**Data, Log Files, and Data Dependencies**

## Online Appendix

## In this appendix the types of data collected from the collaborative tasks in virtual environments are described.

## Technology is essential to build a foundation for the assessment of collaborative skills through higher fidelity simulations without manually and painstakingly recording the details of a collaborative performance, without the use of pre-and post-tests, and without self-reports (Rosen & Foltz, 2014; Hao, Liu, von Davier, & Kyllonen, 2015). Instead we are now able to record the actions of multiple participants in real time. Moreover, the technology also allows for the collection of different types of data, including multimodal data, which can record affective behavior participants during the collaborative experience but also inform the measurement of the collaborative performance (see Luna Bazaldua et al., 2015). However, the richness of the data that we *can collect* is beyond the scope of standard psychometric models. Collaborative data requires different types of analyses, from time series models and dynamic models from statistics and economics, to algorithms from data mining and machine learning. Similarly, the know-how for task and test development for these constructs is emerging.

## *Process Data.* A challenge in analyzing log file data is determining the meaning of individual actions and chats. There may be some process variables that are relatively easy to measure, such as participation level of each team member and turn taking. However, beyond these kinds of variables interpreting actions and chats may be much more complex because of the dynamics and the sheer volume of data generated in log files.

In collaborative problem-solving, interactions will change over time and will involve time-lagged interrelationships. If there are two people on a team, the actions of one of them will depend both on the actions of the other and on his or her own past actions. These dynamics, which are defined by the interdependence between the individuals on the team, could also offer information that could be used to build a hypothesis about the strategy of the team. For example, by analyzing the covariance of the observed variables (the events), one might hypothesize that an unknown variable, such as the team’s type of strategy, explains why the team chose a particular response and avoided the alternative.

Regarding volume and complexity of interpreting actions and chats, note that collaborative interactions in computerized educational environments produce data of high dimensionality (often containing more variables than people for whom those variables are measured). Extracting key features from the noise surrounding such data is crucial not only to make analysis computationally tractable (Masip, Minguillon, & Mor, 2011), but also to extract relevant features of individuals’ performance.

In addition, with the technological advantages of systems for recording, capturing, and recognition (e.g., Kinect for Windows) of multimodal data, the data from collaborative interactions contain discourse, actions, gestures, tone, body language that result in a deluge of data. One way to attempt to find patterns among these different types of data is to make use of data mining techniques. This is discussed in more detail in the next section.

If the collaboration is measured in a collaborative task by allowing test takers to interact with a computer agent as it is the case in PISA then the challenge of analyzing the dynamic and complex dialogs of humans is avoided and the agent actions are deterministic by having multiple-choice options for each human turn in the chat. In essence, in PISA task design and analyses, the multi-turn and multi-action interactions were converted to multiple-choice scoring events. The multiple-choice approach also can be engineered to guarantee a coherent computer response to each alternative available to the human. There are tools for the item writers to visualize the conversational flow and to make sure there are no missing links in the space of possible conversations.

*Outcome Data.* In the case of collaborative problem solving (CPS) in a specific domain, the outcome data are collected through the evaluative scoring—that is, the scoring of step-by-step responses given by individuals— throughout the process (collaboration). For example, an individual’s actions during the collaboration can be scored as *correct* or *incorrect* by a human rater or an automatic scoring engine. Pretests or posttests, if available, also result in individual outcome data. If either of these tests is available, then the test scores that contain information about the test-taker ability can be corroborated with the information contained in the actions scored throughout the CPS task.

The team level outcome data are straightforward to collect. These data indicate whether a team solved the task successfully or whether parts of the problem were solved correctly. For example, a CPS science simulation task may require the team members to solve a complex problem that requires multiple steps. In the collaborative science simulation called the *tetralogue* the team members are expected to gather information about the activity of a volcano using seismometers and then, make predictions about the probability of an eruption based on the data collected using the seismometers (see Andrews et al, this special issue). The team members may be successful in placing the seismometers and collecting the data, but may make mistake in using this information for prediction.

References

Hao, J., Liu, L., von Davier, A., & Kyllonen, P. (2015), Assessing collaborative problem solving with simulation based tasks, In Oskar Lindwall, Päivi Häkkinen, Timothy Koschmann, Pierre Tchounikine, and Sten Ludvigsen (Eds.) *Exploring the Material Conditions of Learning: Computer Supported Collaborative Learning. Proceedings of 11th international conference on computer supported collaborative learning, Gothenburg, Sweden*

<http://www.isls.org/cscl2015/papers/CSCL2015ProceedingsVolume2.pdf>

Kinect for Windows (2014) *Kinect for Windows*. Retrieved from

<http://www.microsoft.com/en-us/kinectforwindows/>

Luna Bazaldua, D.A., Khan, S., von Davier, A.A., Hao, J., Liu, L., & Zhang, Z. (2015). On convergence of cognitive and noncognitive behavior in collaborative activity. In Olga C. Santos, Jesus G. Boticario, Cristobal Romero, Mykola Pechenizkiy, Agathe Merceron, Piotr Mitros, José María Luna, Cristian Mihaescu, Pablo Moreno, Arnon Hershkovitz, Sebastian Ventura, and Michel Desmarais (Eds.) *Proceedings of the 8th international conference on educational data mining,* Madrid, Spain

<http://www.educationaldatamining.org/EDM2015/uploads/papers/paper_124.pdf>

Masip, D., Minguillon, J., & Mor, E. (2011). Capturing and analyzing student behavior in a virtual learning environment. In C. Romero, S. Ventura, M. Pechenizkiy, & R. S. J. d. Baker (Eds.), *Handbook of educational data mining* (pp. 339–351). Boca Raton, FL: CRC Press.

Rosen, Y., & Foltz, P. W. (2014). Assessing collaborative problem solving through automated technologies. *Research and Practice in Technology Enhanced Learning*, *9*(3), 389–410.