On Convergence of Cognitive and Noncognitive Behavior in Collaborative Activity

Luna Bazaldua, Diego A.
Teachers College, Columbia University
dal2159@tc.columbia.edu

Khan, Saad
Educational Testing Services
skhan002@ets.org

von Davier, Alina A.
Educational Testing Services
avondavier@ets.org

Hao, Jiangang
Educational Testing Services
jhao@ets.org

Liu, Lei
Educational Testing Services
lliu@ets.org

Wang, Zuowei
University of Michigan
zwwang@umich.edu

ABSTRACT
We present results from a pilot study to investigate the evidence for convergence and synchrony in cognitive and noncognitive behavior of dyads engaged in a collaborative activity. Our approach utilizes multimodal data including video and participant action log files retrieved from the collaborative activity, an online educational simulation on science topics. The log files captured cognitive behavior including frequency and content of chat messages between dyads, as well system help requests. The video data recorded participant nonverbal behavior that was processed on a frame-by-frame basis using automated facial expression classifiers and coded by trained human raters on high-level noncognitive behaviors including: affect display gestures, engagement, anxiety and curiosity. The data were analyzed at individual and dyad levels and results using hierarchical clustering analysis demonstrate evidence of cognitive and noncognitive behavioral convergence among dyads.

Keywords
Collaborative Assessment, Human-Computer Interaction, Multimodal Data, Noncognitive states, Cluster Analysis

1. INTRODUCTION
Behavioral convergence refers to the unintentional imitation process of gestures, facial expressions, behaviors, moods, postures, or verbal patterns of coparticipants on a range of different time-scales [4, 12]. In literature it has been referred to by a variety of terms e.g., behavioral matching, mimicry, interpersonal coordination, entrainment, interactional synchrony and the Chameleon effect [4, 12, 17, 19]. While previous studies have explored its impact on interpersonal skills, coordinated activity, negotiations, and how individuals influence the behaviors of others [2, 4, 21], little research has focused on finding evidence for behavioral convergence in collaborative activity [24].

Collaboration is a complex activity that constitutes an interplay between *cognitive processes* such as knowledge acquisition, content understanding, action planning, and execution [7, 8, 10, 18, 26] and *noncognitive processes* such as social regulation, adaptability, engagement and social affect, such as boredom, confusion, and frustration [1, 3, 6]. Collaborative activity may take place in face-to-face interactions or through the medium of online distance learning technologies and collaboration platforms [20]. In either context collaboration is more effective when participants are engaged in the task and exhibit behaviors that facilitate interaction [25].

Our hypothesis is that behavioral convergence occurs during collaborative activity and it manifests in both cognitive and noncognitive processes. Based on this premise, we expect that people will tend to synchronize their behaviors (consciously or nonconsciously) while they are engaged in a collaborative activity. To test our hypothesis, a pilot study was conducted involving 12 unique dyads collaborating in an online game-like science assessment: ETS' online collaborative research environment ─the Tetralogue [15, 27]. Multimodal data including video and activity log files of each participating dyad were captured. The log files contain cognitive behavior including frequency and content of chat messages between dyads, as well as system help request (i.e., the participant requests to view educational videos on the subject matter to better answer assessment questions). The video data, on the other hand, recorded participant nonverbal behavior which was analyzed on a frame-by-frame basis using automated facial expression classifiers and annotated by trained human raters on high-level noncognitive behaviors including: affect display gestures, engagement, anxiety, and curiosity. Along with recent studies [17, 20, 24], in this paper we describe one of the first attempts to capture and analyze multimodal data in the context of studying behavioral convergence in collaborative activities.

2. Methodology
2.1 Collaborative Activity Platform
As mentioned earlier, our study used an online collaboration assessment platform: ETS' online collaborative research environment –the Tetralogue. This platform includes a set of multiple-choice items on general science topics, a simulation based assessment, a personality test, and a set of background questionnaires. The simulation task is on geology topics. The simulation-based task was developed as a task for individual test takers who will interact with two avatars and as a collaborative task that requires the collaboration among two human participants and two avatars in order to solve geology problems.

The participants, who may be in different locations, interact through an online chat box and system help requests (selecting to view educational videos on the subject matter). The main avatar, Dr. Garcia, introduces information on volcanoes, facilitates the simulation, and requires the participants to answer a set of individual and group questions and tasks. A second avatar, Art, takes the role of another student, in order to contrast his information with that produced by the dyad.

The system logs activity data of the participants in structured XML files, which capture participant actions including: identification of the user who performed each action, the number of chat messages, the content of those chat messages, the number of times the participants request additional information on subject matter from the system, the answer selected for each individual and group question, and the time at which each action occurred.

While the dyads interacted with the task, we captured the video of each individual participant. The video data were used for both annotating noncognitive behavior of the participants and automated facial expression analysis (see section 2.3 for further details). It should be noted that the only form of direct communication between the dyads was through the Tetralogue text-based chat interface and the dyads were not able to see or hear each other. Figure 1 illustrates the collaborative activity and data capture while participants interact in the system.

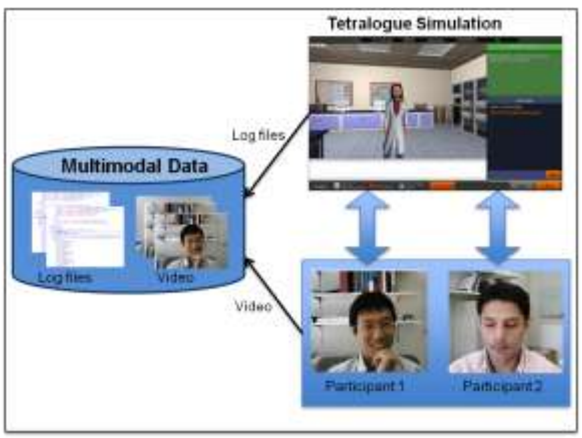

Figure 1. Multimodal data capture including video and action log files while participants engage in collaborative activity on the Tetralogue platform.

2.2 Study Participants and Data Collection

Twenty-four subjects participated in this study and were paired in dyads using random selection. Information about the study was provided to each participant individually and consent forms were obtained from them.

The length of the experiment sessions varied from 15 minutes to 48 minutes, with an average length of 25 minutes. Although there were time variations among sessions, all dyads reviewed the same material and completed the same tasks in Tetralogue. This resulted in approximately 600 minutes of video data and associated participant action log file data. The data stored in the log files were parsed using the 'XML' package [13]. The features extracted from the log files were: number of chat messages sent to the partner and number system helps (viewing educational videos on the subject matter) requested at each stage of the simulation, answer to each individual question, and answer to each group question.

Our focus on "number of messages" and "number of help requests" was driven by former research in the field that associates both features with the performance in learning-oriented tasks, cognitive states, and collaborative interactions [6, 17]. However, more features associated with cognitive activity can be mined from the log files, such as the time length between actions or the content of the chat messages and will be addressed in future studies.

2.3 Video Data Processing and Coding

Facial expression analysis of the video data was performed using the FACET SDK, a commercial version of the Computer Expression Recognition Toolbox [14]. This tool recognizes fine-grained facial features, or facial action units (AUs), described in the Facial Action Coding System [9]. FACET detects human faces in a video frame, locates and tracks facial features, and uses support vector machine based classifiers to output frame-by-frame detection probabilities of a set of facial expressions: anger, joy, contempt and surprise.

In addition, seven trained coders reviewed and coded the videos using the Anvil software [11]. The video data of each participant were assigned to two raters for annotation; however, in three cases there were three raters coding the same video file, and in two cases only a single rater was available for annotation. The raters followed the same coding scheme during the annotation process, which included the next categories: having their hand on their face, expressing engagement, anxiety, or curiosity. As an outcome of the annotation process, the Anvil software produced XML files that were parsed using the 'XML' package [13] in R [22].

Engagement, anxiety, and curiosity were included in the annotation scheme because of the incidence and relevance of these three noncognitive states in simulation games and online learning systems [1, 5]. The coding also included "hand touching face", an affect display gesture that has been linked to affective and cognitive states such as boredom, engagement, and thinking [16].

3. Results

3.1 Behavioral Convergence within Dyads

In order to study evidence of behavioral convergence, features from log files and video data of each of the 24 study participants were represented as a multidimensional behavioral feature vector composed of both the cognitive behaviors: *number_of_messages, number_of_help_requests* and the noncognitive behaviors (i.e. fraction of the time each participant exhibited the behavior): *engagment, hand_on_face, anxiety, curiosity, anger, joy, contempt and surprise*.

An agglomerative hierarchical cluster analysis using an average linkage function was performed on an Euclidean distance matrix (i.e., a similarity matrix) computed from the multidimensional behavioral feature data of the study participants. Our hypothesis is that behavioral convergence will manifest in the cognitive and noncognitive features such that members of the same dyad will tend to group together from the beginning of the clustering process (i.e., they will be closer to each other in the feature space than to others).

Figure 2 depicts the dendrogram plot produced from the cluster analysis. In the plot, members of the same dyad are depicted by consecutive numbers and identical color; for instance, the first dyad includes coparticipants d1.1 and d1.2 colored in red, the second dyad consists of coparticipants d2.1 and d2.2 colored in blue, and so on. The plot shows that participants in 7 of the 12 dyads grouped together in the clustering process (i.e. they were closest to each other in the multidimensional feature space), indicating a high degree of behavioral convergence. Still, some participants (e.g., d10.1 and d4.2) showed a distinctive pattern of values in the variables used to calculate the distances, which prevented them to be grouped with their respective peers.

In addition, we analyzed the similarity matrix of behavioral feature distances for participants within and outside dyads. Behavioral convergence would imply that for dyad members the average distances in feature space is smaller in a statistically significant manner than those of non-dyad members. To study the relative impact of cognitive and noncognitive features we computed two additional similarity matrices: one using exclusively the cognitive features from log files (number of chats messages and number of system help requests) and the other using exclusively noncognitive features produced from the video data (the four facial expression detectors, and the four features from the coding scheme). All features were normalized to present equivalent scaled values between zero and one.

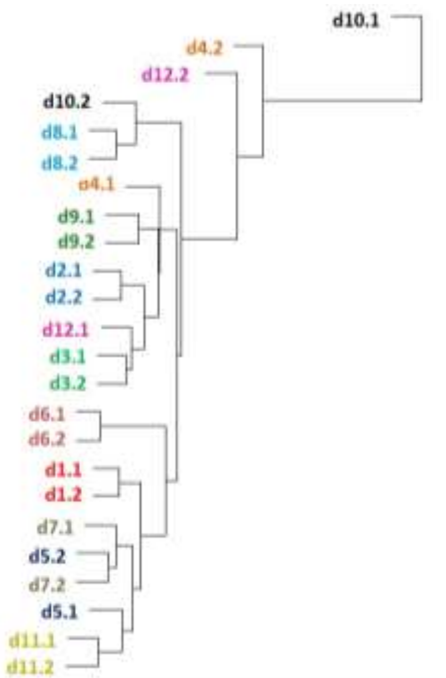

Figure 2. Agglomerative Cluster Dendrogram.

Table 1 shows the mean and standard deviations of feature similarity distances of participants when compared with their dyad partners and others. The results consistently show smaller average distances for the dyads (i.e., members within dyads displayed behavior that was more similar to each other than others), supporting the convergence premise. Additional analysis to test the significance of these differences using the Student's *t*-test demonstrated that using both cognitive and noncognitive features the average distances are statistically significant (t-value = 2.33, df = 11.7, *p*-value < 0.02).

A final analysis was computed on the correlation of the total group scores in the task with the distances of participants with their respective dyad partners and with other users except for their teammate. The group score showed a mild correlation with the distance between dyad members of -0.19 ($s.e._r = 0.21$). Note that the negative correlation is a consequence of using similarity distances (smaller distance values indicate more convergence) and the group score values (higher values indicate a better performance on the task). Nevertheless, as will be underscored in Section 4, the small sample size in the study produced large standard errors for this correlation estimate and do not imply statistically significant patterns.

Table 1. Average and standard deviation of behavioral feature distances within and outside dyads

Features		Mean	S.D.
Cognitive and noncognitive	**Dyad**	0.57	0.22
	Others	0.73	0.24
Cognitive only	**Dyad**	0.36	0.21
	Others	0.57	0.20
Noncognitive only	**Dyad**	0.41	0.17
	Others	0.41	0.22

4. Discussion and Conclusions

Seminal work from Roschelle [23] in his seminal work made the argument that the crux of learning by collaboration is convergence and showed empirical evidence of the convergence occurring at the linguistic level. Our study provides further empirical evidence of behavioral convergence gleaned from multimodal data. As pointed out in [8], cognitive and noncognitive processes occur simultaneously throughout the collaborative task, and both dimensions cannot be separated in practice. The results from cluster analysis in our experimental study support this idea and the pattern of agglomeration of the participants could be interpreted as evidence of convergence of cognitive and noncognitive states when people interact in a collaborative task.

As reported in table 1, the degree of behavioral similarity within dyads tended to be significantly higher than the similarity between non-dyad members, which is good evidence for behavioral convergence in collaborative interactions [4, 12]. In addition, we observed a mild correlation (of approximately 0.2) between the measure of convergence (i.e., the level of similarity between dyads) and the dyad task scores. This might be interpreted as a scaffolding effect that convergence during interaction can have in group performance outcomes. Similar results were reported in [24], underscoring that specific types of convergence have a positive effect in learning and collaboration.

Further research using these data will address topics such as the synchrony of behavior and noncognitive states between members within dyads, machine learning and classification analyses to detect and predict specific cognitive and noncognitive states from facial action units, and more detailed analysis on the impact of cognitive and noncognitive states on the individual-level and group-level assessment outcomes.

There are certain limitations of this study that should be pointed out. First, the current sample size is small ─24 participants─ despite the rich amount of information gathered from each participant. Second, the current collaboration platform neither allows participants to view each other nor uses face-to-face audio-visual interfaces to communicate. This limits how participants are able to mirror each other's behavior and may also explain why we observed weaker convergence in noncognitive features. Third, the study has utilized a very limited set of behaviors both cognitive and noncognitive. We aim to extend our behavior feature set and sources of data (e.g., audio data) in future studies as well as utilize the content of participant chat messages to glean features like shared vocabulary, turn-taking etc.

5. REFERENCES

[1] R. Baker, S.K. D'Mello, M.M.T. Rodrigo, & A.C. Graesser. 2010. Better to Be Frustrated than Bored: The Incidence, Persistence, and Impact of Learners' Cognitive-Affective States during Interactions with Three Different Computer-Based Learning Environments. *International Journal of Human-Computer Studies*, 68, 4, 223-241.

[2] S. Barsade. 2002. The ripple effect: Emotional contagion and its influence on group behavior. *Administrative Science Quarterly*, 47, 644–675.

[3] M. Ben Ammar, M. Neji, A. M. Alimi, & G. Gouardères. 2010. The affective tutoring system. *Expert Systems with Applications*, 37, 4, 3013-3023.

[4] S. Bilakhia, S. Petridis, & M. Pantic. 2013. Audiovisual detection of behavioural mimicry. In *Proceedings of the Humaine Association Conference on Affective Computing and Intelligent Interaction*. 123-128.

[5] R.A. Calvo, & S. D'Mello. 2010. Affect detection: An interdisciplinary review of models, methods, and their applications. *Transactions on Affective Computing*, 1, 1, 18-37.

[6] A.C. Graesser, S.K. D'Mello, S.D. Craig, A. Witherspoon, J. Sullins, B. McDaniel, & B. Gholson. 2008. The relationship between affective states and dialog patterns during interactions with AutoTutor. *Journal of Interactive Learning Research*, 19, 2, 293-312.

[7] S. Greiff. 2012. From interactive to collaborative problem solving: Current issues in the Programme for International Student Assessment. *Review of Psychology*, 19, 2, 111–121.

[8] J. Hao, L. Liu, A.A. von Davier, & P. Kyllonen. 2015. Assessing collaborative problem solving with simulation based task. Paper to be presented at the *11th international conference on computer supported collaborative learning*.

[9] J.C. Hager, P. Ekman, & W.V. Friesen. 2002. *Facial action coding system*. A Human Face, Salt Lake City, UT.

[10] A. Hron, & H.F. Friedrich. 2003. A review of web-based collaborative learning: factors beyond technology. *Journal of Computer Assisted Learning*, 19, 1, 70-79.

[11] M. Kipp. 2012. Multimedia Annotation, Querying and Analysis in ANVIL. In *Multimedia Information Extraction*. M. Maybury, Ed., Wiley-IEEE Computer Society Press, 351-367.

[12] J.L. Lakin, V.E. Jefferis, C.M. Cheng, & T.L. Chartrand. 2003. The chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry. *Journal of nonverbal behavior*, 27, 3, 145-162.

[13] D. T. Lang. 2013. *XML: Tools for parsing and generating XML within R and S-Plus*. R package version 3.98-1.1. http://CRAN.R-project.org/package=XML.

[14] G. Littlewort, J. Whitehill, T. Wu, I. Fasel, M. Frank, J. Movellan, & M. Bartlett. 2011. The computer expression recognition toolbox (CERT). In *Proceedings of the IEEE International Conference on Automatic Face & Gesture Recognition*. 298-305.

[15] L. Liu, J. Hao, A.A. von Davier, P. Kyllonen, & D. Zapata-Rivera. In press. A tough nut to crack: Measuring collaborative problem solving. In *Handbook of Research on Computational Tools for Real-World Skill Development*. Y. Rosen, S. Ferrara, & M. Mosharraf, Eds., IGI-Global, Hershey.

[16] M. Mahmoud, & P. Robinson. 2011. Interpreting Hand Over Face Gestures. In *Proceedings of the International Conference on Affective Computing and Intelligent Interaction*, 248-255.

[17] R. Martinez, J. Kay, J. Wallace, & K. Yacef. 2011. Modelling and identifying collaborative situations in a collocated multi-display groupware setting. In *Proceedings of the 15th International Conference on Artificial Intelligence in Education*, 196-204.

[18] H.F. O'Neil, S.H. Chuang, & G.K.W.K. Chung. 2003. Issues in the computer-based assessment of collaborative problem solving. *Assessment in Education*, 10, 361-373.

[19] J.S. Pardo. 2006. On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, 119, 4, 2382-2393.

[20] D. N. Prata, R. Baker, E. Costa, C. P. Rosé, Y. Cui, & A.M.J.B. de Carvalho. 2009. Detecting and Understanding the Impact of Cognitive and Interpersonal Conflict in Computer Supported Collaborative Learning Environments. In *Proceedings of the 2nd International Conference on Educational Data Mining*. 131-140.

[21] A. Pentland. 2008. *Honest Signals: How they shape our world*. MIT Press, Cambridge, MA.

[22] R Core Team. 2012. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

[23] J. Roschelle. 1992. Learning by collaborating: Convergent conceptual change. *The journal of the learning sciences*, 2, 3, 235-276.

[24] B. Schneider. 2014. Toward Collaboration Sensing: Multimodal Detection of the Chameleon Effect in Collaborative Learning Settings. In *Proceedings of the 7th International Conference on Educational Data Mining*, 435-437.

[25] A.A. Tawfik, L. Sanchez, & D. Saparova. 2014. The Effects of Case Libraries in Supporting Collaborative Problem-Solving in an Online Learning. *Environment. Technology, Knowledge and Learning*, 19, 3, 337-358.

[26] A.A. von Davier, & P.F. Halpin. 2013. *Collaborative Problem Solving and the Assessment of Cognitive Skills: Psychometric Considerations*. ETS Research Report No.RR-13-41. Educational Testing Service, Princeton, NJ.

[27] D. Zapata-Rivera, T. Jackson, L. Liu, M. Bertling, M. Vezzu, & I.R. Katz. 2014. Assessing science inquiry skills using trialogues. In *Intelligent Tutoring Systems*. S. Trausan-Matu, K. Boyer, M. Crosby, & K. Panourgia, Eds., Springer International Publishing, 625-626.