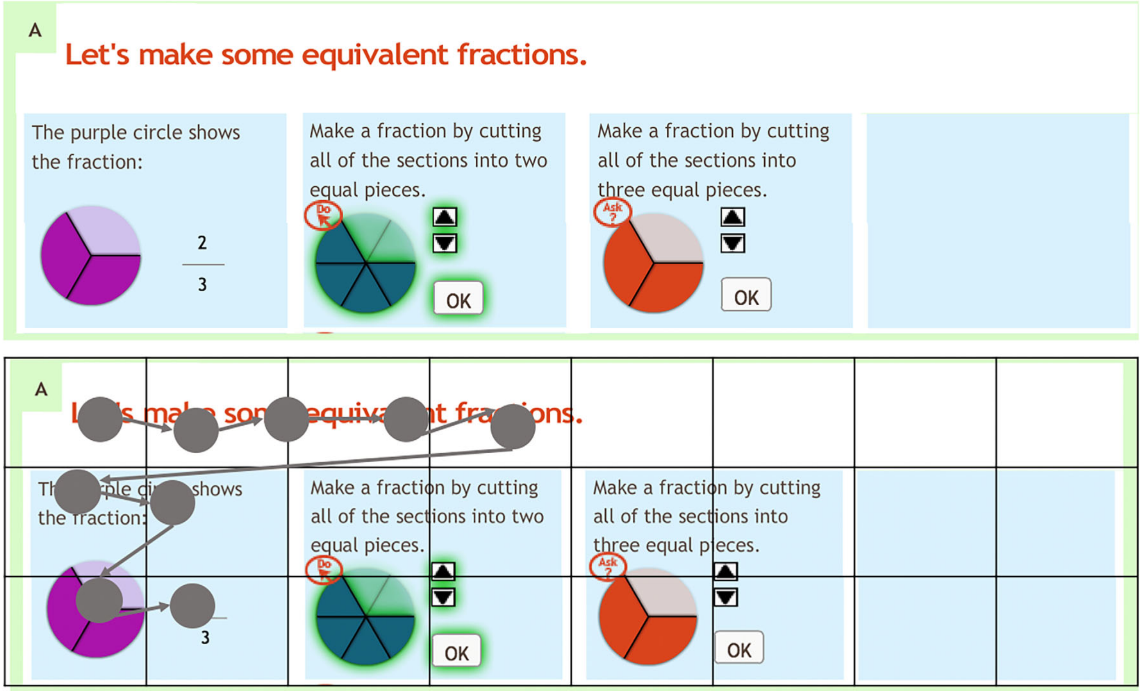

FIGURE 1 (a) Example of a fractions interface showing incremental step reveals, feedback and hint requests. Students had roles assigned that were displayed through their icon. (b) A typical example of stimulus overlaid with a 50-by-50 pixel grid and the gaze data [Colour figure can be viewed at wileyonlinelibrary.com]

previous *p*-values of both *x* and *y*. Mathematically, following is a bivariate linear autoregressive model for two variables *x* and *y*:

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

$$y(t) = \sum^{p} 21j() + \sum_{j=1}^{p} \alpha_{22j} y(t-j) + \varepsilon_2(t) \quad (2)$$

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.

We can conclude that *x Granger-causes y* if the coefficients in α_{12} are jointly significantly different from zero. Statistically, this can be tested using an *F*-test with the null hypothesis α_{11} = 0. Also, the value of *p* can be decided based on the AIC (Akaike, 1974) or BIC (Schwarz, 1978) model estimation values. The *F*-values are calculated using a Wald-test (Kodde & Palm, 1986), which is also called 'Wald's Chi-square test'. This is a test to examine if the explanatory variables of a model (in our case, these are the various lags included in the model) are significant or not. This test can work both for continuous and categorical variables. Therefore, using this test, one can establish Granger causality with both types of the variables. A significant variable in the model would have a non-zero parameter in the Wald-test. The *p*-values in the Wald-test are calculated using a variant of likelihood optimisation such as AIC or BIC. This is a non-parametric test, so it can be used without the knowledge of the underlying distributions.

4 | DATA SET

During the experiment, the researchers collected gaze measures, student dialogue, and test scores. For our analysis, we investigated how analysing these multiple data streams in relation to each other provided additional insight into the collaborative learning process. The main aim of this analysis is to contribute to existing literature that has mainly analysed the processes from a correlational point of view.

4.1 | Gaze data

For our gaze measures, we analysed the students' individual focus and their combined similarity for each dyad. Focus (how concentrated/scattered the individual gaze is) and similarity (what proportion of time the peers spent looking at similar set of objects) have been used in recent research concerning collaborative eye-tracking (Schneider et al., 2016; Sharma et al., 2013, 2015) to combine and analyse gaze behaviour at individual and collaborative levels. In our case, we chose these variables because the use of both individual and collaborative measures could provide insights into a dyad's collaborative process depending upon if the individual measures were causing the collaborative or vice versa.

For all of our gaze measures we used a grid to divide the screen into smaller segments opposed to using areas of interest (AOIs) so that we could capture any interactions with the screen rather than only the main components of the screen. We divided the screen in a 50-by-50 pixel grid (Figure 1b). We chose this size grid because it was roughly the size of the individual components within the tutoring system. Additionally, we divided the whole problem-solving session into 10-second time windows. The 10-second time windows allowed enough time for students to look at the same place if there was lag between references but not so high as to obscure changing gaze patterns.

4.1.1 | Individual focus

This measures how many objects were looked at in a given time window. In other words, this is a measure of *how concentrated or scattered the gaze of an individual is*. This measure is computed in terms of the entropy of the gaze. To compute the entropy, we needed to divide both the area of the screen and the time the students were working into discrete measurements. We then computed the proportion of the time spent in each block in the spatial grid for each 10-s time window. This resulted in a series of two-dimensional proportionality vectors. Finally, we computed the Shannon Entropy for each of the vectors. A low entropy value (the minimum possible value is zero) depicts that the student was looking at only a few elements on the screen, which we call focused gaze (see Figure 2). On the other hand, a high value of entropy indicates more elements are being looked at in a given time window, which we call unfocused gaze. The individual focus indicates whether the information processing is global (many items looked at in the given time period) or local (only a few items looked at in the given time period). This is analogous to the definition provided by Poole and Ball (2006) to the ratio of global and local information processing (defined using the fixation durations and saccade lengths). To compute a single measurement of the individual focus, we computed the probability of both the participants in a pair having a low focus size.

Although focus and attention are related concepts, focus, as we define it here, does not contain the idea of processing the stimulus, as is required in the definition of attention (Hoffman, 1998). Attentive gaze indicates a certain level of processing of the sensory input. Focused gaze simply indicates a small number of elements looked over a fixed time period. For example, when a student is daydreaming, their gaze is often very focused but they are not being attentive to that stimulus.

4.1.2 | Collaborative gaze

This is a measure of *how similar the two individual gaze patterns are*. In order to compute the similarity between the gaze patterns of the collaborating students, we divided the screen space and the interaction time in the same manner as we did for the focus computation. We computed the similarity between the two proportionality vectors by using the reverse function (1/[1 + x]) of the correlation matrix of the two vectors. A similarity value of one will show no similarity between the two gaze patterns during a given time window (see Figure 3). On the other hand, a lower value of similarity will show that the two participants spent time looking at a similar set of objects on the screen during the same time window. Gaze similarity is an alternative measure of gaze convergence (Jermann & Nüssli, 2012; Richardson & Dale, 2005), the only difference between gaze similarity and gaze convergence comes from the mathematical formulation. Gaze convergence uses cross-recurrence (a concept borrowed from dynamical systems) measures to measure how much time the peers spent looking at the same objects (usually coded through Areas of Interest) at the same time (with some lag). This requires an additional step of defining the areas of interest (which might be too context dependent) as an additional preprocessing step. Gaze similarity, as gaze cross-recurrence, is a measurement of joint-attention which might be the result of pointing gesture or verbal references during the collaborative problem-solving (Jermann & Nüssli, 2012; Richardson & Dale, 2005). Therefore, increased gaze similarity might indicate an improved common ground (Sharma et al., 2013, 2015).

4.2 | Dialogue data

For our analysis, we used student dialogues as an indicator of the processes that the students were engaged in during problem-solving. To understand the cognitive processes that students engaged in while

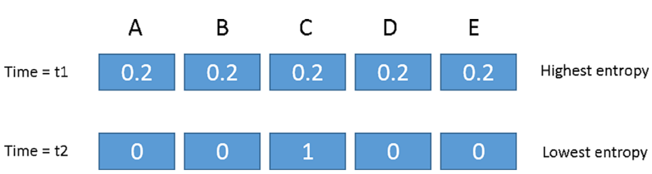

FIGURE 2 Individual focus [Colour figure can be viewed at wileyonlinelibrary.com]

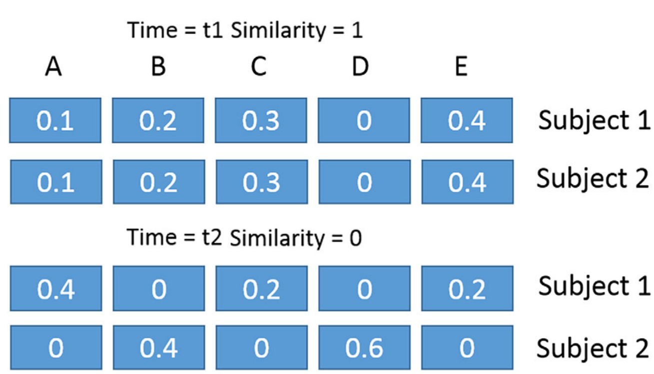

FIGURE 3 Collaborative similarity [Colour figure can be viewed at wileyonlinelibrary.com]

learning, a widely used method is verbal reporting (Van Gog, Kester, Nievelstein, Giesbers, & Paas, 2009). When students engage in concurrent reporting, such as when students vocalize their thoughts while performing the task, they provide information on their actions and outcomes that can be used to assess the cognitive processes that the students are using (Taylor & Dionne, 2000). However, students may decrease their vocalization as the task gets longer or they become absorbed, and, additionally, the vocalization may influence their cognitive processing (Ericsson & Simon, 1984; Schooler, Ohlsson, & Brooks, 1993). When students are working in a collaborative environment, the students naturally vocalize as they work with their partner with less risk of a reduction in the vocalization. Additionally, because the dialogue is part of the collaborative task, we are not asking the students to engage in outside tasks that could change their process. Therefore, the collaborative dialogue allows us to assess student processes through vocalizations without the limitations that arise when students are working individually.

Each of the student dialogues were transcribed and coded for the type of problem-solving process the students were engaged in. For our coding, we did not intend to infer all mental processes, but to fully depend on what occurs in the dialogue. Within our transcripts, we coded for the problem-solving process at the utterance level (Table 1 shows the definitions and examples for the different codes). This allowed us to have a fine-grained coding for each second of the dialogue without losing the context of the words. Our codes consisted of five different activities: acknowledgement, read out loud, interface, problem-solving, and metacognitive. For the coding, all statements that were off-task or were discussions with a researcher were marked as 'not applicable' and were discarded from the analysis. An inter-rater reliability analysis was performed to determine consistency among raters (Kappa = 0.78).

TABLE 1 Dialogue coding scheme with examples

Code	Definition	Example
NA	The student engages in off-task behaviour, converses with the experimenter, or vocalizations without any context	'Just refresh it?' (speaking with the experimenter)
ACK *N* = 321	The student acknowledges their partner, or they request acknowledgement or a repeat of what the partner has said	'Oh'
ROL *N* = 843	The student is reading information provided within the problem and presented on the screen	'OK, is 3/4 in its most reduced form?' (reading from screen)
INTF *N* = 433	The student discusses actions that can be taken in the interface or engage in work coordination	'Do we click the up or down arrow?'
PRO *N* = 2,473	The student is providing an answer to the problem or showing evidence of think aloud as they solve the problem	'OK, uh, by 3, okay, it's 9…'
META *N* = 167	The student verbally expresses their understanding of their current knowledge/ problem-solving state	'I'm so, wait, wait, I'm so confused now'

Note: Codes were applied at the utterance level.
Abbreviations: ACK, Acknowledgement; INTF, Interface; META, Metacognitive; NA, Not applicable; PRO, Problem-solving; ROL, Read out loud.

4.3 | Test data

Finally, to measure students' learning of equivalent fractions, the students took pretests and posttests. The pretest and posttest scores allowed us to understand the relation of the causality to student knowledge. The tests were computer-based and developed to closely align with the target knowledge covered in the tutors. The tests consisted of five procedural and six conceptual test items. Two isomorphic sets of questions were developed, and there were no differences in performance on the test forms across all participants in the original study, ($t[79] = 0.96$, $p = .34$). The presentation of these forms as pretests and posttests was counterbalanced.

TABLE 2 Descriptive statistics for the variables used in the article

	Mean	*SD*	Min	Max
Similarity	0.04	0.06	0	1 (baseline < 0.0001)
Focus	0.07	0.06	0	1 (baseline = 0.003)
Individual posttest	2.84	2.20	0	10 (theoretical)
Average posttest	2.84	1.80	0	10 (theoretical)
Individual pretest	2.87	2.23	0	10 (theoretical)
Average pretest	2.79	1.80	0	10 (theoretical)

5 | RESULTS

In this section, we provide the analyses to arrive at a causal relationship between the gaze focus and similarity mentioned in section Variables. To answer the first research question (What is the direction of causality between the collaborative and individual gaze patterns?), we provide an example for how to determine the Granger causality between two variables using the method explained in section Analyses. This example uses the results from the overall causality analyses between the focus and similarity (descriptive values are shown in Table 2). Next, to answer the second research question (How do dialogue, prior knowledge and success relate to this causality?), we present the causality analyses with prior knowledge, success and dialogues (distribution of the different dialogue codes is shown in Table 3) as covariates. Finally, to analyse the effect of prior

TABLE 3 Distribution of the different dialogue codes across two different groups of collaborative successes

	Dialogue proportions				
	ACK	PRO	ROL	META	INTF
Mean low posttest	7.64	54.69	19.57	5.76	12.33
SD low postest	4.55	39.46	12.93	7.67	7.95
Mean low pretest	8.26	58.03	22.19	2.17	9.23
SD low pretest	7.92	31.32	17.12	3.81	10.15

Abbreviations: ACK, Acknowledgement; INTF, Interface; META, Metacognitive; PRO, Problem-solving; ROL, Read out loud.

TABLE 4 The Granger causality model, across different data types, for collaborative similarity and probability that both participants have high focus

Model	Order	*F*-value	*P*-value
Overall data (1)			
Similarity → focus	4	2.51	.03*
Focus → similarity	4	2.04	.09
Participants engaged in dialogues (2)			
Similarity → focus	8	2.12	.03*
Focus → similarity	8	0.93	.47
Participants engaged in dialogues with INTF dialogue (3)			
Similarity → focus	6	2.83	.009*
Focus → similarity	6	1.01	.41
Participants engaged in dialogues with PRO dialogue (4)			
Similarity → focus	5	0.21	.95
Focus → similarity	5	2.52	.02*
Dyads with high average posttest score (5)			
Similarity → focus	2	3.91	.02*
Focus → similarity	2	1.70	.18
Dyads with low average posttest score (6)			
Similarity → focus	3	7.04	.00001*
Focus → similarity	3	2.04	.11
Dyads with high average posttest score with PRO dialogues (7)			
Similarity → focus	2	2.81	.05*
Focus → similarity	2	1.01	.31
Dyads with low average posttest score with PRO dialogues (8)			
Similarity → focus	3	0.54	.44
Focus → similarity	3	2.74	.05*
Dyads with high average pretest score (9)			
Similarity → focus	3	6.49	.0002*
Focus → similarity	3	0.04	.98
Dyads with low average pretest score (10)			
Similarity → focus	3	0.11	.95
Focus → similarity	3	4.42	.004*

Note: The direction of causality is denoted with an asterisks (*).
Abbreviations: INTF, Interface; PRO, Problem-solving.

knowledge, performance and dialogue on the nature of causality in time, we present the temporal causal analysis for the pairs with high/low average posttest scores.

5.1 | Example

In this analysis, we are checking the Granger causality in both directions without a strong hypothesis for either direction to be more likely.

1. Whether the concentrated/scattered gaze of an individual predicts the chances of looking at the same thing. Or,
2. Looking at similar things makes the degree of 'being concentrated/scattered' similar.

Let us take the case of 'focus' (the probability that both the participants have low gaze entropy, that is, low focus values) and 'similarity' (the extent to which the peers looked at a similar set of objects in a given time window). Table 4, comparison 1 shows the Granger causality results for the overall data. The order of the model (Table 4, column 2) denotes how much lag was used to compute the causal relationship (p in Equations (1) and (2)). In other words, how much prior information from the 'cause' we need to reliably predict the 'effect'.

In the case of Table 4, comparison 1, the lags used are four time windows (each time window corresponds to 10 s). To check if similarity Granger causes focus, we create two models given by Equations (1) and (2) and compare them using an F-test. The F- and p-values denote the test statistic and significance of the model (Table 4, columns 3 and 4, respectively). We repeat the same process for checking if focus Granger causes similarity. We can see in Table 4, comparison 1 that 'similarity Granger causes focus' has a higher F(2.51) and lower (and significant) p-value (.03) than 'similarity Granger causes focus' (F = 2.04, p = .09). This shows that the history of similarity values predicts the current value of the focus in a more reliable manner than the history values of focus predict the current value of similarity. Thus, we can conclude that 'similarity Granger causes focus'.

5.2 | The direction of causality between the collaborative and individual gaze patterns

Continuing from the example above, we can see in Table 4, comparison 1 that 'similarity Granger causes focus' has a higher F(2.51) and lower (and significant) p-value (.03) than 'similarity Granger causes focus' (F = 2.04, p = .09). Thus, we can conclude that 'similarity Granger causes focus'.

5.3 | Causal relation between individual focus and collaborative similarity considering covariates (dialogue, prior knowledge and success)

The previous set of results is based on the overall time series values of individual focus and collaborative similarity. The causal relation might be misleading without considering covariates in the analysis since the causal direction may be affected by the prior knowledge, dialogues and/or the success of the pair. Therefore, we present the analysis using these factors as covariates in our analysis.

We observe that *similarity Granger causes focus* during the whole interaction (Table 4, Comparison 1). This causality also remains when the dyads are engaged in dialogue (Table 4, Comparison 2). Considering the data from the individual dialogue categories, the same causality holds when the peers are talking about interface issues (INTF, Table 4, Comparison 3). However, the causality changes polarity (that is *focus Granger causes similarity*) while the peers are talking about problem-solving (Table 4, Comparison 4). Additionally, there is no conclusive causality for ACK and META.

However, when we divide the data into pairs with high and low average posttest scores, we observe a few different relations. For the pairs with high posttest averages, *similarity Granger causes focus* (Table 4, Comparison 5). This polarity does not change for 'PRO' abstraction (Table 4, Comparison 7). For the pairs with low posttest averages, *focus Granger causes similarity* (Table 4, Comparison 6) and the polarity changes for 'PRO' abstraction (Table 4, Comparison 8).This result shows that there is an interaction between the focus, similarity and performance in addition to an interaction between the focus, similarity and dialogue.

Finally, we considered the relation between the pretest and posttest scores. There is a positive significant correlation between the average pretest and the posttest scores for the pairs ($r[27] = 0.57$, $p = .001$), indicating that prior knowledge also contributes to the success. Therefore, we divided the data set into dyads with low and high average pretest scores and found that *similarity Granger causes focus* for the pairs with high average pretest scores (Table 4, Comparison 9); whereas, *focus Granger causes similarity* for the pairs with low average pretest scores (Table 4, Comparison 10).

5.4 | Temporal causalities

We chose to divide students by their posttest scores since it is an indicator of the students' target knowledge at the end of the intervention. Figure 4 shows the temporal direction of the Granger causality between focus and similarity for pairs with high average posttest scores. We can see that for students with high posttest scores, for a major part of the collaboration (63.33%), the similarity is causing focus, followed by periods of time (episodes) with no clear causality between focus and similarity (33.33%). We observe that there are a few episodes where the causality changes the direction, and focus seems to cause the similarity. However, the percentage of such episodes is low (4.33%).

Figure 5 shows the temporal direction of the Granger causality between focus and similarity for pairs with low average posttest scores. We can see that for a major part of the collaboration (46.67%) the focus is causing similarity, followed by episodes where the causality changes direction and similarity causes the focus (33.33%). Finally, the episodes with no clear causality between focus and similarity (23.33%) are the lowest for the pairs with low average posttest scores.

Figure 6 shows the temporal Granger causality between focus and similarity for the episodes where the peers were engaged in dialogue. The top panel of the Figure 6 shows similar information as Figure 4, while the bottom panel (inverted bar chart) shows the percentage of the most frequent dialogue code during the given time window of 200 s. We observe that when we consider the episodes with the dialogue, pairs with high average posttest show similar behaviour as in Figure 4. When the pairs with high average posttest score are engaged in 'PRO' dialogues, similarity causes focus for most of the episodes (71.42%), while for 21.42% episodes do not exhibit any causal relations. We also observe that the first dialogue episode is 'ROL' (read out loud) where the peers' focus is causing their similarity.

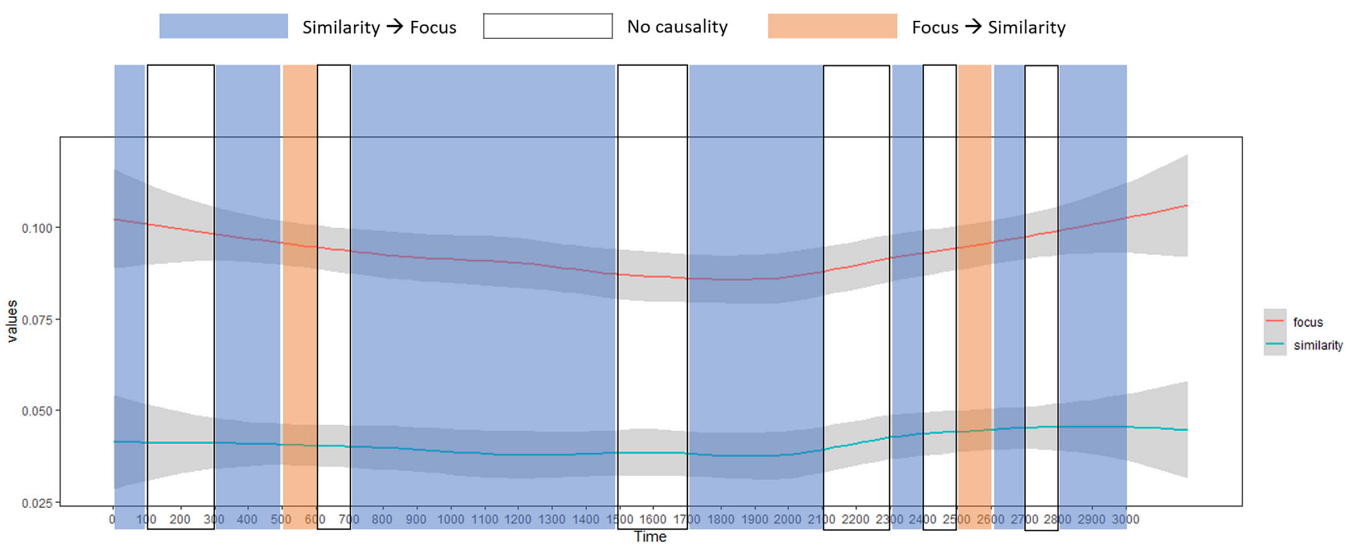

FIGURE 4 High posttest temporal causality [Colour figure can be viewed at wileyonlinelibrary.com]

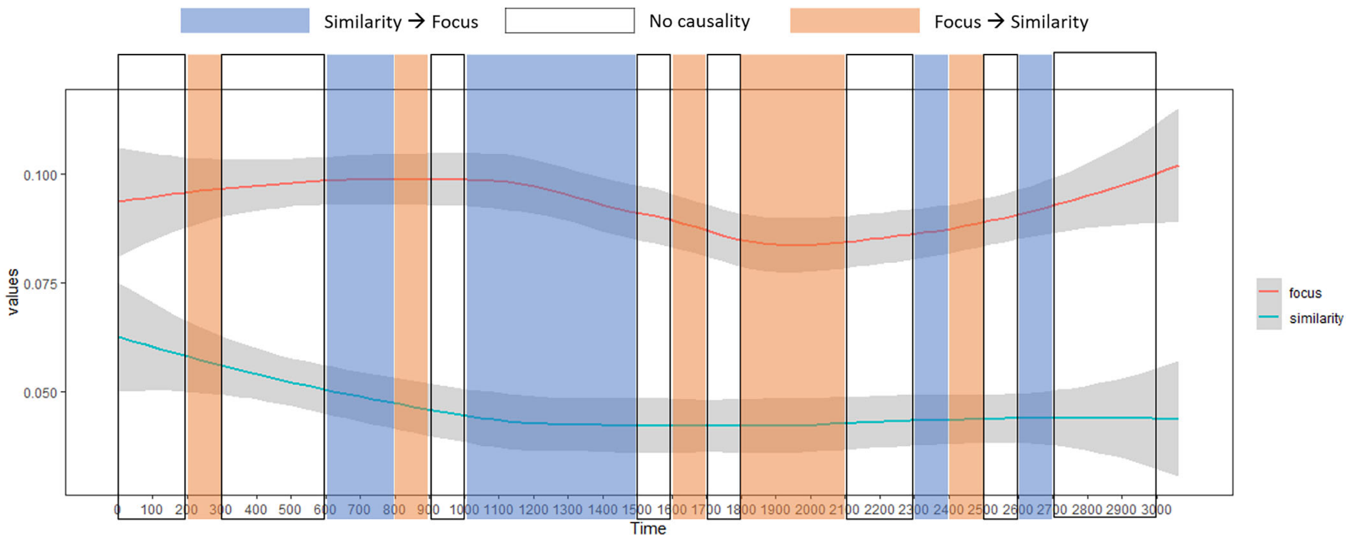

FIGURE 5 Low posttest temporal causality [Colour figure can be viewed at wileyonlinelibrary.com]

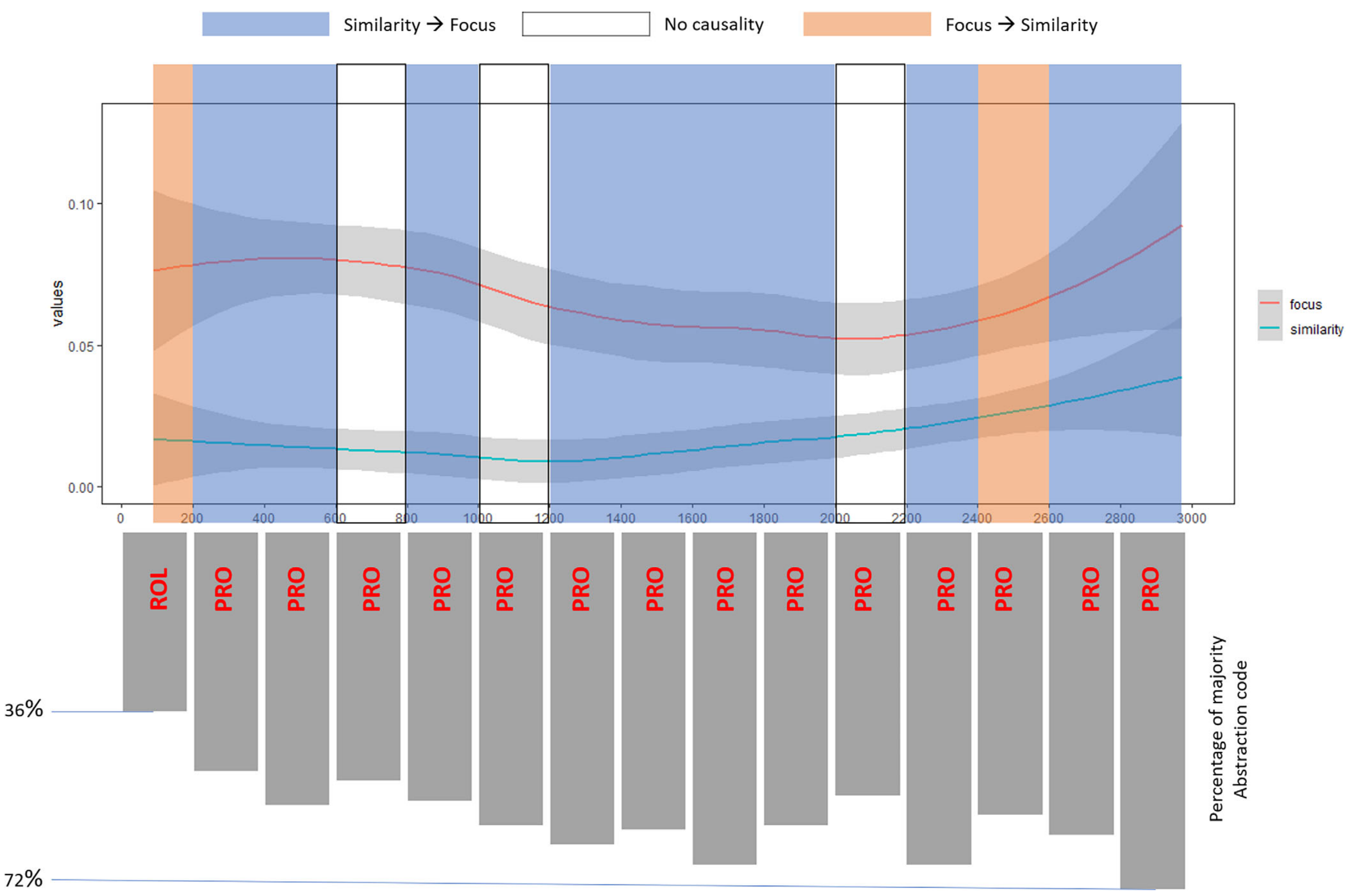

FIGURE 6 High posttest temporal causality with most frequently occurring abstraction codes [Colour figure can be viewed at wileyonlinelibrary.com]

We observe a dissimilar behaviour for the pairs with a low average posttest score. Figure 7 shows the temporal Granger causality between focus and similarity for the episodes where the peers were engaged in dialogue. The top panel of Figure 7 shows similar information as Figure 5, while the bottom panel (inverted bar chart) shows the percentage of the most frequent dialogue code during the given time window of 200 s. We observe that when we consider the episodes with the dialogue considered, pairs with low average posttest show similar behaviour as in Figure 5. When the pairs with low average posttest score engaged in 'PRO' dialogues, either focus causes

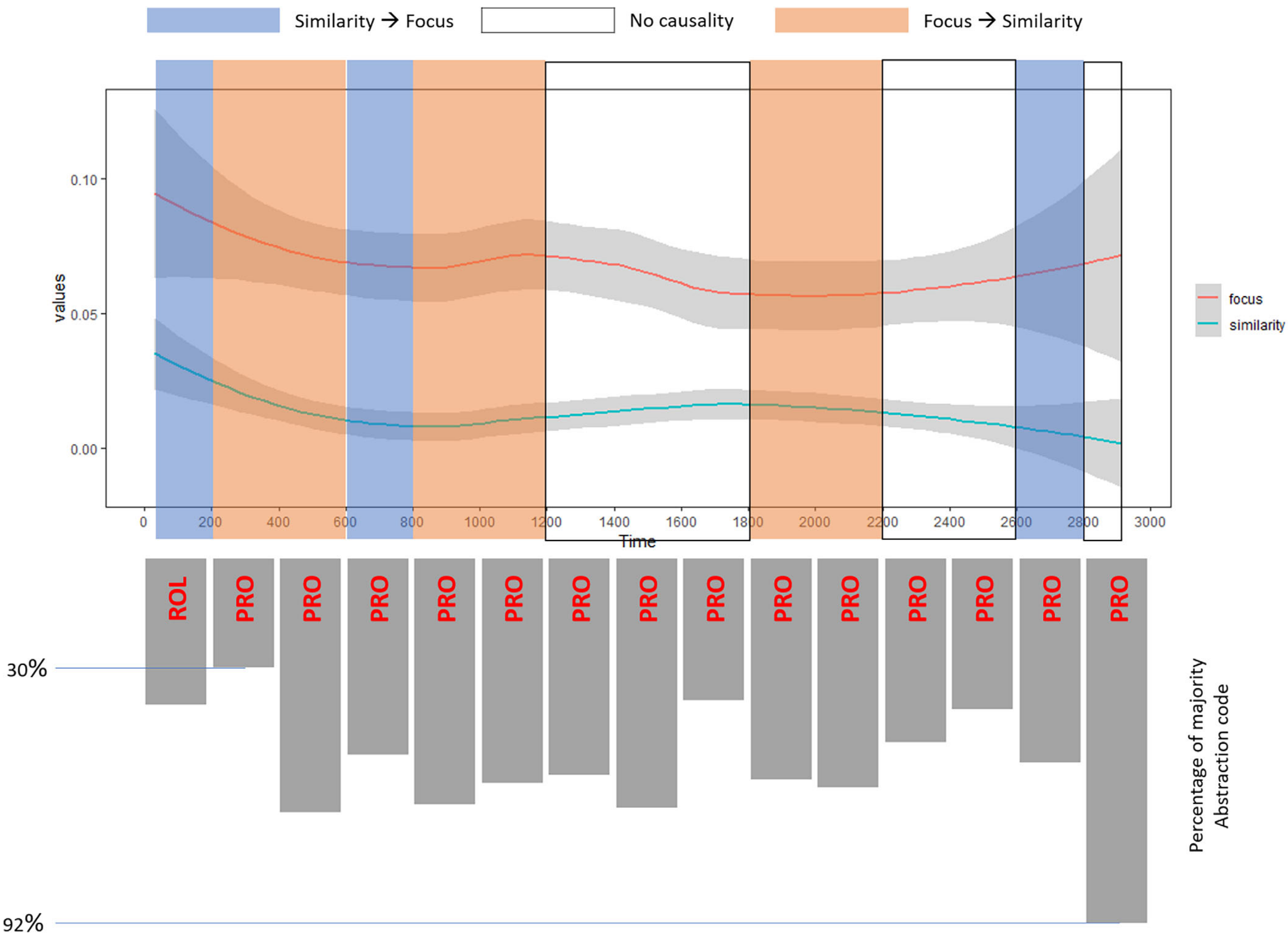

FIGURE 7 Low posttest temporal causality with most frequently occurring abstraction codes [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 5 Comparing the proportions of different dialogue codes in the different causal episodes for the two pair success levels

	ACK		PRO		INTF		META		ROL	
Dialogue code	*F*	*P*	*F*	*P*	*F*	*P*	*F*	*P*	*F*	*p*
Pair success	0.51	.47	1.43	.23	0.01	.99	4.81	.03	0.21	.64
Causality direction	1.25	.26	23.64	.0001	7.56	.008	1.78	.18	4.53	.03
Interaction (success direction)	8.57	.005	11.75	.001	3.89	.05	0.33	.56	9.01	.004

Abbreviations: ACK, Acknowledgement; INTF, Interface; META, Meta-cognitive; PRO, Problem-solving; ROL, Read out loud.

similarity (42.85%) or there is no causality exhibited (42.85%) for most of the episodes (42.85%) while only for 14.28% episodes similarity causes focus. We found that the first dialogue episode is 'ROL' during which the peers' similarity is causing their individual focus.

5.5 | Pair success, causality direction and dialogues

Finally, we compare the proportion of the different dialogue codes for the two causality directions (focus causing similarity and similarity causing focus) and the success levels (high and low). We observe that there is a significant interaction effect (Table 5) of the causality direction and a pair's success on the proportions of PRO (F[1,26] = 11.75, p = .001) INTF (F[1,26] = 3.89, p = .05) and ROL (F[1,26] = 9.01, p = .004). The successful pairs have significantly more PRO dialogues in the episodes where the similarity is causing the focus than the unsuccessful pairs (Figure 8). Successful pairs also have significantly less ROL and INTF dialogues in the episodes where focus is causing similarity than the unsuccessful pairs (Figure 8).

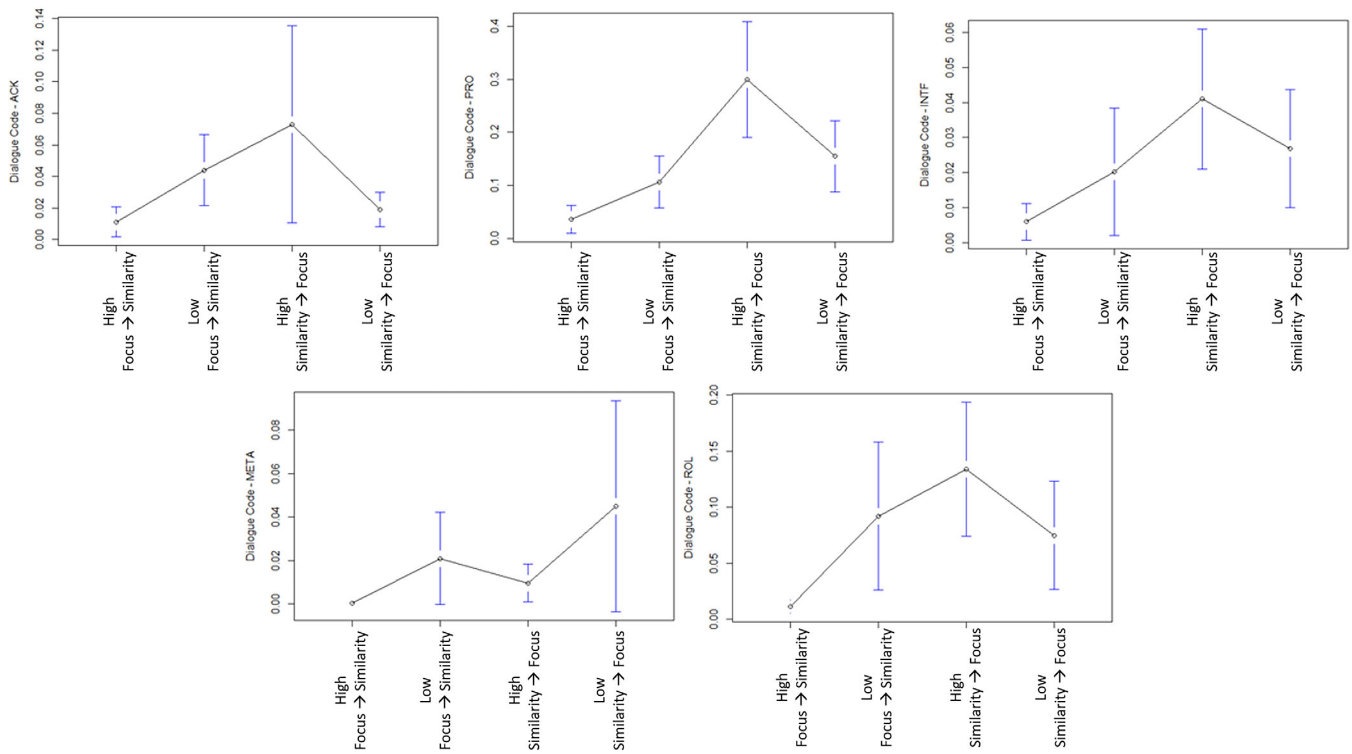

FIGURE 8 Proportions of different dialogue codes in the different causal episodes for the two pair success levels [Colour figure can be viewed at wileyonlinelibrary.com]

6 | DISCUSSION

In this article, we propose using eye-tracking to assess student collaboration behaviours by investigating the causal relationships between different process variables as proxies for cognitive processes to find indicators that can be tracked and measured in real-time within a collaborative setting. Specifically, we investigate the causal relationships between students' individual and collaborative cognitive processes, using gaze patterns (i.e. focus and similarity) as a proxy, for elementary school students working on a collaborative fractions ITS and examine how their dialogue plays a role in this relationship over time. From our results, we contribute to the work on ACLS by proposing a new metric that can be used to identify moments for intervention during real-time student collaboration, even when students are communicating face-to-face.

We found that, depending on the context, individual focus sometimes caused collaborative similarity while at other times, collaborative similarity caused individual focus. What this means in terms of our data is that when the individual focus is causing the collaborative similarity, the higher focus in students leads to students looking at a similar place on the screen as their partner. This could be analogous to the emergent coordination as defined by van Ulzen, Lamoth, Daffertshofer, Semin, and Beek (2008) and Chartrand and Bargh (1999) as the coordination between two individuals having no plan to work together. Emergent coordination has shown to not influence goal achievement (Semin & Smith, 2008). On the other hand, when the collaborative similarity is causing the individual focus, then when the students in a dyad are looking at similar parts of the problem, they are more likely to then focus on a part of the problem. This is analogous to the planned coordination, where the peers use the context (dialogues) and knowledge to achieve a common goal with others (Vesper, Soutschek, & Schubö, 2009). The key difference between the two causalities, 'looking at the same place hence focused' and 'focused hence looking at the same place', might explain whether collaborative processes are driving the individual processes or the other way around.

Although, overall, we found that similarity caused focus, when we included covariates into our models, the nature of the causal relationship changed. These changes in relationship indicate that not all gaze patterns are the same and when we only analyse the correlations, we lose some information. By analysing the causal relationship between the focus and similarity, we may be able to provide some insight into the cognitive process of the students as they are working collaboratively based upon how the nature of the causal relationship changes. Richardson, Dale and Kirikham (2007) have also shown that the social context (in our case the dialogues and performance) can influence the gaze and memory processes. Richardson et al. (2012) argue that the different relations (causality direction in our case) between the individual and joint/collaborative gaze can stimulate behavioural and cognitive performance.

Specifically, when analysing the causal relationship in relation to the students' dialogue, we found that collaborative similarity caused the individual focus for overall dialogue and interface dialogue, but the nature of the causality switched when the pairs were discussing 'how to solve the problem' (i.e. problem-solving dialogue). One plausible explanation for the change in causality is that when two peers are solving a problem, they start to focus on the step they should solve, which leads to them looking at the same part of the screen and having a higher similarity. Additionally, we did not find any conclusive causality during the episodes when the peers were using acknowledgements or metacognitive dialogue. These findings may be explained by the fact that there is no need for the stimulus support (no specific requirement to look at a particular part of the screen) when acknowledging a partner's dialogue or a requirement to reflect upon a peer's own state of understanding.

When we take into consideration how students performed on their tests, we find contrasting causality relations. In the case of students that had high scores on their posttests and are discussing the problem-solving, we found that the collaborative similarity seems to drive the individual focus, while in the case of students that scored low on their posttests and are discussing the problem-solving, the relationship seems to be reversed. We found similar results for students with high and low prior knowledge. When students are engaged in an effective collaboration, they are more likely to ground their discussion (Laughlin, Hatch, Silver, & Boh, 2006), which can lead to a greater focus on an area of the problem (Richardson et al., 2007). This collaboration can result in a successful individual performance (Nokes, 2009). For students with a high posttest score (high performance), we saw that the collaboration was driving the focus during problem-solving, which could be explained by the students grounding their discussion more, leading to focus and higher performance. However, when students are not in an effective collaboration, they may only share information when they are focused. Previous studies have reported worse group performance than individual performance (Weldon & Bellinger, 1997) when there is a lack of coordination (Steiner, 1972) or a disruption of individual contribution (Diehl & Stroebe, 1987). This may be the case for the low performing students who may have struggled to collaborate well when they were not already focused, as seen through their gaze.

Additionally, the different causal relations for students with high/low prior knowledge or success show that collaborative gaze causing the individual gaze is indicative of a 'top-down' approach while individual gaze causing the collaborative gaze points to a 'bottom-up' approach. Having coordinated gaze is a result of deeper sociocognitive mechanisms (Jermann & Nüssli, 2012; Richardson et al., 2007; Sangin et al., 2011; Schneider et al., 2016; Sharma et al., 2015) than just looking at a few elements on the screen (high focused gaze). In this case, for the high performing students to have their similarity cause their focus, the students would need to have a shared knowledge space. Similar gaze may be cognition-driven (referential gestures, familiarity with the interface, or prior knowledge), which is top-down behaviour (Connor, Egeth, & Yantis, 2004). On the other hand, one can hypothesize that individual focus may be caused by the student reacting to a stimulus (on the screen or in their partner's dialogue), which is bottom-up behaviour (Connor, Egeth, 7 Yantis, 2004). The low performing students may have been reacting to a stimulus on the screen, leading to focus and similarity. This similarity could then allow them to share information and collaborate to overcome their difficulties. Our results show that examining the causality between collaborative and individual gaze patterns may unveil intriguing cognitive mechanisms underlying the collaborative learning with tutoring systems.

To develop collaboration interventions, it is not enough to understand the overall trends within a learning session. The temporal causalities also need to be analysed to be able to pinpoint moments where an intervention would be beneficial. Like with the overall trends, the pairs with high posttest score show a 'top-down' gaze behaviour for most of the collaborative session, whereas 'bottom-up' gaze behaviour takes precedence for the pairs with low posttest scores after a small period of 'top-down' gaze behaviour. The moments when the causality changes its direction are the key moments to intervene. For example, in their study of macro-cognition in teams, Fiore, Smith-Jentsch, Salas, Warner, and Letsky (2010) proposed interventions/feedback to be at both the individual and team levels to increase the team knowledge similarity. By focusing our intervention on moments when the individual focus is driving the similarity, we can put interventions in place to support the collaboration.

However, it is not enough to only identify moments when an intervention may be beneficial. From this work, it is still an open question of what that intervention should entail, which must be based on more than a proxy measure and instead align with the underlying cognitive processes that the students are engaged in. What the specific intervention should be depends upon how the students are struggling. To intervene you need to know what actions you expect to see of the students and at this point in time, which ones they are not displaying so you can prompt for those specifically. Otherwise, the intervention may not be beneficial as you will be providing scaffolding for an indicator of collaboration, such as gaze, rather than the underlying issue. Our results showed that students with lower test scores had more moments of individual focus causing similarity, indicating that these moments may not have been beneficial to the students. Moreover, the interaction effect of pairs' success and the causal direction on the dialogue types also indicates different collaborative behaviour by the pairs in similar causality episodes. Such episodes might serve as the key episodes for the intervention. To provide an effective intervention, it is important to understand what the underlying learning processes are so that an intervention can be put into place to address these processes rather than trying to address a symptom. In future work, we aim to explore what the causal patterns may be indicative of in the students' learning to be able to put into place productive interventions that go beyond gaze. We can then combine the analysis presented in this article to find moments for interventions with the designed interventions to investigate their impact on learning.

One limitation of our work is the small student sample size that we used for analysis. With a total of 28 dyads, the results may not be indicative of what we would find with more students participating and should be addressed in future work. Additionally, our sample was gathered in a lab setting and the interactions of the students may differ from how they would act within a classroom with less supervision. However, a limitation of eye-tracking is that it is not necessarily very feasible to use in a classroom setting yet. To expand this work and address the feasibility of using gaze data in the classroom, we propose that the causal relationship between other student actions can also be explored. For example, in addition to the two students, the ITS also provides feedback and could be a driving force for different student cognitive processes.

Despite these limitations, by using the nature of the causal relationship between the gaze measures, we are able to provide adaptation for collaborative moments that are not necessarily captured in the system. While the students are problem-solving, they are not necessarily interacting with the system, meaning that there may be gaps in the log data. However, the speech can be hard to parse to provide an intervention and gaze data can then help to fill in these gaps by providing a continuous data stream. In this article, we have shown that the gaze of the student can be used to assess the collaboration. In future work, we aim to apply our findings to provide an intervention to support student collaborative learning based upon their gaze.

7 | CONCLUSION

In this article, we investigated the causal relationship between individual and collaborative cognitive processes with gaze measures as a proxy to provide more insight into the collaborative learning process. Our work contributes to the use of eye-tracking to enhance our understanding of computer-supported collaborative learning. Much of the previous work in dual eye-tracking has only investigated the correlational relationships involving eye-tracking in a collaborative learning environment (Belenky et al., 2014; Jermann & Nüssli, 2012; Sangin et al., 2011; Sharma et al., 2013). Our work adds to this field by investigating the causal relationship between individual and collaborative gaze measures. Specifically, our finding that the nature of the causal relationship changes depending upon the context of the learning provides insight into the use of eye-tracking in future research.

Additionally, by understanding the causality, we can better use these measures to assess the collaborative state of students as they work with an ITS and develop interventions to guide the collaborative process. This work contributes to adaptive learning by revealing temporal causality relations between individual and collaborative gaze measures that can be used to assess the collaboration of a group so that interventions can be applied at the correct moments. The analysis in this article brings us closer to providing collaboration support to students between interactions with the system without having to parse student speech allowing students to collaborate more effectively.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

As it is possible to identify participants from the data, ethical requirements do not permit us to share participant data from this study.

ETHICS STATEMENT

Participation was voluntarily, and all the data collected anonymously. Appropriate permissions and ethical approval for the participation requested and approved.

ORCID

Kshitij Sharma https://orcid.org/0000-0003-3364-637X

REFERENCES

Adamson, D., Dyke, G., Jang, H., & Rosé, C. P. (2014). Towards an agile approach to adapting dynamic collaboration support to student needs. *International Journal of Artificial Intelligence in Education*, *24*(1), 92–124.

Adamson, D., & Rosé, C. P. (2012). Coordinating multi-dimensional support in collaborative conversational agents. In *International conference on intelligent tutoring systems* (pp. 346–351). Berlin, Heidelberg: Springer.

Akaike, H. (1974). A new look at the statistical model identification. In *Selected papers of Hirotugu Akaike* (pp. 215–222). New York, NY: Springer.

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. *Journal of Memory and Language*, *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), 056221.

Baghaei, N., Mitrovic, A., & Irwin, W. (2007). Supporting collaborative learning and problem-solving in a constraint-based CSCL environment for UML class diagrams. *International Journal of Computer-Supported Collaborative Learning*, *2*(2–3), 159–190.

Baker, M., & Lund, K. (1997). Promoting reflective interactions in a CSCL environment. *Journal of Computer Assisted Learning*, *13*(3), 175–193.

Belenky, D., Ringenberg, M., Olsen, J., Aleven, V., & Rummel, N. (2014). *Using dual eye-tracking to evaluate students' collaboration with an intelligent tutoring system for elementary-level fractions*, Quebec City, Canada: Grantee Submission.

Bergner, Y., Walker, E., & Ogan, A. (2017). Dynamic Bayesian network models for peer tutoring interactions. In *Innovative assessment of collaboration* (pp. 249–268). Cham, Switzerland: Springer.

Bixler, R., & D'Mello, S. (2016). Automatic gaze-based user-independent detection of mind wandering during computerized reading. *User Modeling and User-Adapted Interaction*, *26*(1), 33–68.

Bollen, K. A., & Pearl, J. (2013). Eight myths about causality and structural equation models. In *Handbook of causal analysis for social research* (pp. 301–328). Dordrecht, the Netherlands: Springer.

Chambliss, D. F., & Schutt, R. K. (2018). *Making sense of the social world: Methods of investigation*, CA, USA: Sage.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*(6), 893–910.

Chen, Y., Rangarajan, G., Feng, J., & Ding, M. (2004). Analyzing multiple nonlinear time series with extended Granger causality. *Physics Letters A*, *324*(1), 26–35.

Cherubini, M., Nüssli, M. A., & Dillenbourg, P. (2008). Deixis and gaze in collaborative work at a distance (over a shared map): A computational model to detect misunderstandings. In *Proceedings of the 2008 symposium on eye tracking research & applications* (pp. 173–180). Georgia, USA: ACM.

Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*(4), 219–243.

Connor, C. E., Egeth, H. E., & Yantis, S. (2004). Visual attention: Bottom-up versus top-down. *Current Biology*, *14*(19), R850–R852.

D'Angelo, S., & Begel, A. (2017). Improving communication between pair programmers using shared gaze awareness. In *Proceedings of the 2017 CHI conference on human factors in computing systems* (pp. 6245–6290). Denver, Colorado, USA: ACM.

D'Mello, S., Olney, A., Williams, C., & Hays, P. (2012). Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of Human-Computer Studies*, *70*(5), 377–398.

Dănăcică, D. E., Belaşcu, L., & Ilie, L. (2010). The interactive causality between higher education and economic growth in Romania. *International Review of Business Research Papers*, *6*(4), 491–500.

Diehl, M., & Stroebe, W. (1987). Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of Personality and Social Psychology*, *53*(3), 497–509.

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. *Biological Cybernetics*, *83*(1), 35–45.

Ding, M., Chen, Y., & Bressler, S. L. (2006). Chapter 17: Granger causality: Basic theory and application to neuroscience. In *Handbook of time series analysis: Recent theoretical developments and applications* (p. 437). Wienheim, Germany: Wiley.

Diziol, D., Walker, E., Rummel, N., & Koedinger, K. R. (2010). Using intelligent tutor technology to implement adaptive support for student collaboration. *Educational Psychology Review*, *22*(1), 89–102.

Dowell, N. M., Cade, W. L., Tausczik, Y., Pennebaker, J., & Graesser, A. C. (2014, June). What works: Creating adaptive and intelligent systems for collaborative learning support. In *International conference on intelligent tutoring systems* (pp. 124–133). Cham, Switzerland: Springer.

Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, *5* (2), 155–174.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*, MA, USA: MIT Press.

Evans, A. C., Wobbrock, J. O., & Davis, K. (2016). Modeling collaboration patterns on an interactive tabletop in a classroom setting. In *Proceedings of the 19th ACM Conference on computer-supported cooperative work & social computing* (pp. 860–871). San Francisco, USA: ACM.

Feng, S., D'Mello, S., & Graesser, A. C. (2013). Mind wandering while reading easy and difficult texts. *Psychonomic Bulletin & Review*, *20*(3), 586–592.

Fiore, S. M., Smith-Jentsch, K. A., Salas, E., Warner, N., & Letsky, M. (2010). Towards an understanding of macrocognition in teams: Developing and defining complex collaborative processes and products. *Theoretical Issues in Ergonomics Science*, *11*(4), 250–271.

Freiwald, W. A., Valdes, P., Bosch, J., Biscay, R., Jimenez, J. C., Rodriguez, L. M., … Singer, W. (1999). Testing non-linearity and directedness of interactions between neural groups in the macaque inferotemporal cortex. *Journal of Neuroscience Methods*, *94*(1), 105–119.

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. *Magnetic Resonance Imaging*, *21*(10), 1251–1261.

Goudas, M., Biddle, S., & Fox, K. (1994). Perceived locus of causality, goal orientations, and perceived competence in school physical education classes. *British Journal of Educational Psychology*, *64*(3), 453–463.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, *37*, 424–438.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, *11*(4), 274–279.

Harsley, R., Di Eugenio, B., Green, N., Fossati, D., & Acharya, S. (2016). Integrating support for collaboration in a computer science intelligent tutoring system. In *International conference on intelligent tutoring systems* (pp. 227–233). Cham, Switzerland: Springer.

Hesse, W., Möller, E., Arnold, M., & Schack, B. (2003). The use of time-variant EEG Granger causality for inspecting directed interdependencies of neural assemblies. *Journal of Neuroscience Methods*, *124*(1), 27–44.

Hiemstra, C., & Jones, J. D. (1994). Testing for linear and nonlinear Granger causality in the stock price-volume relation. *The Journal of Finance*, *49*(5), 1639–1664.

Hill, A. B. (1965). *The environment and disease: Association or causation*, 58 (1), 295–300. London, UK: Proceedings of the Royal Society of Medicine.

Hoffman, J. E. (1998). Visual attention and eye movements. *Attention*, *31*, 119–153.

Islam, T. S., Wadud, M. A., & Islam, Q. B. T. (2007). Relationship between education and GDP growth: A multivariate causality analysis for Bangladesh. *Economics Bulletin*, *3*(35), 1–7.

Janssen, J., & Bodemer, D. (2013). Coordinated computer-supported collaborative learning: Awareness and awareness tools. *Educational Psychologist*, *48*(1), 40–55.

Jaques, N., Conati, C., Harley, J. M., & Azevedo, R. (2014). Predicting affect from gaze data during interaction with an intelligent tutoring system. In *International conference on intelligent tutoring systems* (pp. 29–38). Cham, Switzerland: Springer.

Jermann, P., & Nüssli, M. A. (2012). Effects of sharing text selections on gaze cross-recurrence and interaction quality in a pair programming task. In *Proceedings of the ACM 2012 conference on computer supported cooperative work* (pp. 1125–1134). Washington, USA: ACM.

Joerding, W. (1986). Economic growth and defense spending: Granger causality. *Journal of Development Economics*, *21*(1), 35–40.

King, A. (1999). *Discourse patterns for mediating peer learning*, London, UK: Lawrence Erlbaum Associates Publishers.

Kodde, D. A., & Palm, F. C. (1986). Wald criteria for jointly testing equality and inequality restrictions. *Econometrica: Journal of the Econometric Society*, *54*, 1243–1248.

Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research*, *86*(1), 42–78.

Kumar Narayan, P., & Smyth, R. (2006). Higher education, real income and real investment in China: Evidence from Granger causality tests. *Education Economics*, *14*(1), 107–125.

Laughlin, P. R., Hatch, E. C., Silver, J. S., & Boh, L. (2006). Groups perform better than the best individuals on letters-to-numbers problems: Effects of group size. *Journal of Personality and Social Psychology*, *90*(4), 644–651.

Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology*, *106*(4), 901–918.

Magnisalis, I., Demetriadis, S., & Karakostas, A. (2011). Adaptive and intelligent systems for collaborative learning support: A review of the field. *IEEE Transactions on Learning Technologies*, *4*(1), 5–20.

Martinez, R., Wallace, J. R., Kay, J., & Yacef, K. (2011). Modelling and identifying collaborative situations in a collocated multi-display groupware setting. In *International conference on artificial intelligence in education* (pp. 196–204). Berlin, Germany: Springer.

Martinez-Maldonado, R., Kay, J., & Yacef, K. (2013). An automatic approach for mining patterns of collaboration around an interactive tabletop. In *International conference on artificial intelligence in education* (pp. 101–110). Berlin, Germany: Springer.

McLaren, B. M., Scheuer, O., & Mikšátko, J. (2010). Supporting collaborative learning and e-discussions using artificial intelligence techniques. *International Journal of Artificial Intelligence in Education*, *20*(1), 1–46.

McManus, M. M., & Aiken, R. M. (2016). Supporting effective collaboration: Using a rearview mirror to look forward. *International Journal of Artificial Intelligence in Education*, *26*(1), 365–377.

Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition*, *66* (2), B25–B33.

Murcia, J. A. M., Coll, D. G. C., & Garzón, M. C. (2009). Preliminary validation in Spanish of a scale designed to measure motivation in physical education classes: The perceived locus of causality (PLOC) scale. *The Spanish Journal of Psychology*, *12*(1), 327–337.

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.

Nokes, T. J. (2009). Mechanisms of knowledge transfer. *Thinking & Reasoning*, *15*(1), 1–36.

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. *International Society of the Learning Sciences (ISLS)*, *1*, 689–696.

Olsen, J. K., Belenky, D. M., Aleven, V., & Rummel, N. (2014). Using an intelligent tutoring system to support collaborative as well as individual learning. In *International conference on intelligent tutoring systems* (pp. 134–143). Cham, Switzerland: Springer.

Olsen, J. K., Rummel, N., & Aleven, V. (2016). *Investigating effects of embedding collaboration in an intelligent tutoring system for elementary school students*, Singapore: Grantee Submission.

Poole, A., & Ball, L. J. (2006). Eye tracking in HCI and usability research. In *Encyclopedia of human computer interaction* (pp. 211–219). Pennsylvania, United States: IGI Global.

Prieto, L. P., Sharma, K., & Dillenbourg, P. (2015). Studying teacher orchestration load in technology-enhanced classrooms. In *Design for teaching and learning in a networked world* (pp. 268–281). Cham, Switzerland: Springer.

Raca, M., & Dillenbourg, P. (2013). System for assessing classroom attention. In *Proceedings of 3rd international learning analytics*, Leuven, Belgium: ACM.

Rau, M., Scheines, R., Aleven, V., & Rummel, N. (2013). Does representational understanding enhance fluency-or vice versa? Searching for mediation models. In *Educational data mining* (p. 2013). TN, USA: International Educational Data Mining Society.

Rau, M. A., Michaelis, J. E., & Fay, N. (2015). Connection making between multiple graphical representations: A multi-methods approach for domain-specific grounding of an intelligent tutoring system for chemistry. *Computers & Education*, *82*, 460–485.

Richardson, D. C., & Dale, R. (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive science*, *29*(6), 1045–1060.

Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The art of conversation is coordination. *Psychological Science*, *18*(5), 407–413.

Ritter, S., Anderson, J. R., Koedinger, K. R., & Corbett, A. (2007). Cognitive tutor: Applied research in mathematics education. *Psychonomic Bulletin & Review*, *14*(2), 249–255.

Rodríguez, F. J., & Boyer, K. E. (2015). Discovering individual and collaborative problem-solving modes with hidden Markov models. In *International conference on artificial intelligence in education* (pp. 408–418). Cham, Switzerland: Springer.

Rosatelli, M. C., & Self, J. A. (2004). A collaborative case study system for distance learning. *International Journal of Artificial Intelligence in Education*, *14*(1), 97–125.

Rummel, N., Mullins, D., & Spada, H. (2012). Scripted collaborative learning with the cognitive tutor algebra. *International Journal of Computer-Supported Collaborative Learning*, *7*(2), 307–339.

Rummel, N., & Spada, H. (2005). Learning to collaborate: An instructional approach to promoting collaborative problem-solving in computer-mediated settings. *Journal of the Learning Sciences*, *14*(2), 201–241.

Sangin, M., Molinari, G., Nüssli, M. A., & Dillenbourg, P. (2011). Facilitating peer knowledge modeling: Effects of a knowledge awareness tool on collaborative learning outcomes and processes. *Computers in Human Behavior*, *27*(3), 1059–1067.

Schneider, B., Sharma, K., Cuendet, S., Zufferey, G., Dillenbourg, P., & Pea, R. (2016). Using mobile eye-trackers to unpack the perceptual benefits of a tangible user interface for collaborative learning. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *23*(6), 39.

Schooler, J. W., Ohlsson, S., & Brooks, K. (1993). Thoughts beyond words: When language overshadows insight. *Journal of Experimental Psychology: General*, *122*(2), 166–183.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464.

Semin, G. R., & Smith, E. R. (2008). *Embodied grounding: Social, cognitive, affective, and neuroscientific approaches*, Cambridge, UK: Cambridge University Press.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*, Boston New York: HOUGHTON MIFFLIN COMPANY.

Sharma, K., Alavi, H. S., Jermann, P., & Dillenbourg, P. (2016). A gaze-based learning analytics model: In-video visual feedback to improve learner's attention in MOOCs. In *Proceedings of the sixth international conference on learning analytics & knowledge* (pp. 417–421). Edinburgh, Scotland: ACM.

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 (ISLS)*, *1*, 260–267.

Sharma, K., Jermann, P., Nüssli, M. A., & Dillenbourg, P. (2013). Understanding collaborative program comprehension: Interlacing gaze and dialogues. In *Proceedings of computer supported collaborative learning (CSCL 2013)* (Vol. 1, pp. 430–437). Wisconsin, USA: ISLS.

Slavin, R. E. (1996). Research on cooperative learning and achievement: What we know, what we need to know. *Contemporary Educational Psychology*, *21*(1), 43–69.

Soller, A., Martínez, A., Jermann, P., & Muehlenbrock, M. (2005). From mirroring to guiding: A review of state of the art technology for supporting collaborative learning. *International Journal of Artificial Intelligence in Education*, *15*(4), 261–290.

Steiner, I. D. (1972). *Group process and productivity*, (393–422). New York, NY: Academic Press.

Taylor, K. L., & Dionne, J. P. (2000). Accessing problem-solving strategy knowledge: The complementary use of concurrent verbal protocols and retrospective debriefing. *Journal of Educational Psychology*, *92*(3), 413–425.

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.

Van Gog, T., Kester, L., Nievelstein, F., Giesbers, B., & Paas, F. (2009). Uncovering cognitive processes: Different techniques that can contribute to cognitive load research and instruction. *Computers in Human Behavior*, *25*(2), 325–331.

van Ulzen, N. R., Lamoth, C. J., Daffertshofer, A., Semin, G. R., & Beek, P. J. (2008). Characteristics of instructed and uninstructed interpersonal coordination while walking side-by-side. *Neuroscience Letters*, *432*(2), 88–93.

VanLehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education*, *16*(3), 227–265.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, *46*(4), 197–221.

Vesper, C., Soutschek, A., & Schubö, A. (2009). Motion coordination affects movement parameters in a joint pick-and-place task. *The Quarterly Journal of Experimental Psychology*, *62*(12), 2418–2432.

Viswanathan, S. A., & Vanlehn, K. (2017). *High accuracy detection of collaboration from log data and superficial speech features*. Philadelphia, PA: International Society of the Learning Sciences.

Viswanathan, S. A., & VanLehn, K. (2018). Using the tablet gestures and speech of pairs of students to classify their collaboration. *IEEE Transactions on Learning Technologies*, *11*(2), 230–242.

Walker, E., Rummel, N., & Koedinger, K. R. (2011). Designing automated adaptive support to improve student helping behaviors in a peer tutoring activity. *International Journal of Computer-Supported Collaborative Learning*, *6*(2), 279–306.

Walker, E., Rummel, N., & Koedinger, K. R. (2014). Adaptive intelligent support to improve peer tutoring in algebra. *International Journal of Artificial Intelligence in Education*, *24*(1), 33–61.

Walker, E., Walker, S., Rummel, N., & Koedinger, K. R. (2010). Using problem-solving context to assess help quality in computer-mediated peer tutoring. In *International conference on intelligent tutoring systems* (pp. 145–155). Berlin, Germany: Springer.

Weinberger, A., Ertl, B., Fischer, F., & Mandl, H. (2005). Epistemic and social scripts in computer-supported collaborative learning. *Instructional Science*, *33*(1), 1–30.

Weldon, M. S., & Bellinger, K. D. (1997). Collective memory: Collaborative and individual processes in remembering. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(5), 1160.

How to cite this article: Sharma K, Olsen JK, Aleven V, Rummel N. Measuring causality between collaborative and individual gaze metrics for collaborative problem-solving with intelligent tutoring systems. *J Comput Assist Learn*. 2020;1–18. https://doi.org/10.1111/jcal.12467