Measuring Dynamical Interdependence in Small-Group Collaborations

Peter Halpin[1]

Abstract

This paper addresses dynamical interdependence among the actions of group members. I assume that the actions of each member can be represented as nodes of a dynamical network and then collect the nodes into disjoint subsets (components) representing the individual group members. Interdependence among group members' actions can then be defined with reference to a K-partite network, in which the partitions correspond to the group member components. Independence among group members' actions can be defined with reference to a network in which the group member components are disconnected from one another. The degree to which the interactions of actual groups correspond to either of these theoretical network structures can be characterized using modified versions of existing network statistics. Taking this approach, I propose a number of network-based measures of dynamical interdependence, discuss the interpretation of the proposed measures, and consider how to assess their reliability and validity. These ideas are illustrated using an example in which dyads collaborated via online chat to complete a grade 12 level mathematics assessment.

Notes for Practice
- This article shows how time-intensive process data can be used to measure the degree of interdependence among the actions of group members.
- The proposed measures of interdependence predict group performance in mathematics above and beyond (a) the mathematical ability of the group members and (b) the total number of interactions among group members, which suggests the importance of these measures for understanding small-group collaborations.
- The reliability of the proposed measures increases as the number of interactions among members increases, implying that short-duration learning activities are not well suited for studying interdependence.
- The study of groups with three or more members is important for understanding the contributions that individual members make to group dynamics; in dyads, the influence of one group member is not distinguishable from the responsiveness of the other.

[1] Email: peter.halpin@unc.edu Address: University of North Carolina at Chapel Hill, School of Education, Office 1070G, CB 3500 Peabody Hall, Chapel Hill, NC 27599-3500, USA.

1. Introduction

One challenge that arises in the study of small-group collaborations is how to describe interdependence among the actions of group members. The notion of interdependence among group members has been variously interpreted and is undeniably central to the study of small groups in educational settings. For example, positive interdependence (Johnson & Johnson, 2009) and group-worthy tasks (Cohen, 1994) have been theorized in terms of groups whose members depend upon one another in order to achieve shared goals. The concepts of transactivity (Teasley, 1997) and interactivity (Chi & Wylie, 2014) emphasize how learners build upon one another's contributions. In each of these theoretical approaches, interdependence among group members, as opposed to independence, is hypothesized to be a key requirement for realizing the learning benefits of small-group interactions.

In this paper I take a quantitative approach to measuring interdependence that is applicable to process data (traces) collected in technology-mediated settings. In this context, the challenge may be characterized as follows: (a) we have multiple time-intensive sources of data about the actions of each group member, and (b) the focus is to understand not only how the actions of each

group member depend on their own past actions, but also (c) how each group member's actions depend on the past actions of their partners. I refer to the latter in terms of dynamical interdependence. The goals of this paper are to propose a relatively general blueprint for measuring dynamical interdependence and to discuss how to evaluate the quality of such measures from a psychometric perspective. These measures can, in turn, be useful for empirically evaluating theories about the role of interdependence in collaborative learning, informing the design of group-worthy learning activities, and providing feedback to students and educators about small-group interactions.

This paper can be situated more broadly in the context of temporal analytics (B. Chen, Knight, & Wise, 2018; Knight, Friend Wise, & Chen, 2017). While process data are increasingly available to researchers, making sense of time-intensive, low-level variables (features) with respect to higher-level behavioural and cognitive constructs is a thorny problem. One aspect of this problem is measurement, i.e., how to make inferences from the features (e.g., group members' actions) to the concepts of interest (e.g., interdependence). Psychometric theory provides a well-established toolkit for describing the reliability, validity, and fairness of such inferences, but these tools have not been widely applied outside the context of conventional educational assessments. Two broader goals of the present research are to provide a methodological framework for porting over some useful tools from psychometric theory to the temporal analysis of process data and to discuss the relevance of those tools in this new context.

It is also important to clarify the terminology used in this paper. I use the term *dynamical* to mean change over time, broadly conceived, rather than a specific statistical framework for studying change over time (e.g., dynamical systems). This usage is consistent with the literature on dynamical networks (e.g., Barrat, Barthelemy, & Vespignani, 2011; Jost, 2007) and group dynamics (e.g., Kozlowski & Chao, 2018), both of which I draw on in this paper. I also use the words *dynamical* and *temporal* interchangeably throughout the paper, and, in places, somewhat redundantly (e.g., *temporal dynamics*). This looseness of terminology is consistent with the idea that dynamical statistical models are essentially about temporal statistical dependence, e.g., how the present depends on the past. Accordingly, readers may conceptualize the paper as being about temporal interdependence, and this would convey the intended meaning. The term *dynamical* connotes the literature I draw on but is not intended to convey a contrast with the term *temporal*.

The proposed approach to measuring dynamical interdependence can be explained with reference to Figure 1, which depicts a fully connected network among three types of actions recorded for each of three group members. The nodes are process variables or features representing group members' actions. Their colour represents the group member who performed the action, and the node name represents the type of action (e.g., sending chats, requesting hints, in-task responses, gaze, gesture). The double-arrowed edges represent temporal dynamics, such that the history of the one node predicts the future of the other, and vice versa. More details on this representation are given in the following section. For now, let's simply note that this network has many low-level connections among the individual group members' actions but does not provide an obvious way to talk about the overall interdependence among the group members, or the roles of specific individuals or specific types of actions therein. These higher-level characteristics have been described as emergent properties of group process dynamics, meaning that they describe global regularities that arise from many individual interactions (Kozlowski & Chao, 2018).
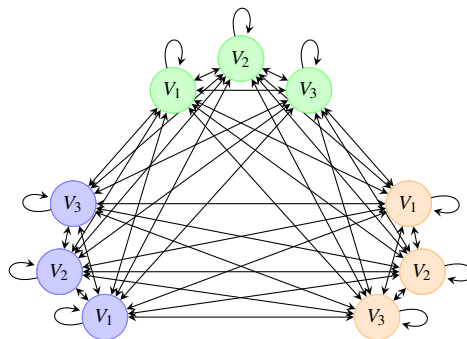
Figure 1. Representation of collaborative process data as a fully connected dynamical network. Nodes represent group members' actions, the node colour represents the group member who performed the action, and the node name represents the type of action. The edges represent temporal statistical dependence among the nodes.

Following this rationale, this paper considers some ways to summarize person-, action-, and group-level aspects of dynamical networks like the one depicted in Figure 1. In terms of data analysis, this boils down to a post-processing step for dynamical models fitted to process data obtained from small groups. The post-processing consists of generalized versions of the usual in- and out-degree statistics for weighted directional networks (e.g., Hanneman & Riddle, 2014). This paper focuses on the interpretation of these statistics, as well as how to evaluate their reliability and validity. Specifically, I address the following

topics:

1. Interpretation: How can dynamical networks be interpreted in the context of small-group collaborations, and how can they be used to define process-based measures that describe interdependence among group members?

2. Reliability: Why is the reliability of these process-based measures important, and how can it be estimated?

3. Validity: Do process-based measures explain variation in group outcomes beyond the traits of individual members (e.g., student knowledge) or simple static summaries of process (e.g., word count)?

In this paper, I do not give a lot of attention to specific types of statistical models for dynamical networks or their estimation. There are extensive literatures on such models, including vector auto-regression (e.g., Lütkepohl, 2005), point processes (e.g., Daley & Vere-Jones, 2003), Markov models and state-space models (e.g., Durbin & Koopman, 2012), dynamic factor analysis (e.g., Molenaar, 1985), and multimodal analytics (e.g., Oviatt et al., 2018). From a purely technical perspective, these models can be ported over more or less directly to small-group settings. Here the focus is on how to use the information provided by such models to describe the emergent dynamics of interactions among group members. I also draw connections with some related approaches specifically developed for small-group dynamics (e.g., Butts, 2008; Dowell, Nixon, & Graesser, 2019; Halpin, von Davier, Hao, & Liu, 2017; Quera, 2018; Suthers, Dwyer, Medina, & Vatrapu, 2010).

These ideas are illustrated using a dataset in which $N = 104$ dyads collaborated via online chat to complete a grade 12 level mathematics assessment. I focus on three types of actions for each member of a dyad: (1) responding to test questions and sending chat messages (2) with and (3) without math content. I use a standard Markov chain to model temporal dependence in the sequence of events for each group. While the modelling approach is entirely standard, the network statistics introduced in the following section provide a novel approach to describing the interdependence among group members' actions. The example data have a number of limitations, notably that group size is limited to dyads, and I discuss the significance of this limitation throughout the paper. The paper closes with discussion of some future research directions.

2. Some Measures of Dynamical Interdependence for Collaborative Process Data

This section first develops a notation for representing the network in Figure 1 and then uses this notation to decompose the network's adjacency matrix into two disjoint components, one representing interpersonal dynamics and the other representing intra-personal dynamics. The former is used to develop a number of summary measures that describe a group's interpersonal dynamics in terms of the individual group members, the different types of actions performed by the group, and the group as a whole. The focus of this section is the interpretation of the measures, and the following sections address their reliability and validity.

2.1 Preliminaries

Let $i = 1, \ldots, I$ index a set of features encoded from the process data of each group member $j = 1, \ldots, J$. The features are denoted V_{ij} and correspond to the nodes in Figure 1. It is assumed that features are organized so that the vector $\mathbf{u}_i = [V_{i1} \cdots V_{iJ}]$ describes the same type of action (e.g., sending a chat message, responding to a problem), which will be referred to as the ith action type component, or the ith action for short.

It is assumed that all features are measured over time, so that the data consist of a sequence of realizations of each feature. The sequential nature of the data is not explicitly represented in Figure 1 — it is implicit in the interpretation of the edges as representing temporal dependence. This interpretation of networks is exemplified by the state-transition diagrams commonly used with Markov chains and other state-space models. Network diagrams can also be used to explicitly represent time, e.g., by subscripting each node with a time index t and representing the relationships among the nodes as unfolding along a time axis. The latter representation is useful when there are relatively few variables in play and the goal is to represent specific types of temporal relationships (e.g., lag 1, lag 2). The choice to represent time implicitly was made to simplify presentation and avoid tangential details of specific modelling approaches.

Let the vector $\mathbf{v}_j = [V_{1j} \cdots V_{Ij}]$ contain the features of group member j, which will be referred to as the jth group member component, or the jth member for short. The vector $\mathbf{v} = [\mathbf{v}_1, \ldots, \mathbf{v}_J]$ contains the full set of $N = IJ$ features. When indexing the individual features within $\mathbf{v}$, the subscripts $m, n \in \{(j-1)I + i \mid i = 1, \ldots, I, j = 1, \ldots, J\}$ will be used.

Let $\mathbf{A} = \{\alpha_{mn}\}$ denote an $N \times N$ adjacency matrix describing statistical dependence among the feature set $\mathbf{v}$. The elements α_{mn} correspond to the directed edges in Figure 1. In the context of dynamical networks, these describe how the probability distribution of feature n depends on the past values of feature m. For example, these might be state transitions in a Markov-type model, regression coefficients in a vector auto-regression, rates or intensities in a point process, and so on. I assume that the α_{mn} are scaled to allow for meaningful comparisons of their magnitudes.

Regardless of the particular modelling framework applied, α_{mm} describes the dependence of a feature on its own past (i.e., auto-dependence). When $m \neq n$, α_{mn} describes the dependence of a feature on the past of some other feature (i.e., cross-dependence). There are two types of cross-dependence in Figure 1 — cross-dependence among the actions of a single group member (i.e., among nodes of the same colour) and cross-dependence among the actions of different groups members (i.e., among nodes of different colours). As mentioned above, I refer to the latter in terms of dynamical interdependence. I refer to the former as dynamical intra-dependence, and also include auto-dependence under this rubric. These inter- and intra-personal temporal dependencies collectively describe the dynamics of the network.

2.2 Separating Inter- and Intra-personal Dependence

Next, let's consider how to decompose the adjacency matrix **A** into parts describing the inter- and intra-personal dynamics of the group. The overall idea is depicted in Figure 2, in which the fully connected network in Figure 1 is decomposed into two sub-networks with the same nodes but different adjacency matrices.
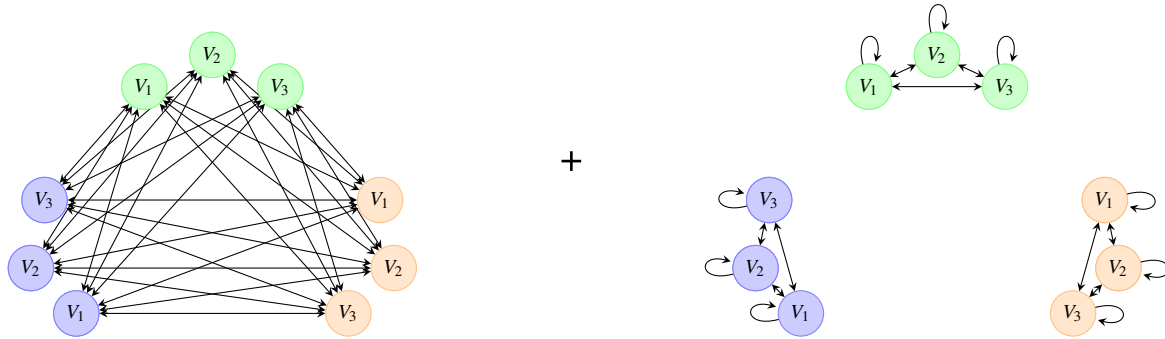
Figure 2. Decomposition of the fully connected network in Figure 1 into two sub-networks. The left panel shows a 3-partite network in which the partitions correspond to the three group members. The right panel shows a disconnected graph with three components corresponding to the group members.

The left-hand panel of Figure 2 depicts a K-partite graph, in which the $K = 3$ partitions correspond to the group members. In this sub-network, only the interdependence among group members' actions is included. The right-hand panel of Figure 2 depicts a disconnected graph, in which the separate components correspond to the group members. In this sub-network, only intra-dependence is considered.

It should be emphasized that the K-partite graph and the disconnected graph in Figure 2 are both extreme cases, and neither is likely to characterize actual groups very well. It is not proposed that either network structure in Figure 2 should be used for data analysis. Rather, it is proposed to summarize the inter- and intra-personal dynamics of a given network, such as that in Figure 1, by decomposing it into these two structures.

Mathematically, the decomposition proceeds as follows. First note that **A** is organized as a block-wise matrix, with blocks of dimension $I \times I$ denoted $\mathbf{A}_{jk}$:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \cdots \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}.$$

The diagonal blocks $\mathbf{A}_{jj}$ contain all of the edges among the features of the jth group member component. The off-diagonal blocks $\mathbf{A}_{jk}$ contain all of the edges running from the jth member to the kth member.

Next, define the block-wise indicator matrix $\mathbf{I} = \{\mathbf{I}_{jk}\}$ with blocks $\mathbf{I}_{jj} = \mathbf{1}$ and $\mathbf{I}_{jk} = \mathbf{0}$, $j \neq k$, each of size $I \times I$. Letting $\circ$ represent the Hadamard or element-wise product, the adjacency matrix $\mathbf{A}^{\sim} = \mathbf{A} \circ \mathbf{I}$ describes a network in which the group-member components are completely disconnected. This corresponds to the right-hand panel of Figure 2. Also define the matrix $\mathbf{A}^{*} = \mathbf{A} \circ (\mathbf{1} - \mathbf{I})$, where $\mathbf{1}$ is an $N \times N$ matrix of ones. $\mathbf{A}^{*}$ is a K-partite network with partitions corresponding to the group-member components, as depicted in the left-hand panel of Figure 2. As shown in the figure, $\mathbf{A}^{*}$ and $\mathbf{A}^{\sim}$ decompose **A**:

$$\mathbf{A}^{*} + \mathbf{A}^{\sim} = (\mathbf{A} - \mathbf{A} \circ \mathbf{I}) + \mathbf{A} \circ \mathbf{I} = \mathbf{A}. \tag{1}$$

This decomposition is the basis of the proposed measures of dynamical interdependence.

2.3 Some Measures of Interdependence

A large number of descriptive summaries can be used to characterize networks (e.g., Hanneman & Riddle, 2014), many of which have found useful applications in the analysis of small-group interactions (e.g., Butts, 2008; Dowell et al., 2019; Leenders, Contractor, & DeChurch, 2016; Swiecki, Ruis, Farrell, & Shaffer, 2019). Here I outline the interpretation of some summaries built from the out-degree and in-degree of individual nodes in the interpersonal dynamics matrix $\mathbf{A}^*$. A visual summary of these statistics is presented in Figure 3.
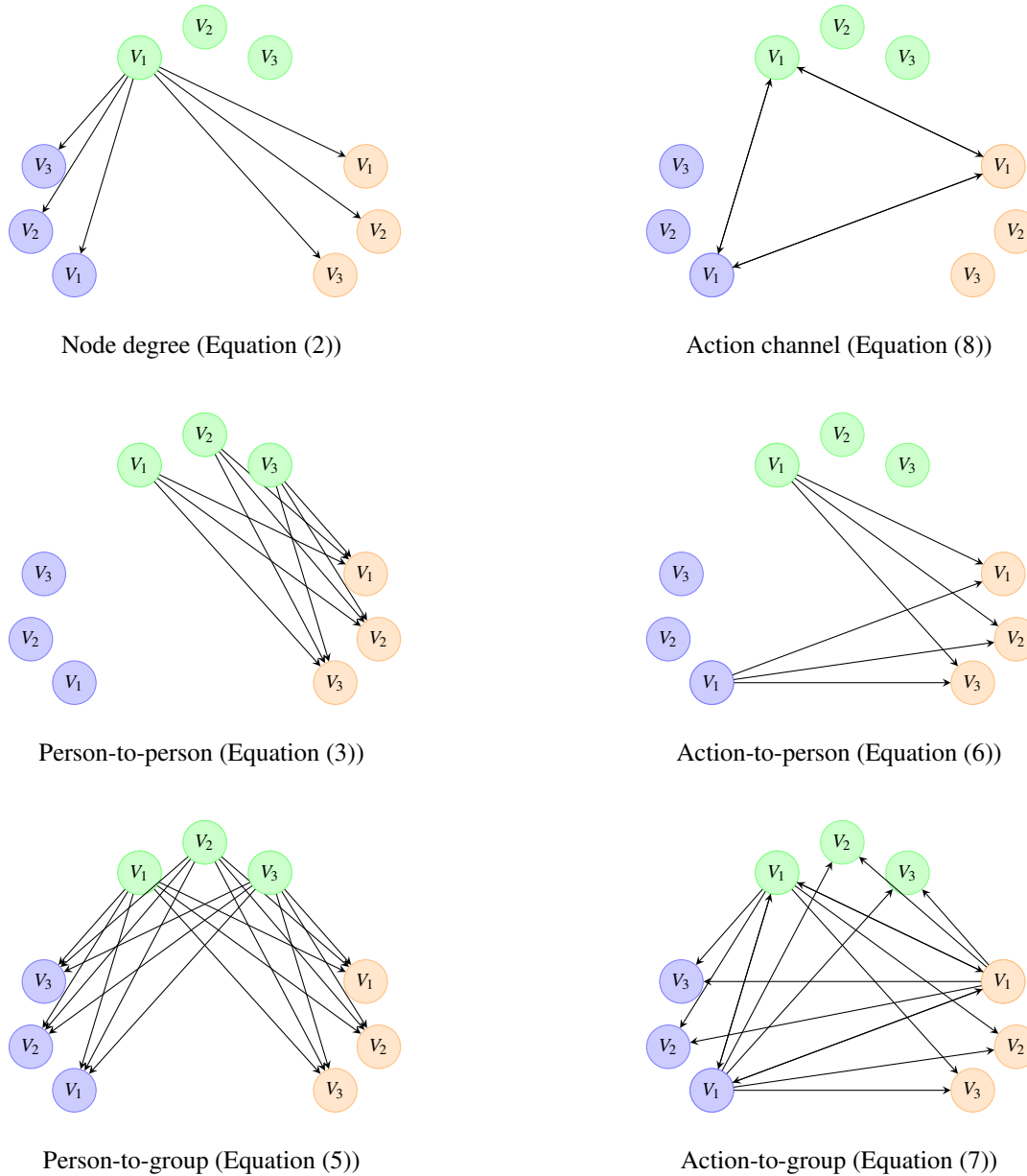
Figure 3. Network representation of some out-degree (influence) measures of dynamical interdependence built from the left-hand panel of Figure 2. The proposed measures are computed by averaging the indicated edges. For the corresponding in-degree (responsiveness) measures, the direction of the edges is reversed. The panel labels note the equations where the algebraic representations of the measures are provided.

Node out-degree is depicted in the top left-hand panel of Figure 3, and in-degree reverses the direction of the edges. In application to dynamical networks, out-degree describes how a node influences the future of the network, and in-degree describes how the node is influenced by the past of the network. Applied to the interdependence matrix $\mathbf{A}^*$, the out- and

in-degree for node V_{ij} are, respectively,

$$D^+(V_{ij}^*) = \frac{1}{N-I} \sum_{n=1}^{N} \alpha_{mn}^* \quad \text{and} \quad D^-(V_{ij}^*) = \frac{1}{N-I} \sum_{m=1}^{N} \alpha_{mn}^*. \tag{2}$$

Research addressing temporal dynamics in small-group settings has led to various interpretations of out-degree, including interpersonal influence (e.g., Madhyastha, Hamaker, & Gottman, 2011), talkativeness (Leenders et al., 2016), uptake (Suthers et al., 2010), and social impact (Dowell et al., 2019). I retain the generic dynamical interpretation of influence in this paper, while recognizing that more nuanced interpretations may be appropriate in a given application. Note that this interpretation of influence is not causal in a strict statistical sense — it may be interpreted along the lines of Granger causality, which is essentially temporal precedence combined with statistical association.

There are also a variety of ways of interpreting in-degree in small-group contexts, again depending on the specific types of interactions being considered. Some examples include prestige or popularity (Butts, 2008; Leenders et al., 2016), contingency (Suthers et al., 2010), and responsiveness (Dowell et al., 2019; Halpin et al., 2017). As mentioned above, in-degree can also be interpreted in terms of influence, but rather than a node influencing the future (i.e., out-degree), it describes how the node is influenced by the past. In this paper I adopt the term *responsiveness* as a generic term for in-degree, because it has a natural interpretation in group settings and avoids overloading the term *influence*.

The other summaries depicted in Figure 3 are essentially variations on this same theme, showing different ways of aggregating influence and responsiveness over the interpersonal dynamics matrix $\mathbf{A}^*$. To write out these measures, I adopt the following slightly more general notation for degree:

$$D^\pm (\mathbf{u} , \mathbf{w})$$

Degree valence

Focal features

Reference features
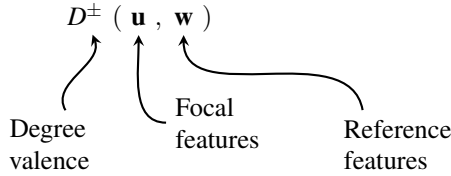
As above, the valence, $+$ or $-$, indicates whether the degree statistic is interpreted as measuring influence or responsiveness, respectively. The focal features are those whose degree we wish to describe, which can be any subset of the full set of features $\mathbf{v}$ defined above. This extends the usual definition of degree, which is applied to individual nodes. The reference features make explicit the scope of the influence or responsiveness of the focal features, which can also be any subset of $\mathbf{v}$. This also extends the usual definition of degree, in which the reference nodes include all nodes other than the focal node.

Using this notation, Equation (2) can be rewritten as

$$D^+(V_{ij}^*, \mathbf{v}^*) = D^+(V_{ij}^*) \quad \text{and} \quad D^-(V_{ij}^*, \mathbf{v}^*) = D^-(V_{ij}^*).$$

I make use of this more general notation throughout the following sections.

2.3.1 Person-Aligned Measures

First, let's consider some measures that arise when the focal features correspond to an individual group member. These types of measures are most appropriate when the goal is to describe the contributions of individuals to the overall interpersonal dynamics of the group. As noted below, some important limitations arise in the context of dyads.

Recall that features denoting the actions of group member j are collected in the vector $\mathbf{v}_j = [V_{1j}, V_{2j}, \ldots, V_{Ij}]$. The person-to-person measures in the middle-left panel of Figure 3 can be written

$$D^+(\mathbf{v}_j^*, \mathbf{v}_k^*) = \frac{1}{I^2} \sum_{m \in \mathcal{I}_j} \sum_{n \in \mathcal{I}_k} \alpha_{mn}^* \quad \text{and} \quad D^-(\mathbf{v}_j^*, \mathbf{v}_k^*) = \frac{1}{I^2} \sum_{m \in \mathcal{I}_k} \sum_{n \in \mathcal{I}_j} \alpha_{mn}^*, \tag{3}$$

with $\mathcal{I}_j = \{(j-1)I + i \mid V_{ij} \in \mathbf{v}_j\}$. Person-to-person out-degree describes the overall influence of one group member's actions on those of another. Person-to-person in-degree describes the overall responsiveness of one group member's actions to those of another.

There is an important equivalence between person-to-person influence and responsiveness:

$$D^+(\mathbf{v}_j^*, \mathbf{v}_k^*) = \frac{1}{I^2} \sum_{m \in \mathcal{I}_j} \sum_{n \in \mathcal{I}_k} \alpha_{mn}^* = D^-(\mathbf{v}_k^*, \mathbf{v}_j^*). \tag{4}$$

Equation (4) means that, when considering two group members at a time, it is not possible to distinguish the influence of one member from the responsiveness of the other.[1]

When there are more than two group members, this equivalence can be avoided by aggregating over referent persons. The resulting person-to-group out-degree describes the overall influence of one group member's actions on those of the entire group, and similarly for in-degree:

$$D^+(\mathbf{v}_j^*, \mathbf{v}^*) = \frac{1}{J-1} \sum_{k \neq j} D^+(\mathbf{v}_j^*, \mathbf{v}_k^*) \quad \text{and} \quad D^-(\mathbf{v}_j^*, \mathbf{v}^*) = \frac{1}{J-1} \sum_{k \neq j} D^-(\mathbf{v}_j^*, \mathbf{v}_k^*). \tag{5}$$

Clearly, the person-to-person and person-to-group measures are identical for dyads (i.e., when $J = 2$ in Equation (5)).

From these considerations, it can be concluded that person-aligned measures of influence and responsiveness are not well defined for dyads. This observation is in line with past research showing that many fundamental social network properties are not observable for dyads (e.g., Hanneman & Riddle, 2014, pp. 11 ff). This also leads to a specific recommendation for research in collaboration analytics: groups of at least size three must be studied in order to unambiguously define individual group members' influence or responsiveness within the group. Below, I introduce some action- and group-aligned summaries that remain applicable with dyads.

2.3.2 Action-Aligned Measures

Next, consider the measures in the right-hand column of Figure 3. Whereas person-aligned measures were defined with individual group members as the focal point, action-aligned summaries turn the focus to types of actions, regardless of who performed them. These measures are useful for describing the degree to which specific types of actions contribute to the group's overall interpersonal dynamics.

Recall that features denoting actions of type i are collected in the vector $\mathbf{u}_i = [V_{i1}, V_{i2}, \ldots, V_{iJ}]$. First I consider the middle left and lower left panels of Figure 3, which are directly analogous to the person-level measures defined above.

The action-to-person measures are

$$D^+(\mathbf{u}_i^*, \mathbf{v}_j^*) = \frac{1}{I(J-1)} \sum_{m \in \mathcal{J}_i} \sum_{n \in \mathcal{J}_j} \alpha_{mn}^* \quad \text{and} \quad D^-(\mathbf{u}_i^*, \mathbf{v}_j^*) = \frac{1}{I(J-1)} \sum_{m \in \mathcal{J}_j} \sum_{n \in \mathcal{J}_i} \alpha_{mn}^*, \tag{6}$$

with $\mathcal{J}_i = \{(j-1)I + i \mid V_{ij} \in \mathbf{u}_i\}$. Action-to-person out-degree describes how a type of action (e.g., sending chat messages) influences the future actions of a given group member. The action-to-person in-degree describes the extent to which the actions of a given member elicit a certain type of action in response from the other members of the group.

As illustrated in the lower right panel of Figure 3, action-to-group degree extends these measures to address how a specific type of action influences, or occurs in response to, the actions of the entire group:

$$D^+(\mathbf{u}_i^*, \mathbf{v}^*) = \frac{1}{J} \sum_{j=1}^{J} D^+(\mathbf{u}_i^*, \mathbf{v}_j^*) \quad \text{and} \quad D^-(\mathbf{u}_i^*, \mathbf{v}^*) = \frac{1}{J} \sum_{j=1}^{J} D^-(\mathbf{u}_i^*, \mathbf{v}_j^*). \tag{7}$$

Note that when only one action type is considered, the action-aligned measures are equivalent to corresponding person-aligned measures and consequently suffer from the ambiguity described in connection with Equation (4). Thus, a second specific recommendation for collaboration analytics is that research involving dyads should focus on recovering multiple process features from each member.

Lastly, consider a measure that is specific to a type of action, which I will refer to as an action channel. This idea is depicted in the top right panel of Figure 3 and can be written as

$$D(\mathbf{u}_i^*, \mathbf{u}_i^*) = \frac{1}{J(J-1)} \sum_{m \in \mathcal{J}_i} \sum_{n \in \mathcal{J}_i} \alpha_{mn}^*. \tag{8}$$

An action channel describes the extent to which a group's interpersonal dynamics are conveyed via a specific type of action. It contains the subset of edges from the action-to-group measures that are bi-directional, and therefore can be thought of as a pruned version of action-to-group measures. Also note that the degree-sum formula for directed graphs, which states that the sum over nodes of out-degree and in-degree are both equal to the sum total of the edges, implies that action channels do not have a valence. Rather, an action channel summarizes the role of a specific type of action in conveying both influence and responsiveness over the entire group.

---

[1]This equivalence arises whenever the valence of the degree is switched *and* the focal and reference features are switched. However, of the measures proposed in this paper, person-to-person degree is the only case of this duality that does not also change the type of focal feature. For brevity, I do not discuss it more generally.

2.3.3 Group-Aligned Measures

The foregoing has presented some ways of summarizing interpersonal dynamics in terms of group members or the types of actions they perform. In many research settings, this will be more informative than a single overall summary of a group's interpersonal dynamics. Nonetheless, group-level extensions of the person-aligned and action-aligned measures show how they are related and help address the previously noted issues that arise when working with dyads.

Table 1. Four equivalent ways of analyzing group dynamical interdependence

	Influence (out-degree)	Responsiveness (in-degree)
Persons Eqs. (3) & (5)	The influence of each group member on the group's future actions	The responsiveness of each group member to the group's past actions
Actions Eqs. (6) & (7)	The degree to which a specific type of action influences the group's future actions	The degree to which the group responds to past actions with a specific type of action

Table 1 summarizes the relationships among the person-aligned and action-aligned measures depicted in the bottom two rows of Figure 3. As mentioned above, the degree-sum formula implies that we may analyze the overall interpersonal dynamics of a group either in terms of influence or in terms of responsiveness. Similarly, aggregating person-to-group measures over focal persons and feature-to-group measures over focal features yields the sum total of the edges. Thus, there are four ways of breaking down the overall interpersonal dynamics of the group, as described in Table 1 and shown in the following equation:

$$D(\mathbf{v}^*, \mathbf{v}^*) = \frac{1}{J} \sum_{j=1}^{J} D^+(\mathbf{v}_j^*, \mathbf{v}^*) = \frac{1}{J} \sum_{j=1}^{J} D^-(\mathbf{v}_j^*, \mathbf{v}^*) = \frac{1}{I} \sum_{i=1}^{I} D^+(\mathbf{u}_i^*, \mathbf{v}^*) = \frac{1}{I} \sum_{i=1}^{I} D^-(\mathbf{u}_i^*, \mathbf{v}^*). \tag{9}$$

In particular, the first two equalities in Equation (9) lead to the following dyad-aligned measure, which is unvalenced and therefore avoids the ambiguities discussed in defining person-to-person measures of influence and responsiveness:

$$D_2 = \frac{1}{2}(D^+(\mathbf{v}_j^*, \mathbf{v}_k^*) + D^+(\mathbf{v}_k^*, \mathbf{v}_j^*)) = \frac{1}{2}(D^-(\mathbf{v}_j^*, \mathbf{v}_k^*) + D^-(\mathbf{v}_k^*, \mathbf{v}_j^*)). \tag{10}$$

This measure summarizes the degree to which two group members mutually influence one another or, equivalently, the degree to which they are mutually responsive to one another.

The interpretation of Equation (9) can be further facilitated by considering its relation to the overall temporal dependence among a groups' actions (i.e., Figure 1). The decomposition in Equation (1) suggests various ways to accomplish this. For example, the overall proportion of temporal dependence among a group's actions that is attributable to interpersonal sources can be defined as the weighted density of the interpersonal dynamics matrix $\mathbf{A}^*$:

$$G = \frac{I(J-1)D(\mathbf{v}^*, \mathbf{v}^*)}{ND(\mathbf{v}, \mathbf{v})}. \tag{11}$$

Rather than a proportion-of-total formulation, we might instead consider a signal-to-noise ratio, $G' = G/(1-G)$, where interpersonal dynamics are the signal and intra-personal dynamics are the noise.

While the interpretation of the group-level measures G and G' is facilitated by their similarity to other widely used statistical summaries, they each involve ratios of dependent random variables, which can be difficult to work with. However, in some dynamical models, the sum of the edges in $\mathbf{A}$ is a fixed quantity. In particular, this is the case when the matrix $\mathbf{A}$ represents conditional probabilities, which means that a sum-to-one constraint is applied to the rows (or columns) of $\mathbf{A}$ and so $D(\mathbf{v}, \mathbf{v}) = 1/N$. This yields the simplified expression

$$G = I(J-1)D(\mathbf{v}^*, \mathbf{v}^*). \tag{12}$$

2.4 Summary

The foregoing provides a relatively general blueprint for building network-based measures that can be used to describe the interpersonal dynamics of small groups, as well as several examples of such measures. Sections 2.1 and 2.2 sketched out the groundwork in a way that is applicable to methodological approaches in which group dynamics can be represented as a network or graph. Section 2.3 showed how some commonly used network summaries can be extended to characterize the contributions of specific persons or specific types of actions to the interpersonal dynamics of a group, and to describe the overall proportion of group dynamics that is attributable to interpersonal sources. The remainder of the paper addresses some ways to evaluate the quality of the proposed measures or other summaries derived using this blueprint.

3. Reliability

In the psychometric literature, reliability describes the extent to which a statistic (e.g., a test score) varies systematically over individuals (Joint Committee for Educational and Psychological Testing, 2014). In this section I briefly discuss why reliability is important in the context of collaboration analytics and show how it can be computed for the measures introduced above.

3.1 Why Is Reliability Important?

If process-based measures do not systematically vary over groups or their members, then, by definition, those measures will not be predictive of differences among groups or their members. In the context of purely predictive modelling with massive feature sets and sample sizes, this may not matter much. For example, if the goal is to automatically extract features from high-throughput sensors and "see what sticks" when predicting learning outcomes, unreliable features will be culled during variable selection, because they are, by definition, not predictive.

However, if the goal is to interpret, understand, or intervene in group processes, the reliability of the features describing those processes takes on a much more significant role. In such cases, features are often engineered to provide information about specific behavioural or cognitive processes hypothesized to be important for collaboration. If such features are not predictive of relevant group outcomes, we can distinguish two cases:

1. the behavioural or cognitive processes are not related to the outcome of interest, or
2. the process-based features were not sufficiently reliable to learn the relationship.

These two options can be likened to true negatives and false negatives, respectively. If we do not know the reliability of the features used to describe group processes, we cannot know the correct interpretation of a negative outcome. In short, we can't know whether the problem was the theory or the data.

In addition to providing an overall gauge of measurement quality, reliability can also be treated as a function of task or group characteristics. For example, a standard practice in conventional psychometric settings is to treat the reliability of a test score as a function of test length. This tells us how long a test needs to be in order to reliably estimate the trait being assessed. Applying this rationale to group processes, it may be useful to consider the reliability of a measure as a function of the total number of feature-events used to compute that measure. This would provide insight about the overall "volume" of group interactions required to reliably estimate quantities of interest, which can in turn be used to inform the design of group activities. This basic idea of using reliability to inform design underlies much of psychometric research (e.g., Lord, 1980).

Treating reliability as a function of group characteristics, rather than task characteristics, can provide information about equity and fairness. For example, if the reliability of a measure differs systematically by the gender or racial composition of collaborative groups, this would be described in terms of measurement and/or of predictive bias (Joint Committee for Educational and Psychological Testing, 2014). Similar considerations about equity and fairness have recently made their way into the literature on AI and machine learning (Mehrabi, Morstatter, Saxena, Lerman, & Galstyan, 2019), and an examination of the reliability of measures used for predictive modelling provides a way to begin addressing those issues in research settings.

3.2 How to Compute the Reliability of the Proposed Measures

Above it was assumed that the adjacency matrix $\mathbf{A} = \{\alpha_{mn}\}$ is defined with reference to a statistical model suitable for analyzing group process data (e.g., a state-space model, vector auto-regression, point process). Here I additionally assume that the point estimates, $\hat{\alpha}_{mn}$, and their covariance matrix are available for each group. The quantities introduced in Section 2.3 can be estimated by using $\hat{\alpha}_{mn}$ in place of α_{mn}. The reliability of the measures can then be computed from the covariance matrix of the $\hat{\alpha}_{mn}$ using the following two steps:

Step 1: Compute the standard error of the measure. Except for Equation (11), the degree-based measures introduced in Section 2 are computed as sums of the α_{mn}, and so their standard errors can be obtained from the variance sum law. For example, let $\hat{D}_g$ denote the estimate of a given degree measure in group $g = 1, \dots, G$. The squared standard error of $\hat{D}_g$ is

$$[\text{SE}(\hat{D}_g)]^2 = \sum_{m,n} [\text{SE}(\hat{\alpha}_{mn})]^2 + \sum_{m,n} \sum_{r,s} \text{COV}(\hat{\alpha}_{mn}, \hat{\alpha}_{rs}), \tag{13}$$

where SE denotes standard error, COV denotes covariance, and the range of the sums is the same as for $\hat{D}_g$. This same approach is applicable to the proportion-of-dependence coefficient G given in Equation (12). To obtain the standard error of quantities defined as ratios of degree-based measures, such as Equation (11) or other more complicated measures, one can make use of the multivariate delta method (e.g., Oehlert, 1992) or bootstrapping approaches (e.g., Hastie, Tibshirani, & Friedman, 2009).

Step 2: Average the squared standard errors over groups. This provides an estimate of the overall measurement error variance, which can then be compared to the total variance to obtain various summaries of the measure's reliability. In particular, the

reliability coefficient may be computed as (see, e.g., Thissen & Wainer, 2001)

$$\rho(\hat{D}) = \frac{\text{VAR}(\hat{D}) - \text{MEAN}([\text{SE}(\hat{D}_g)]^2)}{\text{VAR}(\hat{D})}, \tag{14}$$

in which the mean and variance are taken over groups. In practice, the covariance matrix of the $\hat{\alpha}_{mn}$ and the means and variances taken over groups are replaced with their sample estimates. The use of Equations (13) and (14) is illustrated in Section 5.

4. Validity

Psychometric research on validity includes a wide range of topics about the use and interpretation of measures (Joint Committee for Educational and Psychological Testing, 2014). In this section I focus on a relatively narrow question, but one that I think demands special attention when analyzing group process data: What is the added value of process-based features when predicting group outcomes? In this section it is argued that, in order to justify their additional complexity, process-based features should contribute above and beyond (1) individual traits that can be measured without consideration of group membership (i.e., group composition) and (2) static summaries of group processes that "code and count" features without describing their sequential or temporal relationships. Otherwise stated, predictive models that are used to evaluate the importance of process-based features should also include well-established non-process-based features that offer simpler explanations of group performance.

4.1 Group Processes versus Group Composition

There is a substantial body of research addressing how the performance of teams can be explained in terms of the attributes of their members (e.g., Laughlin, 2013; Mathieu, Tannenbaum, Donsbach, & Alliger, 2014; Woolley, Chabris, Pentland, Hashmi, & Malone, 2010). In particular, Lorge and Solomon (1955) proposed a model, simply referred to as Model A, to explain the performance of any group solely in terms of the performance of its individual members working in isolation. In the heyday of social psychological research on small groups, Model A was often regarded as the default explanation of group performance, against which any new proposal would need to justify its additional complexity (e.g., Davis, 1992).

This same rationale seems applicable to theories and measures that emphasize the importance of small-group collaborations. If group processes cannot tell us more about relevant outcomes than do the pre-existing traits of the individual group members, then it seems hard to justify the ontological baggage incurred by theorizing the processes. More pragmatically, if the statistical association between processes and outcomes is due to underlying aspects of group composition, it would be incorrect to infer a causal relationship between the processes and outcomes.

A very wide range of student outcomes have been addressed by research on small groups in educational settings (e.g., Johnson & Johnson, 2009; S. Chen, Lawrence, Zhou, Min, & Snow, 2018). However, one outcome that is of central interest in most research settings is student learning, measured either proximally via a group member's performance on a given task or more distally via group-to-individual transfer (e.g., individual post-testing). It is well known that prior academic performance is a potent predictor of future performance. Consequently, when student learning is considered a relevant outcome, the added value of process-based measures should control for group members' knowledge and skills in the content domain, as measured prior to group interactions.

4.2 Group Processes versus Code and Count

A similar argument applies to static summaries of group interactions, which have been described as code-and-count approaches (Rack, Zahn, & Mateescu, 2018). In this situation, a rubric is applied to categorize the different features that arise from a given data stream, and then person- or group-level summaries are produced by counting the number of occurrences of the different codes.

There are numerous theory-based rubrics for coding chat or other aspects of small-group interactions (e.g., Adams et al., 2015; Andrews-Todd & Forsyth, 2020; Liu, Hao, von Davier, Kyllonen, & Zapata-Rivera, 2015; Hao, Chen, Flor, Liu, & von Davier, 2017). In many cases, though certainly not all, these codes are based on a single turn in a sequence of actions, i.e., they are defined in reference to one action performed by one group member. There is nothing about one-turn codes that is inherently dynamic. Consequently, I argue that merely describing the frequency distributions of such codes is a static, rather than dynamic, summary of group interactions. Of course, such codes may be analyzed for their temporal dependence, but this is not accomplished simply by counting them.

In situations involving text-based interactions among group members, there is a particularly strong precedent for considering raw word count as one such static summary. Word count is known to be a robust predictor of automated essay scores, even when the scoring procedures are not themselves based on word count (e.g., Attali & Burstein, 2014). This has led to the criticism that automated scoring is not actually measuring essay quality at all. An analogous argument holds for text-based features of group

processes —- dynamical interdependence is intended to describe the nature of a group's interactions, not merely whether group members engaged in dialogue. Consequently, process-based measures of small-group collaborations should tell us something beyond static summaries of participation, as measured by word count or similar indices.

5. Empirical Example

This section illustrates the proposed approach to measuring dynamical interdependence with an example in which 104 dyads interacted over online chat to answer questions from a grade 12 level mathematics assessment. I focus on three types of actions for each member of a dyad: (1) responding to test questions and sending chat messages (2) with and (3) without math content. I use a standard Markov chain to model temporal dependence in the sequence of events for each dyad and apply the network statistics introduced in Section 2 to the resulting transition probabilities of each group. As discussed above, person-aligned measures are not uniquely defined for dyads, so this illustration is limited to the feature- and group-aligned measures.

The analyses focus on

(a) a description of the action-aligned measures, including their reliability;

(b) the relationship of the action-aligned measures to one another, to group performance on the math assessment, to individual math ability measured via pre-test, and to the total word count of chat messages sent by each dyad; and

(c) a multiple regression (i.e., a linear, additive, supervised learning) model that isolates the contribution of the process-based measures to group performance, after controlling for group members' ability and the total word count.

The regression model addresses the points made in Section 4 about predictive validity. This section presents the results of the analyses, with the overall interpretation deferred to the Discussion section.

5.1 Sample and Procedure

The sample and data collection procedures are described in detail by Mohan, Bergner, and Halpin (2020) and briefly summarized here. The sampling frame for the study comprised Amazon Mechanical Turk (AMT) workers who self-reported to live in the United States and to speak English as their first language. The median age was 32 years, with an interquartile range of [27, 40]. The majority of the sampling frame (71%) self-identified as "white," 51% as female, and 88% as having at least one year of post-secondary education. The data for this study were obtained as a convenience sample from the sampling frame. However, due to a recording error, demographic characteristics of the realized sample were not available for analysis.

The assessment content was adapted from publicly released grade 12 level math items used by the National Assessment of Educational Progress (NAEP)[2]. Some adaptations of the NAEP items were made to encourage more productive interactions among collaborators. These included changing the response format from multiple choice to numeric response and modifying some items based on templates for dyadic assessments used by Griffin and Care (2015).

A total of 270 respondents were assessed under both individual and group testing conditions. The test forms administered in the two different conditions were counterbalanced and were equated for difficulty using item response theory. The individual testing condition was used to estimate respondents' mathematical ability. After completing the individual assessment, respondents were routed to a second form in which they were randomly paired with an anonymous partner and encouraged to work with their partner to ensure that both individuals arrived at the correct response. In the present analysis, I excluded 31 dyads in which either of the partners did not complete the study. This was operationalized as an individual not responding to at least one third of the test questions. In many cases, this was due to poor Internet connectivity, but in some cases it was apparent from the chat logs that one or both group members did not want to continue. The resulting analytical sample consisted of 104 dyads.

5.2 Measures
5.2.1 Process-Based Features
The time-stamped user log data for each dyad were coded to represent the sequence of each of three actions performed by the two members. These three actions were chosen to be low-inference aspects of group processes that can serve to illustrate the measures introduced in Section 2.3. For discussion of a more extensive set of process-based features, please see Mohan and colleagues (2020).

The actions encoded were as follows, with variable codes indicated in parentheses:

- *Responding to a test question* (`prob`): The correctness of the response was not considered, only whether a response was submitted.

---

[2]https://nces.ed.gov/nationsreportcard/assessments/frameworks.aspx

- *Sending a chat message with math content* (`math`): Messages were parsed for digits and the symbols $+, -, *, /, <, >, \hat{}, =, \%$. If such content was present, the message was coded as a math chat.

- *Sending a chat message without math content* (`chat`): These messages were coded separately.

Note that each of these actions was automatically coded and did not involve qualitative judgments of human raters. It should also be noted that while math content of chat was predictive of performance on this task, simply discussing the content of the questions was not (Mohan et al., 2020). Qualitative readings of the chat transcripts suggested that problem content was associated with discussing the problem, whereas math content was associated with discussing a solution to a problem.

5.2.2 Measures of Dynamical Interdependence

For each dyad, the transition probabilities for the sequence of process-based features were estimated using the `markovchain` package in `R` (Spedicato, 2017). Using the notation of Section 2.1, the three features recorded for each group member correspond to the nodes $\mathbf{v}_j$, and the matrix of transition probabilities corresponds to the adjacency matrix $\mathbf{A}$.

As mentioned above, person-aligned measures are not well defined for dyads. Consequently, the focus of the example was on the set of measures described below:

- *Action-to-group influence (+) measures.* These measures indicate the degree to which the performance of a specific type of action by either partner influenced the probability of the other partner's future actions. For example, if I send a math-based chat message to my partner, does this elicit a response of any kind — another math-based chat message, a non-math chat message, or a problem response? Similarly, if my partner sends a math-based chat, how does this influence my further actions? The math-to-group influence measure aggregates these two sources of influence — my math-based chatting and that of my partner — on our future actions.

- *Action-to-group responsive (−) measures.* These measures indicate the degree to which partners tend to respond to one another's actions in a certain way. For example, perhaps my partner and I mainly interact to share answers to the test questions, in which case we might tend to respond to one another mainly by filling in our problems. This would correspond to a high degree of problem-to-group responsiveness.

- *Action channels.* Unlike the action-to-group measures, which consider the influence or responsiveness of a certain type of action on any other action, action channels consider the influence or responsiveness of a type of action only on actions of the *same* type. If my partner answers a test question, how does this influence the probability that I will answer a test question, and vice versa? If my partner sends a math-based chat message, how does this influence the probability that I will send a math-based chat message, and vice versa?

- *Overall proportion of group dynamics that are due to inter-individual sources (G-coefficient).* This is a group-level measure that describes how likely it is that any action by either partner leads to an action of the other partner, as opposed to another action by the same person.

Because a Markov chain was used to model group dynamics, the action-aligned measures are all interpretable as averages of transition probabilities, whereas the group-aligned measure is scaled as a proportion of total dependence (i.e., the sum of all the transition probabilities).

5.2.3 Group Outcomes

The relationship of the proposed measures with group performance on the math test was considered as a source of evidence about the validity of the measures. Each item on the group test was scored as correct if and only if both members of a dyad submitted the correct response to the item. Each group's overall performance on the test was summarized as the proportion of test items that were responded to correctly. This scoring procedure is compatible with the instructions on the group test, and Halpin and Bergner (2018) provide a more detailed discussion of group-level scoring procedures.

5.2.4 Control Variables

As discussed in Section 4, two control variables were considered in this study. The first was group members' individual performance on the math pre-test. Individual performance was quantified using the same conjunctive scoring procedure as for the group-administered assessment, but now applied to the individually administered pre-test. Conceptually, this represents the proportion of test items that the two group members would have both answered correctly, even without working together. Because the group and individual test forms were equated on difficulty, the conjunctive score on the individual test can be viewed as a proxy for the expected performance on the group test, under the assumption that the group members did not work together (Halpin & Bergner, 2018).

The second control variable was the total word count of chat messages sent between partners. Because word count was moderately skewed, it was log-transformed prior to analysis. Many related measures could also have been used (e.g., number of

messages sent, number of words per message, differences between group members' word counts), but these were either highly correlated with word count or not correlated with group performance on the math test.

5.3 Results

Univariate summary information about the interdependence measures is presented in Figure 4 and Table 2. Starting with the top panel of Figure 4, it can be seen that the measures were roughly symmetrical in distribution and did not exhibit problematic characteristics such as floor or ceiling effects. The three distributions with the least variance (highest peaks) were the action-to-group influence (+) measures. Each of these three measures also had a similar mean in this sample, whereas the locations of the responsive measures (−) and action channels were more spread out.
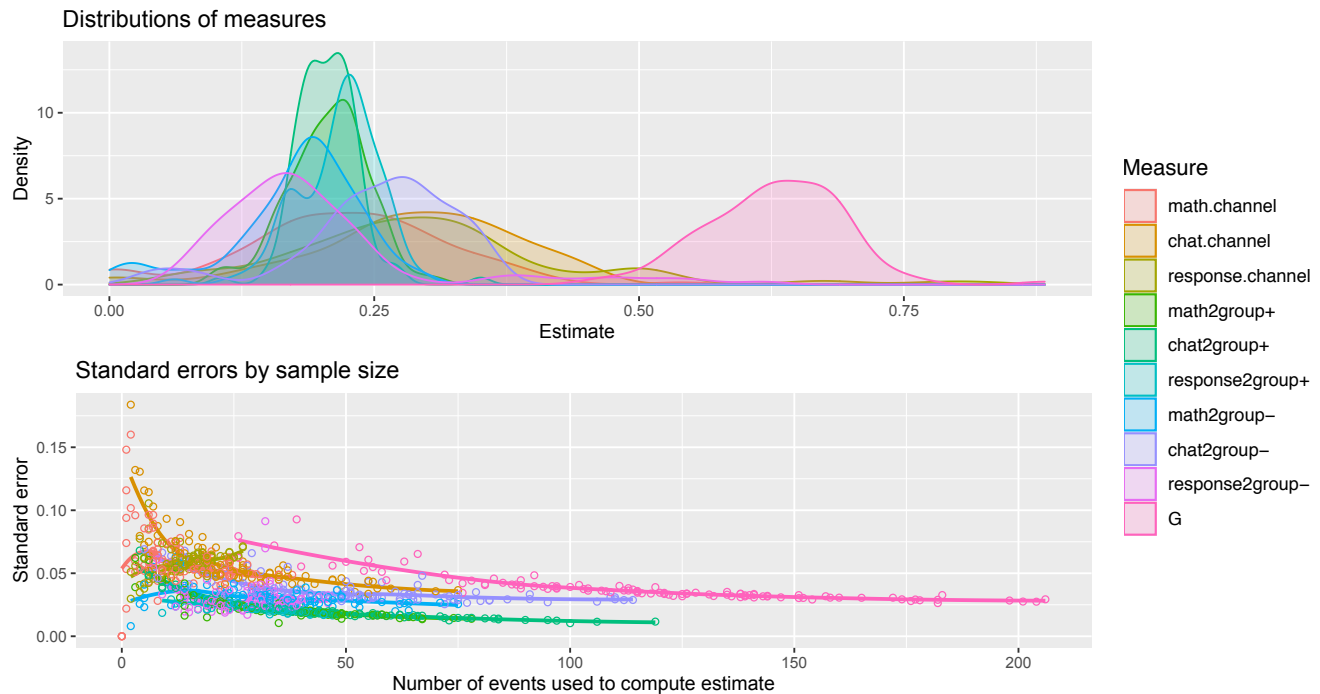
Figure 4. Summaries of interdependence measures. All measures other than G are averages of transition probabilities; G is interpreted as a proportion of the sum total of the transition probabilities.

The bottom panel of Figure 4 shows that the standard errors of the measures were decreasing in the number of observations (feature-events) used to compute each measure. For most measures, the standard errors began to plateau at around 50 events. For the G-coefficient, the standard errors did not level out until about 150 events. These numbers provide some practical guidance about required sample sizes when analyzing group interactions with a Markov chain, but more complicated models are likely to require more data.

Table 2. Estimated reliability coefficients of interdependence measures

Measure	Reliability
chat.channel	0.602
math.channel	0.689
prob.channel	0.717
chat2group+	0.268
math2group+	0.590
prob2group+	0.706
chat2group−	0.745
math2group−	0.726
prob2group−	0.890
G	0.683

All but one of the estimated reliability coefficients in this sample ranged from 0.59 to 0.89. This means that between 59% and 89% of the variation in the measures was systemically associated with differences among groups. The lowest reliability was associated with the influence of (non-math-based) chat on groups' later actions. Although the standard errors of this measure were quite low, it was also the measure with the least variability over groups, as shown in the top panel of Figure 4. The low reliability coefficient is perhaps due to the limited variability of this measure in the present sample. Aside from chat-to-group influence, the other measures generally had reasonable reliability coefficients. The overall conclusion is that proposed measures had sufficient precision to meaningfully consider their relationships with other variables.

Figure 5 shows the zero-order correlations among the interdependence measures and the other variables considered in this study. The following observations can be made:

- Some of the dynamical interdependence measures were quite strongly correlated with group performance. Notably, the response-to-group influence ($r = 0.46$), math-chat-to-group responsiveness ($r = 0.45$), and G-coefficient ($r = 0.48$) all had correlations with group performance that were on par with the control variables (individual performance and number of words).

- The measures that were correlated with group performance also tended to be correlated with the control variables. This indicates that the control variables are potential confounders of the relationship between group performance and the measures of dynamical interdependence.

- The action-to-group responsiveness measures were all strongly correlated with the channels of the same action, indicating that these measures are providing somewhat redundant information. The action-to-group influence measures showed a similar pattern of association with the action channels, but the correlations were not as strong. The influence and responsiveness measures for the same action were not strongly correlated. In general, it appears that the different process-based measures (influence, responsiveness, and channels) are providing related but distinct information about group interactions.
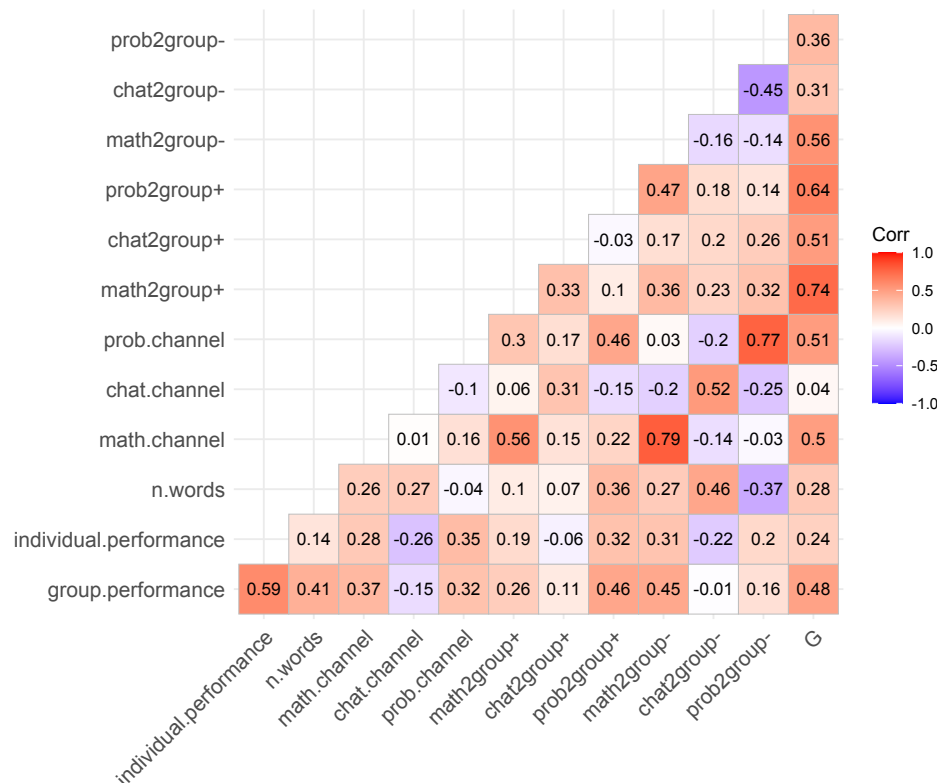

Figure 5. Pearson correlations among study variables.

Finally, Table 3 shows the results of the linear regression analyses. The two control variables together accounted for about 46% of the variation in group performance (Model 2), indicating their importance as statistical controls. Together, all of the interdependence measures collectively added about 10% of explained variance in group performance (Model 8). When the controls were not included in the model (not shown in the table), the proportion of variation associated with the interdependence measures was 42.3% ($F(8,95) = 8.717, p < 0.001$), which, as argued above, would be a drastic overestimate of their added

value in predicting group performance. Except for the chat-to-group features (Model 6), each of the models that included interdependence measures contributed a statistically significant improvement in prediction beyond the model with just the controls. However, of the interdependence measures, the G-coefficient was the most powerful predictor of group performance, and the contributions of the other measures beyond the G-coefficient were not significant ($F(7, 93) = 0.483, p = 0.846$).

Table 3. Summary of regression analyses

	Dependent variable: group performance							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Individual Performance	0.837*** (0.113)	0.771*** (0.104)	0.673*** (0.111)	0.634*** (0.106)	0.814*** (0.106)	0.655*** (0.105)	0.675*** (0.099)	0.592*** (0.117)
n.words		0.175*** (0.031)	0.171*** (0.039)	0.127*** (0.035)	0.173*** (0.042)	0.151*** (0.041)	0.135*** (0.031)	0.115** (0.045)
math.channel			0.504** (0.198)					0.202 (0.381)
chat.channel			0.135 (0.228)					−0.123 (0.317)
prob.channel			0.197 (0.172)					0.192 (0.386)
math2group+				0.834* (0.496)				0.298 (0.717)
math2group−				0.890** (0.344)				0.507 (0.819)
chat2group+					0.956 (0.711)			0.583 (0.879)
chat2group−					0.421 (0.276)			0.475 (0.555)
prob2group+						1.660*** (0.478)		1.001 (0.929)
prob2group−						−0.094 (0.198)		–
G							1.174*** (0.275)	–
R^2	0.349	0.462	0.510	0.527	0.482	0.521	0.545	0.568
ΔR^2	–	0.113	0.048	0.065	0.020	0.059	0.083	0.106
ΔF	54.771***	22.583***	3.172***	6.827***	1.902	6.036***	18.277***	2.854***

Note: Table reports regression coefficients with their standard errors in brackets. To compute ΔR^2, Model 2 is compared to Model 1 and Models 3 to 8 are compared to Model 2. Model 8 omits the last two predictors due to multicollinearity. $^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

6. Discussion

This paper has discussed how to measure dynamical interdependence among group members. I proposed a general blueprint for developing such measures that is applicable to methodological approaches in which group dynamics can be represented as a

network or graph. The blueprint was used to motivate several measures that characterize the contributions of specific persons or specific types of actions to the interpersonal dynamics of a group, as well as group-level measures summarizing the total proportion of a group's dynamics that are due to inter-individual, rather than intra-individual, sources. These considerations also led to the conclusion that it is not possible to unambiguously define person-level measures for dyads, and therefore that the study of groups with three or more members is important for understanding the contributions of individual members to group dynamics. The interpretation, reliability, and validity of some feature- and group-aligned measures were illustrated using an empirical example in which dyads worked together over online chat to complete a grade 12 level mathematics assessment. It was found that a sample size of approximately 50 feature-events was sufficient for the standard errors of the action-aligned measures to reach a lower plateau; however, a sample size closer to 150 events was required for the group-level measure. In the current sample, all of the measures had reasonably good marginal reliability, except for the (non-math) chat-to-group responsiveness, which had limited variability. In general, it is important to design group activities so that the volume of interactions per group can support reliable estimation of the measures of interest. Due to limitations of the sample data, it was not possible to examine reliability as a function of group composition. However, as mentioned in Section 3, this can provide an important avenue for researching equity and fairness in the analysis of group processes.

The different interdependence measures provided related but distinct information about group interactions. Additionally, some of the proposed interdependence measures were quite strongly correlated with group performance on the math assessment. The degree to which group members had continued exchanges of chat messages with math content was moderately correlated with group performance. In fact, responding to math-chat with any action —- more math-chat, non-math-chat, or problem responses — was also a moderately strong predictor of group performance. These two measures (math channel and math-to-group responsiveness) both suggest the importance of partners' building off each other's input about how to solve the math questions. Somewhat surprisingly, problem-to-group influence was also quite strongly correlated with group performance. This measure describes the degree to which one person submitting a problem response was followed by an action of any kind by their partner. Selective reading of some transcripts with low and high values of problem-to-group influence provided some insight here. In particular, low values often resulted when one person worked on the questions without their partner, resulting in a sequence of problem responses from the same individual, sometimes interspersed with chat messages informing their partner of the submitted responses. Thus, low values of problem-to-group responsiveness indicated that one partner was working ahead on the assessment without waiting for input from their partner. Lastly, the *G*-coefficient was also a strong predictor of group performance. This measure describes the overall proportion of a group's dynamics that is associated with inter-individual, as opposed to intra-individual, sources.

Altogether, the proposed measures of dynamical interdependence accounted for about 42% of the variation in group performance. However, stronger evidence about the predictive validity of the proposed measures can be provided by controlling for non-process-based features that are known predictors of group performance. Two such features were included in the example. Individual performance on a math pre-test was used to control for group composition, and total word count was used as a static code-and-count process feature. These two variables together explained about the same amount of variation in group performance as the proposed measures, suggesting their importance as control variables. Together, the proposed interdependence measures added predictive value to the control variables, explaining an additional 10% of variation in group performance, with the full model explaining about 57% of the variation in group performance. Most of this added value could be attributed to the *G*-coefficient, which suggests its usefulness as a summary measure of dynamical interdependence. More generally, it was argued that the predictive value of process-based features should be evaluated after consideration of (1) individual traits that can be measured without consideration of group membership (i.e., group composition) and (2) static summaries of group processes that code and count features without describing their sequential or temporal relationships.

In conclusion, this paper has introduced and illustrated a relatively generic approach to synthesizing low-level, time-intensive group process data in terms of the emergent dynamics among group members and the roles of specific individuals and specific types of actions therein. It is hoped that this work can provide a useful blueprint for building custom measures of dynamical interdependence, or other constructs related to the temporal dynamics of small-group processes, without losing sight of basic questions about the reliability, validity, and overall quality of those measures.

Acknowledgements

Funding

Declaration of Conflicting Interest

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Adams, R., Vista, A., Scoular, C., Awwal, N., Griffin, P., & Care, E. (2015). Automatic coding procedures for collaborative problem solving. In P. Griffin & E. Care (Eds.), *Assessment and Teaching of 21st Century Skills* (pp. 115–132). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-017-9395-7_6

Andrews-Todd, J., & Forsyth, C. M. (2020). Exploring social and cognitive dimensions of collaborative problem solving in an open online simulation-based task. *Computers in Human Behavior*, *104*, 105759. https://doi.org/10.1016/j.chb.2018.10.025

Attali, Y., & Burstein, J. (2014). Automated essay scoring with e-rater® V.2. *The Journal of Technology, Learning and Assessment*, *4*(2), i–21. https://doi.org/10.1002/j.2333-8504.2004.tb01972.x

Barrat, A., Barthelemy, M., & Vespignani, A. (2011). *Dynamical Processes on Complex Networks*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511791383

Butts, C. T. (2008). A relational event framework for social action. *Sociological Methodology*, *38*(1), 155–200. https://doi.org/10.1111/j.1467-9531.2008.00203.x

Chen, B., Knight, S., & Wise, A. F. (2018). Critical issues in designing and implementing temporal analytics. *Journal of Learning Analytics*, *5*(1), 1–9. https://doi.org/10.18608/jla.2018.53.1

Chen, S., Lawrence, J. F., Zhou, J., Min, L., & Snow, C. E. (2018). The efficacy of a school-based book-reading intervention on vocabulary development of young Uyghur children: A randomized controlled trial. *Early Childhood Research Quarterly*, *44*, 206–219. https://doi.org/10.1016/j.ecresq.2017.12.008

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Cohen, E. G. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research*, *64*(1), 1–35. https://doi.org/10.3102/00346543064001001

Daley, D. J., & Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes* (2nd ed.). New York: Springer. https://doi.org/10.1007/b97277

Davis, J. H. (1992). Some compelling intuitions about group consensus decisions, theoretical and empirical research, and interpersonal aggregation phenomena: Selected examples, 1950–1990. *Organizational Behavior and Human Decision Processes*, *52*(1), 3–38. https://doi.org/10.1016/0749-5978(92)90044-8

Dowell, N. M. M., Nixon, T. M., & Graesser, A. C. (2019). Group communication analysis: A computational linguistics approach for detecting sociocognitive roles in multiparty interactions. *Behavior Research Methods*, *51*(3), 1007–1041. https://doi.org/10.3758/s13428-018-1102-z

Durbin, J., & Koopman, S. J. (2012). *Time Series Analysis by State Space Methods*. Oxford, UK: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199641178.001.0001

Griffin, P., & Care, E. (2015). *Assessment and Teaching of 21st Century Skills: Methods and Approach*. New York: Springer. https://doi.org/10.1007/978-94-007-9395-7

Halpin, P. F., & Bergner, Y. (2018). Psychometric models of small group collaborations. *Psychometrika*, *83*, 941–962. https://doi.org/10.1007/s11336-018-9631-z

Halpin, P. F., von Davier, A. A., Hao, J., & Liu, L. (2017). Measuring student engagement during collaboration. *Journal of Educational Measurement*, *54*(1), 70–84. https://doi.org/10.1111/jedm.12133

Hanneman, R. A., & Riddle, M. (2014). Concepts and measures for basic network analysis. In *The Sage Handbook of Social Network Analysis* (pp. 340–369). London, UK: SAGE Publications Ltd. https://doi.org/10.4135/9781446294413.n24

Hao, J., Chen, L., Flor, M., Liu, L., & von Davier, A. A. (2017). CPS-rater: Automated sequential annotation for conversations in collaborative problem-solving activities. *ETS Research Report Series*, *2017*(1), 1–9. https://doi.org/10.1002/ets2.12184

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning* (2nd ed.). New York: Springer. https://doi.org/10.1007/978-0-387-84858-7

Johnson, D. W., & Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational Researcher*, *38*(5), 365–379. https://doi.org/10.3102/0013189X09339057

Joint Committee for Educational and Psychological Testing. (2014). *Standards for Educational and Psychological Testing*. Washington, DC: AERA APA, & NCME.

Jost, J. (2007). Dynamical networks. In J. Feng, J. Jost, & M. Qian (Eds.), *Networks: From Biology to Theory* (pp. 35–62). London, UK: Springer. https://doi.org/10.1007/978-1-84628-780-0_3

Knight, S., Friend Wise, A., & Chen, B. (2017). Time for change: Why learning analytics needs temporal analysis. *Journal of Learning Analytics*, *4*(3), 7–17. https://doi.org/10.18608/jla.2017.43.2

Kozlowski, S. W. J., & Chao, G. T. (2018). Unpacking team process dynamics and emergent phenomena: Challenges, conceptual advances, and innovative methods. *American Psychologist*, *73*(4), 576–592. https://doi.org/10.1037/amp0000245

Laughlin, P. R. (2013). *Group Problem Solving*. Princeton, NJ, USA: Princeton University Press. Retrieved from https://www.jstor.org/stable/j.ctt7rmfz

Leenders, R. T. A. J., Contractor, N. S., & DeChurch, L. A. (2016). Once upon a time: Understanding team processes as relational event networks. *Organizational Psychology Review*, *6*(1), 92–115. https://doi.org/10.1177/2041386615578312

Liu, L., Hao, J., von Davier, A. A., Kyllonen, P., & Zapata-Rivera, D. (2015). A tough nut to crack: Measuring collaborative problem solving. In Y. Rosen, S. Ferrara, & M. Mosharraf (Eds.), *Handbook of Research on Computational Tools for Real-World Skill Development* (pp. 344–359). Hershey, PA, USA: IGI-Global. https://doi.org/10.1136/bmj.330.7485.0-h

Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. New York: Routledge.

Lorge, I., & Solomon, H. (1955). Two models of group behavior in the solution of eureka-type problems. *Psychometrika*, *20*(2), 139–148. https://doi.org/10.1007/BF02288986

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Berlin, New York: Springer. https://doi.org/10.1007/978-3-540-27752-1

Madhyastha, T. M., Hamaker, E. L., & Gottman, J. M. (2011). Investigating spousal influence using moment-to-moment affect data from marital conflict. *Journal of Family Psychology*, *25*(2), 292–300. https://doi.org/10.1037/a0023028

Mathieu, J. E., Tannenbaum, S. I., Donsbach, J. S., & Alliger, G. M. (2014). A review and integration of team composition models: Moving toward a dynamic and temporal framework. *Journal of Management*, *40*(1), 130–160. https://doi.org/10.1177/0149206313503014

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2019). *A survey on bias and fairness in machine learning.* (arXiv:1908.09635)

Mohan, K., Bergner, Y., & Halpin, P. (2020). Predicting group performance using process data in a collaborative assessment. *Technology, Knowledge and Learning*, *25*, 367–388. https://doi.org/10.1007/s10758-020-09439-5

Molenaar, P. C. M. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika*, *50*(2), 181–202. https://doi.org/10.1007/BF02294246

Oehlert, G. W. (1992). A note on the delta method. *The American Statistician*, *46*(1), 27–29. https://doi.org/10.1080/00031305.1992.10475842

Oviatt, S., Schuller, B., Cohen, P., Sonntag, D., Potamianos, G., & Krüger, A. (2018). *The Handbook of Multimodal-Multisensor Interfaces: Signal Processing, Architectures, and Detection of Emotion and Cognition* (Vol. 2). Morgan & Claypool. https://doi.org/10.1145/3107990

Quera, V. (2018). Analysis of interaction sequences. In E. Brauner, M. Boos, & M. Kolbe (Eds.), *The Cambridge Handbook of Group Interaction Analysis* (1st ed., pp. 295–322). Cambridge University Press. https://doi.org/10.1017/9781316286302.016

Rack, O., Zahn, C., & Mateescu, M. (2018). Coding and counting: Frequency analysis for group interaction research. In E. Brauner, M. Boos, & M. Kolbe (Eds.), *The Cambridge Handbook of Group Interaction Analysis* (1st ed., pp. 277–294). Cambridge University Press. https://doi.org/10.1017/9781316286302.015

Spedicato, G. A. (2017). Discrete time Markov chains with R. *The R Journal*, *9*(2), 84. https://doi.org/10.32614/RJ-2017-036

Suthers, D. D., Dwyer, N., Medina, R., & Vatrapu, R. (2010). A framework for conceptualizing, representing, and analyzing distributed interaction. *International Journal of Computer-Supported Collaborative Learning*, *5*(1), 5–42. https://doi.org/10.1007/s11412-009-9081-9

Swiecki, Z., Ruis, A. R., Farrell, C., & Shaffer, D. W. (2019). Assessing individual contributions to collaborative problem solving: A network analysis approach. *Computers in Human Behavior*, *104*, 105876. https://doi.org/10.1016/j.chb.2019.01.009

Teasley, S. D. (1997). Talking about reasoning: How important is the peer in peer collaboration? In L. B. Resnick, R. Säljö, C. Pontecorvo, & B. Burge (Eds.), *Discourse, Tools and Reasoning: Essays on Situated Cognition* (pp. 361–384). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-662-03362-3_16

Thissen, D., & Wainer, H. (2001). *Test Scoring*. Mahwah, NJ, USA: Lawrence Earlbaum Associates.

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science*, *330*(6004), 686–688. https://doi.org/10.1126/science.1193147