Evidence-Centered Design and Its Application to Collaborative Problem Solving in Practice-based Learning Environments

By: Mutlu Cukurova, Manolis Mavrikis, and Rose Luckin,

UCL Knowledge Lab, University College London

January 2017

About Analytics for Learning (A4L)

The availability of data coming from digital learning environments is creating the possibility to measure learning like never before. The Analytics for Learning (A4L) Network is made up of researchers exploring the measurement of student learning behaviors and strategies in digital learning environments. Learn more: http://analytics4learning.org/.

The Improvement Analytics group engages in collaborative data intensive research to help educational organizations improve learning opportunities for all students. Based within SRI Education, the Improvement Analytics group leverages broad domain expertise and diverse methodological approaches to support those working with learners of all types turn promising ideas into improvements. Across multiple partnerships, we have helped educational organizations find actionable insights within complex data sets and develop as well as productively adapt instructional innovations. Learn more: http://improvement-analytics.org

SRI Education, a division of SRI International, is tackling the most complex issues in education to identify trends, understand outcomes, and guide policy and practice. We work with federal and state agencies, school districts, foundations, nonprofit organizations, and businesses to provide research-based solutions to challenges posed by rapid social, technological and economic change. SRI International is a nonprofit research institute whose innovations have created new industries, extraordinary marketplace value, and lasting benefits to society. Learn more: http://www.sri.com/education.

This material is based upon work supported by the National Science Foundation through grant SMA-1338487. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Evidence-Centered Design and Its Application to Collaborative Problem Solving in Practice-based Learning Environments

By: Mutlu Cukurova, Manolis Mavrikis, and Rose Luckin,
UCL Knowledge Lab, University College London

January 2017

Contents

Introduction

Learning analytics is introducing a number of new techniques and frameworks for studying learning, including collaborative problem solving processes. An increasing number of researchers are using data from students' interactions with learning technologies to support the assessment of collaborative problem solving (Bienkowski, Feng, & Means, 2012). Moreover, to create shared understanding among the multiple disciplines involved in learning analytics research (such as psychology, social psychology, the learning sciences, machine learning, statistics, and artificial intelligence) and those studying learning processes like collaborative problem solving, researchers recently started working on theoretical frameworks to share insights, receive feedback, and build on one another's efforts (Bhanot, Cheng, & Krumm, 2016). Evidence-centered design (ECD) (Mislevy & Haertel, 2006), an assessment design process that assessment designers use to articulate design goals and decisions, has been leveraged successfully as such a framework. ECD was shown to be very useful in framing large data sets generated in digital learning environments, offering data on students' interactions with a system that made it possible to track and identify student learning processes such as "gaming the system" (Baker et al., 2009), "engagement" (D'Mello, Lehman, & Person, 2010), "wheel spinning" (Feng, 2015), and "task persistence" (DiCerbo, 2014).

Although ECD has been shown to be useful for assessment design in digital learning environments (Bhanot et al., 2016), we argue that its potential is broader, that it is also applicable to multimodal learning analytics from practice-based learning environments. In this paper, we first present the practice-based learning environment that we are studying, and then we show how ECD could be deployed for the assessment of collaborative problem solving processes in these learning environments.

Collaborative Problem Solving in Practice-based Learning Environments

Policy makers and employers are concerned that students are not graduating with the required 21st century skills in STEM subjects (science, technology, engineering, and mathematics). They argue that learners must go beyond the acquisition of discipline-specific facts and skills to develop an integrated understanding of STEM disciplines in an authentic context of collaborative problem solving. Collaborative problem solving is complex. Both constructs encompassed in it—collaboration and problem solving—have multiple interpretations (see for instance Dillenbourg, 1999; Lin et al., 2015). In this research study, collaborative problem solving refers to the process through which a small number of participants who share a problem state and a goal apply their social and cognitive knowledge and skills to the problem to achieve a solution.

The focal construct of this paper is collaborative problem solving. Recently, much research on collaborative problem solving processes has highlighted their potential for equipping young people with the skills and experiences necessary for successfully participating in and contributing to education and the workforce in the 21st century. Collaborative problem solving differs greatly depending on contextual factors. For instance, within a massive open online course (MOOC), collaborative problem solving involves asynchronous contributions from participants who have broadly varied knowledge and skills. The dynamics of participants' interactions in a digital environment are different from those of a pair of students who have more or less similar knowledge and skills working synchronously and face-to-face in a classroom. Such factors can influence how collaborative problem solving is practiced, and they should be taken into account in discussions about this complex process. We are particularly interested in STEM education environments in schools where most attempts at having students solve problems collaboratively occur in practice-based learning activities in which a small number of students with similar education levels work in the same physical environment synchronously.

Practice-based learning activities are considered to be a significant part of STEM education and are believed to have the potential to help educators achieve high-tier institutional and policy goals, such as developing 21st century skills in STEM subjects at scale. The potential of practice-based learning activities to contribute to students' collaborative problem solving is very clear in theory, but assessing its value in real-life contexts is problematic. As Fredricks et al. (2011) emphasized, interest in researching 21st century skills including collaborative problem solving is increasing, although their assessment remains a challenge. In traditional educational research, the most common approach to assess such processes is to apply self-report measures. However, respondents tend to answer in socially desirable ways rather than expressing their genuine opinions and insights, which jeopardizes the measures' value in eliciting appropriate interpretations (Shechtman, DeBarger, Dornsife, Rosier, & Yarnall, 2013; Shute & Ventura, 2013). Moreover, most of 21st century skills relate to students' ability to function in a social space, and traditional assessment approaches often capture only individual knowledge and skills whose value is open to criticism in group settings.

Thanks to advances in technology, including tracking sensors and mobile tools, it is now possible to capture more complex performance in assessment settings, and automated methods have become available for analyzing complex learning processes in complex learning environments. The data collected from such technologies should be organized in a way that best serves the purposes of the assessment of the focal construct. With this goal in mind, we leveraged ECD and argue that it not only helps researchers organize their thinking, but also provides a common terminology that makes assessment strategies easier to reuse and share. Before we get into the details of ECD, we will present the learning context we studied to make our data collection points clear.

Learning Context

In this paper, we focus on open-ended, hands-on, physical computing design tasks. This type of practice-based learning activity is commonly used to improve collaborative problem solving, and its popularity is increasing exponentially in both secondary and postsecondary learning institutions, after the introduction of the Makers Movement (Halverson & Sheridan, 2014).

We studied practice-based learning at three education levels: Engineering, Interaction Design, and High School using the Arduino physical computing toolkit.[1] We designed learning activities that aligned with the learning objectives of the curriculum for these three education levels. They all had similar initial introductory tasks (such as blinking an LED on/off with a timer, blinking an LED on/off with a button, using a potentiometer to control an LED) in order to familiarize students with the Arduino Visual Programming platform and the learning analytics system (Figure 1).

Figure 1. Learning Context

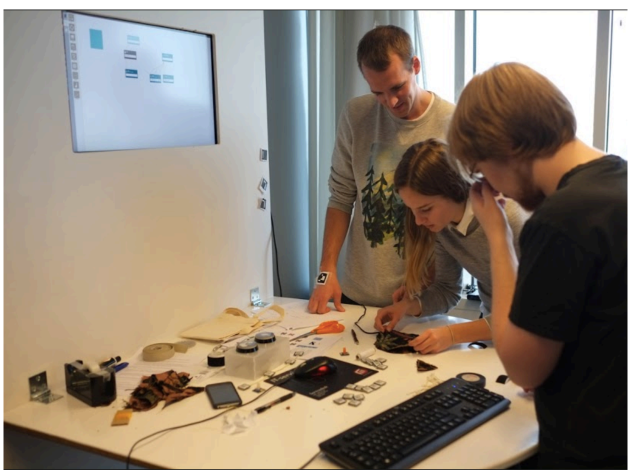


Figure 2. Engineering Learning Task Kit

Then in each educational context, students followed a different open-ended investigation. For instance, in the Engineering context students were asked to build a basic infrastructure for a smart home (Figure 2). Students were provided with different sensors and actuators to control events in their open-ended investigations.

[1] https://www.arduino.cc/

Evidence Centered Design (ECD)

Learning analytics can contribute to the assessment of learning processes including collaborative problem solving, yet it has not yet contributed significantly to practice-based learning (Worsley & Blikstein, 2014). Recently, with the help of multimodal learning analytics that capture streams of data from sensors and from learners' activities, researchers have been able to generate large data sets of students' interactions during collaborative problem solving in practice-based learning environments (Spikol et al., 2016). ECD can be useful as a foundation for the systematic investigation of the collaborative problem solving process in these learning contexts.

ECD is an approach to constructing educational assessments in terms of evidentiary arguments (Mislevy, Almond, & Lukas, 2003). It focuses on three models and their interrelationships:

- The ***student model*** defines one or more variables related to the knowledge, skills, and abilities we wish to measure.
- **The *task model*** describes how to structure the learning situations to obtain the kinds of data needed for the evidence models. It has three parts: the characteristic features of the task, the variable features of the task, and the potential task products.
- **The *evidence model*** provides detailed instructions on how we should update our information about the student model variables given a performance in the form of examinees' work products from tasks. It has two parts: potential observations and potential frameworks (Bhanot et al., 2016).

Data Collection

The data collection tool described in this paper consists of a purpose-built learning environment with multiple sensors to collect data during practice-based activities, with added web and mobile tools for learners to document their learning activities. The learning environment is a workshop that includes a specially designed table with a built-in display that is connected to a freestanding wall (Figure 3). This special work area accommodates a group of up to four students. The workstation collects a range of data, including log files from the programming of physical computing kits, and tracks the different physical components that are being used by the students. Students' mobile inputs are captured (in the current release), and we have created two large buttons (sentiment buttons) that students can push to signify "eureka" and "frustration" moments.

Figure 3. Learning Analytics System

These buttons trigger the system to capture a screen shot of the programming activity and a photo of the action on the table. Students need to plan, document, and reflect on their solution by entering brief text descriptions and capturing photographs and video through a mobile system. The mobile system also provides a tool for research observations to be marked (in a temporal sense) “on the fly,” providing a bridge between the data the system collects and the data the human observer collects. Table 1 illustrates this multimodal learning analytics system, describing the types of data collected, the sensors involved, and the split between machine and human data collection and coding. The annotation data the researchers and the learners create serves as a coarse temporal breakdown of the activities. These web-based mobile tools (cross platform) also provide a key device for researchers to provide coarse real-time encoding based on an analysis framework for collaborative work (Cukurova, Avramides, Spikol, Luckin, & Mavrikis, 2016). We intend to explore how this coarse coding done by researchers and learners could supply more automatic segmentation between the different design activities, thus structuring the machine-collected data.

Table 1. Multimodal Learning Analytics System

Data Collector	Arduino IDE	Mobile System	Sentiment Buttons
Type of data/events	DATA: Number of components and inputs EVENTS: Arduino modules and codes	DATA: Number of posts, transitions between activities EVENTS: Self-documentation	DATA andEVENTS: Critical incidents
Brief description	How students designed and built their designs—types and number of components and manipulation in the IDE	How the students planned, documented, and reflected on their design-critical incident marks, researcher-coded activities	Marking of critical incidents
Type of analysis	Machine	Machine and human	Machine and human

Note: IED is Integrated Development Environment.

This Application of ECD

The backbone of ECD is the systematic articulation of models of the student competencies or abilities to be measured by an assessment, models of the task requirements of an assessment, and models of the evidence used to make inferences about students based on assessment results.

Student Model

In practice-based learning environments where multiple data collectors are used to provide insights into learning processes (see Table 1), task and evidence models need to be rearranged to fit with the type of data collected. Student models do not vary depending on the type of data collected because they are related to the students' knowledge, skills, and abilities that are linked to the focal construct being investigated.

One thing that is very clear is that collaborative problem solving is a very complex process. There could be numerous reasons for a group to present exceptional results in solving problems collaboratively or to fail. A potential failure in collaborative problem solving has been attributed to a number of cognitive and social factors including but not limited to cognitive load (Dillenbourg, 1999), lack of coordination (Steiner, 1974), disruption and production blocking of individual contributions (Diehl & Stroebe, 1987), diffusion of responsibility (Latane, Williams, & Harkins, 1979), fear of evaluation (Goethals & Darley, 1987), management of attention (Barron, 2003), individuals' prior knowledge (Wiley & Jolly, 2003), and individuals' expertise at task (Nokes-Malach, Meade, & Morrow, 2012). Inclusion of any of these variables into models could improve the accuracy in detecting participants' collaborative problem solving processes.

On the other hand, task models and evidence models do vary depending on the type of data being collected. We will now present the task models and evidence models and how they differ depending on the data collected.

Task Model

As stated, task models describe how to structure the learning situation to obtain the kinds of data needed for the evidence models. They have three parts: the characteristic features of the task, the variable features of the task, and the potential task products.

Since in multimodal learning analytics systems there are multiple data collection points, each data collection point has its own task model and evidence model.

Arduino Integrated Development Environment (IDE): For collection of meaningful data from the IDE records, the learning task should involve use of the visual programming interface of the physical computing kit. Students should be stimulated to connect and disconnect the physical computing modules, and all programming and connecting/disconnecting activities should be logged. Within the task, the complexity and challenge of the products and the number and variety of the physical computing toolkit modules and logics provided can be varied. For instance, for novice or younger students, the number and variety of physical computing modules can be kept smaller. This shift in difficulty may elicit different levels of collaboration among participants. Potential task products involve the number of attempts made before or the time spent on reaching a solution or stopping work, as well as how frequently students plug in components and interact with IDE.

Mobile system: Similarly, appropriate mobile systems should be integrated into the learning environment considering all the practical and technical issues related to participants' use of mobile tools as part of the learning tasks. Students should be stimulated and/or reminded to use the mobile tools during the learning activities, and input though mobile tools should be logged. Within the task, number of mobile submits or words required/allowed in text can be varied to increase or decrease expected time and effort students put into reflection. The mobile system can also set a time requirement for different stages of students' collaborative design tasks, such as planning, building, and reflecting, and these variables can be changed. These changes in the amount and quality of students' documentation may elicit different levels collaborative problem solving among participants.

Sentiment buttons: The greatest issue with sentiment buttons in data collection is that participants often ignore or forget to use them during completion of the tasks. Hence, as part of the characteristic features of the task should be periodic reminders to participants to use the buttons. This could be achieved with a color code. For instance, the sentiment box could turn from green to red every 5 minutes for students to provide feedback on their progress. These incidents should be logged, pictured, and recorded. Response time to submit a eureka or a frustration moment, frequency between types of incidents, and duration between types of incidents can be measured as potential task products.

Evidence Model

Arduino IDE: The task products can manifest themselves in many ways. For instance, the attempts students make before reaching a solution or stopping work can be used to interpret more complex collaborative problem solving since the greater number and variety of modules and logics used may represent it. On the other hand, participants' limited or nonexistent access to alternative logics or sets of components can be interpreted as a lack of ability to solve problems collaboratively. Moreover, measuring the amount of time spent on visual programming can be useful for interpretation, with, for example, a group's too fast/slow visual programming interpreted as less effective collaborative problem solving.

Mobile system: The mobile system can provide various types of evidence. For instance, limited/too much time spent on different design stages (such as problem scoping, building, or reflecting) can indicate the quality of collaborative problem solving. Verbosity of responses can be used to interpret the amount of reflective practice that participants went through. Reflective practice data can be used to interpret the complexity of the collaborative problem solving process. More time spent on the problem scoping stage may be interpreted as a sign of better problem solving. Certain text mining could be done on the input provided by the participants using the mobile tool to measure the effort put in to establish a shared understanding, which is considered an essential aspect of collaborative problem solving.

Sentiment buttons: Potential observations in the evidence model include the limited or no accessing of eureka moments, too frequent use of frustration moments or too little time spent between a eureka moment and a frustration moment. For instance, a low ratio of eureka moments to frustration moments may reflect participants' lack of ability to solve problems collaboratively.

In addition to these models, human coders (similar to the BROMP methodology for affect, c.f. Baker et al., 2009) can collect observations across the evidence model, rules linking observations to competencies of collaborative problem solving can be used as a triangulation method (Cukurova, Avramides, Luckin, Mavrikis, 2016), and Bayesian networks can be used for combining evidence from different sources.

Results

We recently conducted our first full test of the system with two groups of three students. The user trial procedure was to set up to test and introduce the students to the system. Each of the students was shown how to use the mobile reporting system. Then the students were guided through a hands-on introduction to the visual programming platform (IDE) that included working with sensors and actuator blocks and programming them. The research observer using the mobile device began coding of the activity with marking events (design stages of project scoping, project realization, and reflection) while the students used the tool to capture planning, documenting, and reflecting.

Figure 4. Initial Visualizations Focusing on IDE and the Sentiment Buttons

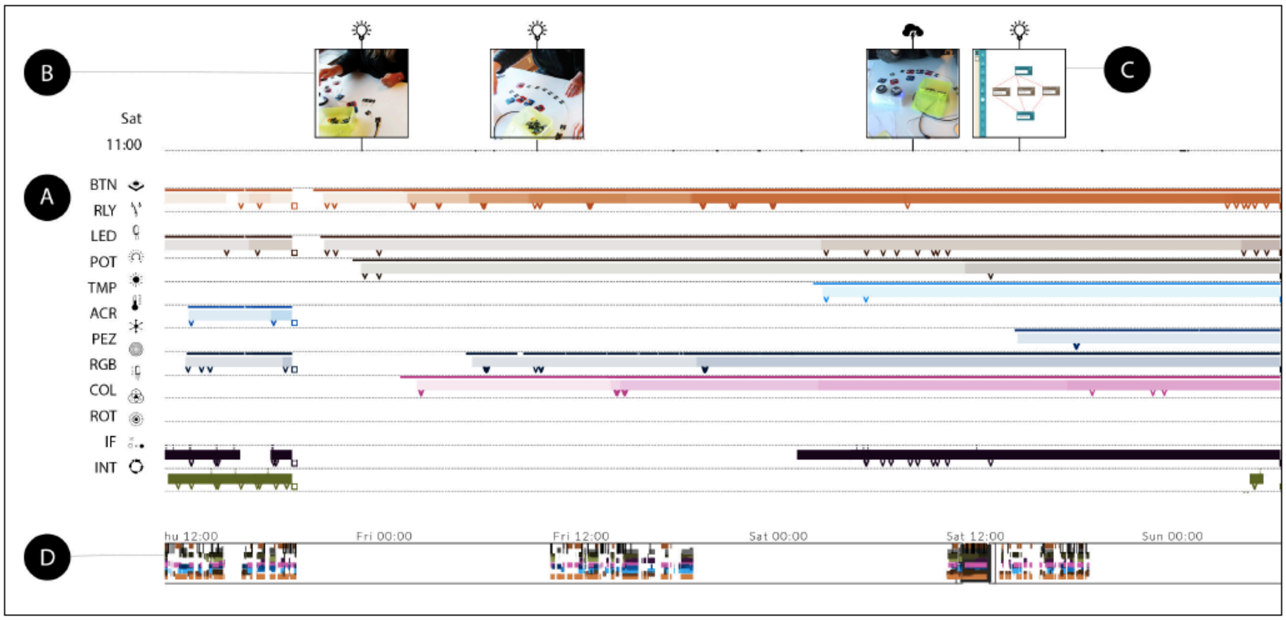


Figure 4 presents one of our initial visualizations. The physical connection of a component is represented as a strong thin line and software use as a rectangle, each extending for the period of time they were either physically or digitally connected. The color of the component's visual representation depends on whether it was an input (button, sensor, etc.) or output. Any connection made is represented as a triangle on the element connected, and each end of the connection on that element is represented as a square at the moment of disconnection. These are placed in line with that element's general linear representation track. Using these visualizations, our objective was to interpret

- repeated attempts students made before reaching a solution or stopping work;
- the amount of time students spent on reaching a solution or stopping work; and

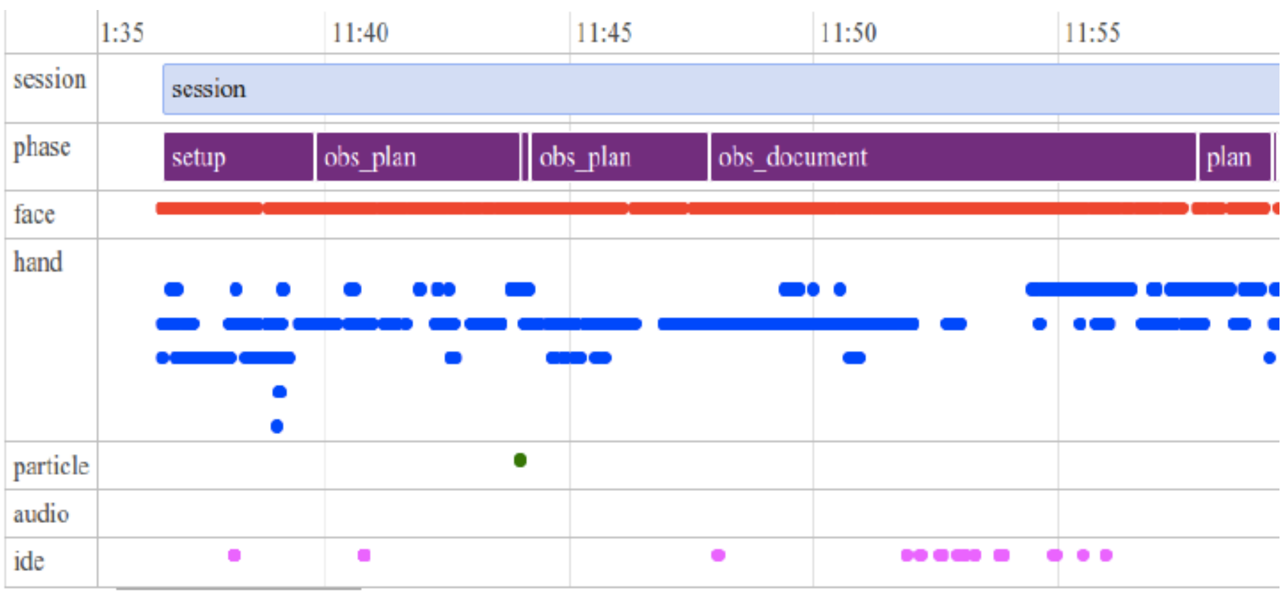
- the type, complexity, and variety of alternative logic or the set of components students used to reach their solutions.

The sentiment button icons and corresponding snapshots act as a marker throughout the stream of data that persists underneath. The sentiment feedback buttons let students input a positive/negative sentiment (represented with a light bulb representing a eureka moment and a storm cloud denoting a frustration moment on the front of the button box). Using these visualizations we are aiming to detect

- limited or no accessing of the eureka sentiment button;
- too frequent use of the frustration sentiment button; and
- too little time spent between a eureka moment and a frustration moment.

Figure 5 presents a visualization that breaks down the planning, building, and reflecting stages across the data generated by the combination of research observations and students' input. As can be seen, we can visualize some data regarding our evidence models including the amount and/or quality of documentation, the amount of time spent in the planning phase. and the quality of the outcome.

Figure 5. Visualization of the Breakdown of Planning, Building, and Reflecting



This visualization helps us to interpret

- the amount and/or quality of documentation; and
- the amount of time spent on different phases.

However, we encountered significant challenges in implementing the ECD task and evidence models for assessing collaborative problem solving in this practice-based learning context. These mainly related to the complexity of our focal construct. Even though ECD is a very useful for well-defined granular learning phenomena, the proximity of the evidence model seems to decrease as the focal construct increases in coarseness (c.f. the “wheel-spinning” vs. “engagement” design pattern). Hence, a next step would be to apply ECD to more fine-grained competencies in the collaborative problem solving process. Rather than trying to create an ECD for complex phenomena in learning such as collaborative problem solving, we believe these phenomena should be investigated at smaller grains and ECDs created to investigate those smaller grains. For instance, it would be more applicable to create an ECD for the assessment of ability to establish and maintain shared understanding, which is considered to be an essential aspect of collaborative problem solving (OECD, 2015), or of ability to vocalize knowledge, an important aspect of ability to establish and maintain shared understanding. With finer grains, the accuracy and flexibility of evidence models increase.

In previous work (Cukurova et al., 2016), we identified nine competencies of collaborative problem solving in practice-based learning environments. Three core competencies are particularly related to collaboration and six are particularly related to problem solving.

Competencies related to collaboration:

1. Establishing and maintaining shared understanding
2. Taking appropriate action to solve the problem
3. Establishing and maintaining team organization

Competencies related to problem solving

1. Identifying facts
2. Representing and formulating
3. Generating hypotheses
4. Planning and executing
5. Identifying knowledge and skill deficiencies
6. Monitoring, reflecting, and applying

We suggest that the assessment of these more granular competencies related to collaborative problem solving would provide valuable evidence for the assessment of the collaborative problem solving process as a whole in practice-based learning environments. That would require the creation of ECDs for each of these competencies and combining them for an overall assessment of the coarser construct. We invite other researchers to investigate these competencies while aiming to make sense of collaborative problem solving process.

Conclusions

We have described how ECD can be leveraged to help investigate collaborative problem solving in practice-based learning environments with multimodal learning analytics. Although initially ECD is designed to frame large data sets from digital learning environments, its use to create a common language among researchers who are trained in various domains related to but distinct from each other can be expanded to practice-based learning environments. Moreover, ECD can be used to generate insights into complex noncognitive constructs as well as shape the design of assessments of them. Collaborative problem solving processes are too complex to be assessed as a whole in the context of practice-based learning environments. However, ECD can be more effectively deployed to generate evidence about the existence of competencies related to collaborative problem solving processes. For future research, we invite researchers to work on the assessment of the competencies of collaborative problem solving. Using the evidence generated and related to these competencies, groups who present exceptional results in solving problems collaboratively could be compared with those who fail. These comparisons would help identify competency patterns that lead to success in collaborative problem solving. In this way, we believe that ECD has the potential to help us understand learning processes that occur in complex learning environments in addition to its use as an assessment design tool.

We are also working on a potential methodology and tools that would help with the observation of evidence for competencies related to collaborative problem solving in practice-based learning. These methods and tools can be used to triangulate results generated from multimodal learning analytics systems. Combining these with the multimodal data collected from the learning analytics system would allow both predictive and diagnostic modeling.

Acknowledgment

The learning analytics system discussed in this paper was created in the PELARS project, and it was funded under the Seventh Framework Programme of the European Commission (GA No. 619738). We particularly thank the Malmo University team, which is led by Dr Daniel Spikol, and the Copenhagen Institute of Interaction Design team, which is led by Dr Nina Valkanova, for providing us the presented visualizations in the paper.

References

Baker, R. S., de Carvalho, A., Raspat, J., Aleven, V., Corbett, A. T., & Koedinger, K. R. (2009). *Educational software features that encourage and discourage "gaming the system."* Paper presented at the 14th International Conference on Artificial Intelligence in Education, Brighton, UK.

Barron, B. (2003). When smart groups fail. *Journal of the Learning Sciences, 12*(3), 307-359.

Bhanot, R. T., Cheng, B. H., & Krumm, A. E. (2016). *Investigating noncognitive constructs as part of learning in digital environments using evidence-centred design.* Paper presented at the Annual meeting of American Educational Research Association, Washington, DC.

Bienkowski, M., Feng, M., & Means, B. (2012). *Enhancing teahcing and learning through educational data mining and learning analytics: An issue brief*. Washington, DC: U.S. Department of Education Office of Educational Technology.

Cukurova, M., Avramides, K., Luckin, R., Mavrikis, M. (2016). Revelaing behaviour pattern differences in collaborative problem solving. K. Verbert et al. (Eds.): *EC_TEL 2016, LNCS 9891*, pp. 563-569, Springer: Switzerland.

Cukurova, M., Avramides, K., Spikol, D., Luckin, R., & Mavrikis, M. (2016). An analysis framework for collaborative problem solving in practice-based learning activities: A mixed-method approach. Proceedings of the Sixth International Conference on Learning Analytics & Knowledge, p. 84-88, ACM New York, NY, USA.

Diehl, M., & Stroebe, W. (1987). Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of personality and social psychology*,*53*(3), 497.

D'Mello, S. K., Lehman, B., & Person, N. (2010). Monitoring affect states during effortful problem solving activities. *International Journal of Artifilcial Intelligence in Education, 20*(4), 361-389.

DiCerbo, K. E. (2014). Game-based assessment of persistence. *Educational Technology & Society, 17*(1), 17-28.

Dillenbourg P. (1999) What do yuo mean by collaborative leraning?. In P. Dillenbourg (Ed) Collaborative-learning: Cognitive and Computational Approaches. (pp.1-19). Oxford: Elsevier.

Feng, M. (2015). Wheel-spinning desing pattern. Retrieved from SRI Education, Data Science Group, website: http://analytics4learning.org/wp-content/uploads/2015/03/wheel-spinning_FINAL.pdf

Fredricks, J., McColskey, W., Meli, J., Mordica, J., Montrosse, B., & Mooney, K. (2011). *Measuring sudent engagement in upper elementary through high school: A description of 21 instruments*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (REL 2011-No. 098).

Goethals, G. R., & Darley, J. M. (1987). Social comparison theory: Self-evaluation and group life. *Theories of group behavior* (pp. 21-47). Springer: New York.

Halverson, E. R., & Sheridan, K. M. (2014). The Maker Movement in education. *Harvard Educational Review, 84*(4), 495-506.

Latane, B., Williams, K., & Harkins, S. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of personality and social psychology, 37*(6), 822.

Lin, K. Y., Yu, K. C., Hsiao, H. S., Chu, Y. H., Chang, Y. S., & Chien, Y. H. (2015). Design of an assessment system for collaborative problem solving in STEM education. *Journal of Computers in Education, 2*(3), 301-322.

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). *A brief introduction to evidence-centered design*. Princeton, NJ: Educational Testing Service, Research & Development Division.

Mislevy, R. J., & Haertel, G. D. (2006). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice, 25*(4), 6-20.

Nijstad, B. A., Diehl, M., & Stroebe, W. (2003). Cognitive stimulation and interference in idea generating groups. *Group creativity: Innovation through collaboration* (pp. 137-159), Oxford University Press, Oxford.

Nokes-Malach, T. J., Meade, M. L., & Morrow, D. G. (2012). The effect of expertise on collaborative problem solving. *Thinking & Reasoning, 18*(1), 32-58.

OECD. (2015). Draft Collaborative Problem Solving Framework, Retrieved from https://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Collaborative%20Problem%20Solving%20Framework%20.pdf

Shechtman, N., DeBarger, A. H., Dornsife, C., Rosier, S., & Yarnall, L. (2013). *Promoting grit, tenacity, and perseverance: Critical factors for success in the 21st century*. Washington, DC: U.S. Department of Education.

Shute, V., & Ventura, M. (2013). Stealth assessment: Measuring and supporting learning in video games. Cambridge, MA: MIT Press.

Spikol, D., Avramides, K., Cukurova, M., Vogel, B., Luckin, R., Mavrikis, M., & Ruffaldi, E. (2016). Exploring the interplay between human and machine annotated multimodal learning analytics in hands-on STEM Activities. Proceedings of the Sixth International Conference on Learning Analytics & Knowledge, p. 522-523, ACM New York, NY, USA.

Steiner, I. D. (1974). Whatever happened to the group in social psychology? *Journal of Experimental Social Psychology, 10*(1), 94-108.

Wiley, J., & Jolly, C. (2003). *When two heads are better than one expert.* Paper presented at the Proceedings of the 25th annual conference of the Cognitive Science Society, Cognitive Science Society Boston, Massachusetts, USA.

Worsley, M., & Blikstein, P. (2014). Analyzing engineering design through the lens of computation. *Journal of Learning Analytics, 1*(2), 151-186.

Appendix: Student Model, Task Models and Evidence Models

Student Model	
Focal construct	**Collaborative problem solving**—the process through which a small number of participants (from dyads to classroom size) who share a problem state and a goal apply their social and cognitive knowledge and skills to the problem to achieve a solution.
Additional knowledge, skills, and abilities	• Cognitive load (Dillenbourg, 1999) • Lack of coordination (Steiner, 1974) • Disruption and production blocking of individual contributions (Nijstad, Diehl, & Stroebe, 2003) • Diffusion of responsibility (Latane, Williams, & Harkins, 1979) • Fear of evaluation (Goethals & Darley, 1987) (Goethals & Darley, 1987) • Management of attention (Barron, 2003) • Individuals' prior knowledge (Wiley & Jolly, 2003) • Individuals' expertise at task (Nokes-Malach, Meade, & Morrow, 2012)

Data Collector	Arduino IDE	Mobile system	Sentiment buttons
Task Model			
Characteristic features of the task	• Students use visual programming in the physical computing kit. • Students connect and disconnect physical computing modules. • Programming interactions are logged.	• Appropriate mobile tools are integrated into the learning environment. • Students are stimulated to use the mobile documentation tool. • Input through the mobile tool is logged.	• Students are reminded to use the buttons with time intervals. • Incidents are logged. • Incident moments are pictured and recorded.
Variable features of task	• Complexity and challenge of products • Number and variety of the physical computing toolkit modules and logics provided	• Number of mobile submits required/allowed • Number of words in text submitted • Required duration of different activities to progress • Content of text and multimedia	• Time intervals of sentiment button reminders
Potential task products	• Attempts students made before reaching a solution or stopping work • Time spent on reaching a solution or stopping work • Accessing of alternative logic or set of components • Frequency of components plugged in • Frequency of interactions with IDE • Consequence of reaching/not reaching a solution	• Detailed documentation of the planning phase • Performance on tasks completed before reaching a solution	• Response time to submit a eureka or a frustration moment • Accessing of supports/hints • Frequency between types of incidents • Duration between types of incidents

Evidence Model			
Potential observations	• Repeated attempts students make before reaching a solution or stopping work • Extended or too little time spent on reaching a solution or stopping work • Limited or no accessing of alternative logic or set of components	• High amount and/or quality of documentation • Low performance on tasks completed before reaching a solution or stopping work • Limited amount of time spent on design phases (such as problem scoping, reflecting) • Low quality of product (through taken pictures) • Verbosity of responses	• Limited or no accessing of eureka sentiment button • Too frequent use of frustration sentiment button • Too little time spend between a eureka moment and a frustration moment
Potential frameworks	• More and greater variety of modules and logics used, more complex problem solving • Algorithmic complexity • Too fast/slow programming	• More time spent on documentation, deeper reflection. • Quality of the documentation, verbosity, deeper reflection • More time spent on problem scoping, better problem solving • More time spent on establishing shared understanding, better collaboration	• Delay of response to sentiment feedback request
	• Human coders (similar to BROMP methodology for affect, c.f. Baker et al., 2009) can collect observations across evidence models • Rules linking observations to competencies of collaborative problem solving (Cukurova et al., 2016) • Bayesian networks for combining evidence from different sources		

A4L Analytics for Learning