Data Analytics and Computational Thinking Skills in Construction Engineering and Management Education: A Conceptual System

Abiola A. Akanmu, Ph.D.[1]; Vincent S. Akligo[2]; Omobolanle R. Ogunseiju[3]; Sang Won Lee, Ph.D.[4]; and Homero Murzi, Ph.D.[5]

[1]Associate Professor, Myers-Lawson School of Construction, Virginia Tech, Blacksburg, VA (corresponding author). Email: abiola@vt.edu
[2]Myers-Lawson School of Construction, Virginia Tech, Blacksburg, VA.
Email: vincentsa@vt.edu
[3]Myers-Lawson School of Construction, Virginia Tech, Blacksburg, VA.
Email: omobolanle@vt.edu
[4]Assistant Professor, Dept. of Computer Science, Virginia Tech, Blacksburg, VA.
Email: sangwonlee@vt.edu
[5]Assistant Professor, Dept. of Engineering Education, Virginia Tech, Blacksburg, VA.
Email: hmurzi@vt.edu

ABSTRACT

Data analytics and computational thinking are essential for processing and analyzing data from sensors, and presenting the results in formats suitable for decision-making. However, most undergraduate construction engineering and management students struggle with understanding the required computational concepts and workflows because they lack the theoretical foundations. This has resulted in a shortage of skilled workforce equipped with the required competencies for developing sustainable solutions with sensor data. End-user programming environments present students with a means to execute complex analysis by employing visual programming mechanics. With end-user programming, students can easily formulate problems, logically organize, analyze sensor data, represent data through abstractions, and adapt the results to a wide variety of problems. This paper presents a conceptual system based on end-user programming and grounded in the Learning-for-Use theory which can equip construction engineering and management students with the competencies needed to implement sensor data analytics in the construction industry. The system allows students to specify algorithms by directly interacting with data and objects to analyze sensor data and generate information to support decision-making in construction projects. An envisioned scenario is presented to demonstrate the potential of the system in advancing students' data analytics and computational thinking skills. The study contributes to existing knowledge in the application of computational thinking and data analytics paradigms in construction engineering education.

INTRODUCTION

The construction industry, one of the industries with the largest labor force in the United States (8% of the total workforce (BLS 2019), has long suffered from productivity loss, premature exits of workers from the workforce, and safety issues. To address these challenges, the construction industry has begun investing in sensing technologies (e.g., laser scanners, camera drones, real-time location sensing systems and radio frequency identification (RFID) tags) that can enhance access to critical information needed for quick and informed decision-making. The need to adopt these sensing technologies is also reinforced by the benefits of improved performance currently realized by other industry sectors (e.g., manufacturing industries). Data analytics and computational thinking are needed to process data from sensors, analyze the data, and present the resulting models in formats suitable for decision-making. Unfortunately, most undergraduate construction engineering and management (CEM) students struggle to understand the computational concepts and workflows required, because they lack the understanding of how to translate data into knowledge for supporting decisions (Shyamala et al. 2017).While a small number of institutions have begun adding sensing technologies and data analytics into their undergraduate CEM programs, some of these courses suffer from low enrollment rates, due to high entry barriers of learning computational concepts that are separate from the target context. This, in turn, has led to a shortage of the future CEM workforce equipped with the required competencies for developing and implementing sustainable solutions with sensor data. This shortage of skilled workers has been identified as an impediment to the technological growth of the construction industry.

Clarke et al. (2005) suggests that computational concepts should be presented in a way that reduces cognitive burden while maximizing pedagogical value. Several studies have provided evidence of the potential of programming environments to stimulate students' interest and enhance their competencies in engineering disciplines. Elements of computation models (e.g., variables, algorithms, functions, and methods) can be represented as interactive visual blocks and elements (Letondal 2006). An end-user programming (EUP) environment, into which interactive objects (e.g., abstracted real-world resources, such as vehicles, inventory, and workers) are integrated, can provide ways to facilitate computational thinking within the target context, enabling students to analyze sensor data in approachable ways and understand essential computational concepts without the burden of knowing detailed programming syntax (Ko et al. 2011). With such EUP environment, undergraduate students will be able to: understand the structures, functions, behaviors, states, and relationships between objects on a construction site and how they are related to the safety and productivity of construction projects by exploring them interactively; quickly prototype analytical methods by specifying algorithms with interactive visual objects; and review analysis results and iteratively develop desirable models for further intelligent algorithms (Ko et al. 2011).

This paper proposes a conceptual system that uses an EUP environment for equipping CEM students with competencies and computational thinking skills for addressing construction challenges with sensor data analytic. The need for sensor data analytics in the construction industry, challenges to equipping CEM students with sensor data analytic skills and opportunities offered by EUP to equip students with sensor data analytic and computational thinking skills are presented. A conceptual system grounded in the four tenets of the Learning-for-Use framework propounded by Edelson (2001) is proposed to demonstrate the potential of EUP for advancing knowledge construction and application design. A potential application of the framework for addressing the complexity of interpreting and analyzing sensor data is presented.

LITERATURE REVIEW

Sensor and Data Analytics in the Construction Industry. In recent years, there has been a rapid increase in the adoption of sensing technologies in the construction industry. The investment in sensing technologies in the past five years have been reported to have doubled the investment of 2008-2012 (Dixon et al. 2018). According to Omar and Nehdi (2016), there have been increasing reports of the use of sensors and other data acquisition technologies for tracking real-time performance levels and physical states of various resources, such as personnel, materials, and heavy equipment. Data captured and recorded from these technologies are becoming overwhelming, and companies are seeking ways to structure and analyze these data to make actionable decisions to improve their bottom line. With the advancements in artificial intelligence and machine learning, the construction companies will be able to use a combination of internal and external data sets to predict future outcomes on projects. Data analytics can help firms (1) determine the most profitable projects to pursue, (2) predict and manage possible safety and performance risks to their projects, and (3) design more sustainable and adaptable buildings and civil infrastructure systems. Some construction companies are already exploring the potential of data analytics. The Forum (2016) has predicted that the rate of adoption of advanced sensing and analytics techniques could result in significant savings in the design, construction, and operations and maintenance phases. This advancement and diffusion of sensing technologies in the construction industry, has triggered the need to equip the future workforce with such skills.

Undergraduate CEM Education Needs. To prepare the future CEM workforce to address the aforementioned challenge, data analytics and computational thinking skills are required. CEM students are better positioned to address the challenges, because of their knowledge and grounding in their domain field. Unlike traditional CEM courses, computational thinking draws on computer science concepts, such as abstraction, decomposition, and simulation. This can play a significant role in the CEM discipline, as an effective approach for addressing the challenging interdisciplinary construction industry problems. Despite these perceived opportunities, efforts to integrate computational thinking in CEM education are lacking. Most CEM students struggle with understanding the required computational concepts and workflows because they lack the theoretical foundations. Furthermore, there is a paucity of CEM programs currently providing these kinds of education to undergraduate students. Some of the programs offering related courses are either focused on hands-on implementation of data acquisition technologies or data processing with prepackaged software, which are only suitable for addressing specific applications and do not allow students to experience the entire pipeline of the problem-solving process. Other programs introduce students to the elements of computational concepts, such as variables, algorithms, functions, and methods, and use traditional programming languages (e.g., MATLAB, Python, and C# functions) to execute data analytics. Students still face the challenge of understanding the detailed programming syntax. Current learning resources for programming are not designed with non-computer science students in mind and may not be appropriate for solving problems in a specific domain (Wang et al. 2018). Therefore, it would take too much effort for instructors and students to reach the point where they can solve problems from practice.

EUP for Providing Data Analytics and Computation Thinking Skills. Providing a platform to empower non-computer science students with the ability to create the analysis of sensor data can stimulate computational thinking and learning within the context of construction. If properly designed, educational EUP environments can give students, with little to no prior programming experience, the skills necessary to rapidly invoke innovations with sensor data. Compared with general programming platforms, such as MATLAB, C++ or Python, EUP environments can provide students with means to carry out sophisticated analysis immediately by employing programming by demonstration (Lieberman 2001) and visual programming language (Burnett and Myers 2014), in place of the often tedious syntax that accompanies traditional programming mechanics. Through these simulations, students can formulate problems, logically organize, and analyze sensor data, represent data through abstractions such as variables and algorithms, and transfer the solution to solve a wide variety of problems. These are the key elements of computational thinking. Also, by employing graphical interactive elements, students are quickly introduced to the logic of programming: graphical object programming has the power to motivate students in the domain of computational modeling in construction (AbouRizk et al. 2011). With EUP environments, students are motivated to gain skills necessary to complete their problems, learn new skills, and attempt more sophisticated problems (Scaffidi et al. 2005). As students gain new expertise, they can address complex problems relating to high-level interactions between their domain and artificial intelligence. Students will be able to apply advanced technologies to their practice and to hold more informed communication with programmers and decision-makers for their future careers with enhanced understanding of the computational concepts (Chilana et al. 2016).

THEORETICAL FRAMEWORK

The proposed system is grounded in Learning-for-Use (LfU) theory. LfU is a technology design theory based on the following four tenets: "(1) knowledge construction is incremental; (2) learning is goal-directed; (3) knowledge is situated; and (4) procedural knowledge needs to support knowledge construction" (Edelson 2001). Tenets 1 and 4 asserts that there is incremental development of new knowledge and procedures when students' prior knowledge is tied to new knowledge. For example, students use foundational CEM and computing knowledge to apply sensor data analytics to address construction problems. By employing the structured process in Figure 1, the system enhances students' engagement in an incremental process. Students incrementally build knowledge by adding new concepts to memory, while making new connections between concepts. This new knowledge equips students to become proactive in their own learning and to construct solutions to construction problems requiring data analytics.

Tenets 2 and 3 asserts that knowledge acquisition is goal-directed and situated. Theorists Greeno (1998) and Lave and Wenger (1991) argue that knowledge should not be delivered in the abstract but in the context. The situative perspective views knowledge "as distributed among people, their environments and the communities of which they are a part" (Greeno and Engeström 2006), and learning is conceptualized as meaningful participation in a community of practice. The realization of gaps in one's knowledge, perhaps because of specific competency demands in the workplace or self-curiosity, can serve as a motivational goal for acquiring new knowledge. For example, the modules within the conceptual system will be developed based on formal competencies established in collaboration with construction industry practitioners. The system will be designed to encourage goal-directed tasks as students will create simulations to learn and develop computational thinking skills, which are needed for sensor data analytics careers in the construction industry. To determine how knowledge is acquired and how computation skills are developed, an epistemological perspective through a LfU framework is therefore appropriate.

PROPOSED SYSTEM

The proposed system, called SensDat, will provide a user experience that facilitates the process of problem-solving through which students can learn (1) the challenges involved in understanding sensor data collected from construction sites, (2) computational approaches to encode domain knowledge, and (3) how to discover meaningful information from low-level sensor data to facilitate decision making. SensDat will be designed to support the workflow of a problem-solving process in a data-driven fashion. The workflow of this process is depicted on the left side of Figure 1. In the existing problem-solving approach, students encounter various challenges, namely, the open-ended nature of a solution, technical barriers involved in manipulating inconsistent data, and the complexity of the domain knowledge. To address these challenges, SensDat will provide a structured process to help students learn how to translate industry stakeholders' requirements into a set of computational rules and extract meaningful information that can support the decision-making process from large volumes of raw sensor data. The modules within SensDat will be capable of addressing diverse construction industry problems with a set of commonly used and readily available sensors. SensDat will comprise three main components, shown in Figure 1 and described below:

- **Module Selection/Practice Set:** An interactive diagnostic with hypothetical datasets to help students select the right sensors and collect data from the field activity per problem.
- **Data Modeling by Demonstration:** A data modeling program which processes collected sensor data into a multimodal representation (e.g., vehicle trajectories, time-series graphs, and time-lapse videos).
- **Multimodal Query:** A query program in which students can specify algorithms to compute metrics of their choice in a multimodal demonstration (e.g., textual, and sketch).

System Architecture. SensDat is being built on a web browser, based on the Python programming language. Python was chosen because it has numerous online resources and is frequently used for data science with open-source Python libraries such as numPy, TensorFlow, or sciPy. Using Python can provide a smooth transition from SensDat to further complex data analytics tasks. SensDat will be designed to target novice users to immediately write code without having to configure/install any program's language. In that regard, the entire system will be built on a web browser. A client-side python interpreter (Mozilla's Pyodide, https://pyodide.org/) will be employed and the run-time state will be maintained in a client machine, a web server serving as a simple static file server. The graphical user interface (or GUI) will be built using HTML/CSS/JavaScript. SensDat will employ model-view-controller pattern where the program state is available in the python environment, and the front-end sends code text to the environment and receives the run-time state. SensDat will implement a python package that contains custom objects (e.g., vehicles, workers) and a library of helper functions. An intermediate layer will interact with the GUI as well as the Python run-time state. The program will be hosted in a web server and host a brief database for authentication (sign-up/in) and storing the code/raw sensor data. The backend will be implemented in MongoDB, Express.js, and Node.js. The architecture of the program is shown in Figure 2.

ENVISIONED SCENARIO

This section introduces a hypothetical scenario in which a group of students applies SensDat to a construction problem. Sasha, Chris, and Victoria are undergraduate students who signed up for a course to gain the knowledge and skills for applying sensor data analytics in construction. The instructor introduced each team in the class to an industry partner, a project manager, Paul, responsible for some local construction sites, and asked them to collaborate with Paul to solve a real-world problem experienced at the construction site. The team met Paul at one of his nearby sites and had a kick-off meeting to learn more about the problems he is interested in solving.

Paul told the students that he needs a means of evaluating the safety of his site layout plans; that is, he wants to understand the safety risks associated with the arrangement of resources at the site. A good site layout boosts the effectiveness and efficiency of the construction work taking place, reduces costs and material travel distances, and improves the safety of the construction site. Paul explained that when site managers face different site layout scenarios, there are no safety risk assessment models to help them make decisions.

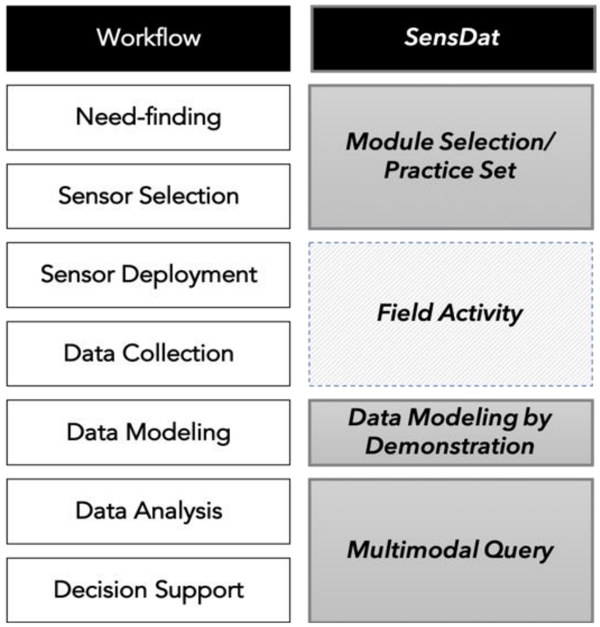

Figure 1. Overview of SensDat.

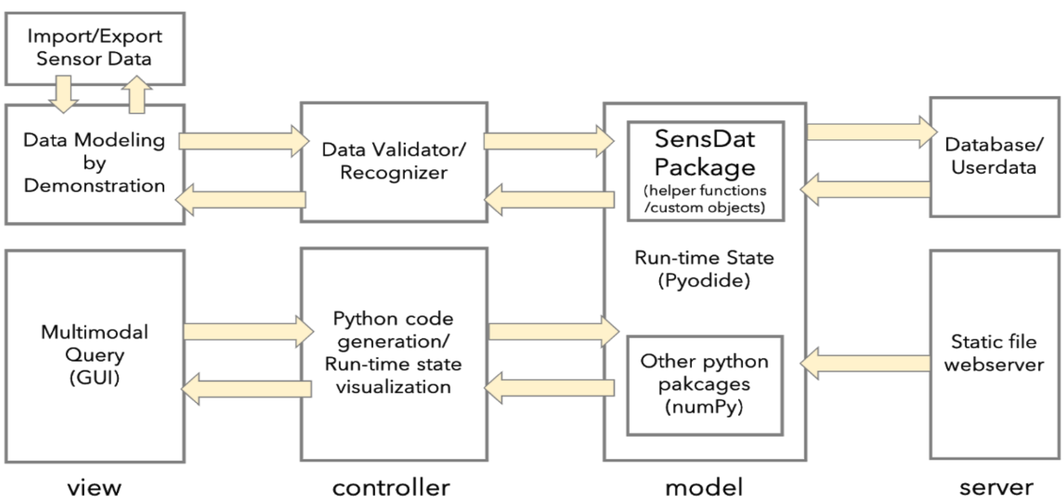

Figure 2. Architecture of SensDat.

To address this problem, he wanted to track the flow of materials, personnel, and equipment (such as excavators, bulldozers, loaders, forklifts, and dump trucks) between facilities and around the site. By tracking these flows, he wanted to identify high worker and equipment activity locations, potential near-misses and close-calls between these resources, the time and frequency of such events, and the resources involved. He told the students that accidents could be very costly for contractors beyond the safety of their workers, in terms of their impact on insurance premiums and the contractor's reputation; and can also have social implications on employees and the society's perception of the profession.

Module Selection. After the meeting, the team set out to address the problem using SensDat. One of the modules of SensDat enables students to evaluate construction sites based on the recorded trajectories of heavy equipment and workers. The module has a tutorial that illustrates the process of computing near-misses and close calls based on the locations of agents (viz., vehicles, and workers) in the model. Using the module, Sasha learned how to generate an occupancy grid for the site, compute the Euclidean distance between two points at a given time, and specify a time window to define the near-misses; for example, a forklift and a worker were within five meters of each other at any given time. The tutorial ends with an interactive guide in which the application asks questions about the nature of the construction site and resources to be tracked before offering guidance on the most suitable sensor. Questions focus on such topics as whether an agent will be performing work indoors or outdoors. When Sasha responded 'indoors' to this question, the SensDat module recommended using an RFID or UWB real-time location sensing system rather than global positioning system tags: while all three works for measuring location, only two are suitable for use indoors. The module also informed the team of the types of data they would need to compute near-misses, such as resource ID, timestamp, and x and y coordinates.

Data modeling by Demonstration. The team went to the construction site and deployed UWB sensors on two workers and one loader. SensDat provided guidelines for installing the sensors to preclude or address potential data discrepancies (e.g., inconsistent timestamps and battery issues). After collecting data for a week, the team obtained three different log files, each containing a week of sensor data for worker 1, worker 2, and loader 1, respectively. Chris ran the SensDat Data Modeling program. The program first reminded him of the kinds of data required by the module to evaluate near-misses. As Chris knew that he needed four different types of data for the module, he wanted to make sure that the raw data included the values that SensDat requires. Chris imported the data, and the program began scanning the log files. The program showed a few sample lines from the log file and asked Chris interactive questions regarding the locations of the required data, such as "Please highlight the column corresponding to the x coordinate of the vehicle." Chris understood how and in what format the sensor logged data and was thus able to comprehend how the raw data represent the trajectory of a vehicle. After he specified the x coordinate column, he indicated the locations of the rest of the data (y coordinate, timestamps, and resource ID) one by one, in response to the interactive questions. Then, the SensDat Data Modeling program found some lines in the log files that did not match the pattern, prompting Chris to specify the data for the newly found pattern. Chris realized that UWB sensors must sometimes estimate coordinates as a fallback solution, producing different values in certain columns and resulting in a new pattern. Still, he was able to recognize data that corresponded to the required coordinates and highlighted the correct region. The log files were then imported completely.

Multimodal Query. Victoria was to analyze the sensor data. A SensDat module visualized the trajectory of three objects in paths on a plane using the imported data. Sasha and Victoria needed to specify the spatial context of the construction site as contextual information. For example, they drew the locations of staging areas, toilets, and workstations on the map, as if creating presentation slides. Clearly, the trajectory of worker 1 and loader 1 had intersected each other, based on the visualization of the trajectories. Sasha needed to present how frequently near-misses occurred during the data collection period. She clicked on two icons representing worker 1 and loader 1, and the edge between the two objects gave a list of functions for analyzing relationships between two objects (See Figure 3). She chose *"distance at time"*, which calculates the Euclidean distance between two objects. She dragged the counter object and revised the expression to counter *(distanceAtTime (worker1, loader1) < 5),* giving the team the number of cases where two objects were within five meters of each other at any given time. Victoria was able to spot those two incidents in the trajectories and started to create a report describing the potential safety hazards on the job site. Sasha wondered if she could also visualize the locations associated with these risks and used another object, identified by the expression heatmap *(distanceAtAnyTime (worker1, loader1) < 5)*, which highlighted regions in which two objects intersected each other. Both realized how sensor data could provide metrics that can support decisions and were surprised how quickly they were able to write a program to analyze their raw data just by demonstration.

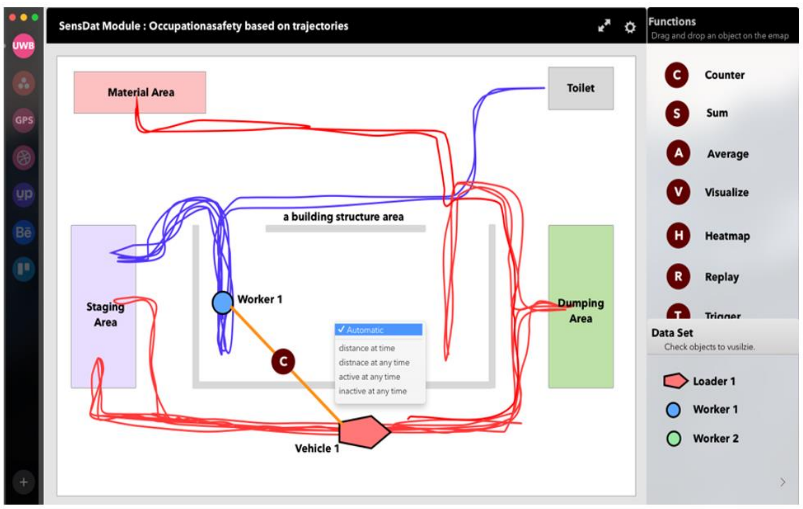

Figure 3. An example sketch of a SensDat EUP environment.

CONCLUSIONS

Large volumes of data are generated on construction sites and throughout the lifecycle of projects. However, the scarcity of CEM graduates with expertise in data analytics and computational thinking to gain insight from the data has not been plausible. Some industries, e.g., manufacturing, and healthcare have successfully utilized insights from data to improve their processes and make informed decisions. The complexity of existing systems makes it rather unappealing for CEM students to learn these skills. With an EUP environment that abstracts the complexity of programming and computational theories, students can be trained to acquire computational and data analytics skills to discover insights from sensor data. The presented conceptual system converges at the creation of a visual, user-friendly platform that eliminates the rubrics of complex programming syntaxes. With the system, students can extract insights from project data and make forecasts to alleviate some of the key challenges faced by the construction industry.

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 2111003.

REFERENCES

AbouRizk, S., Halpin, D., Mohamed, Y., and Hermann, U. (2011). “Research in modeling and simulation for improving construction engineering operations.” *Journal of Construction Engineering and Management*, 137(10), 843-852.

BLS (Bureau of Labour and Statistics). (2019). “Employment by detailed occupation.” <https://www.bls.gov/emp/tables/emp-by-detailed-occupation.htm>. (June 11, 2021).

Burnett, M. M., and Myers, B. A. (2014). “Future of end-user software engineering: beyond the silos.” *Proceedings of the on Future of Software Engineering*, 201-211.

Chilana, P. K., Singh, R., and Guo, P. J. “Understanding conversational programmers: A perspective from the software industry.” *Proc., Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ACM, 1462-1472.

Clarke, T., Ayres, P., and Sweller, J. (2005). “The impact of sequencing and prior knowledge on learning mathematics through spreadsheet applications.” *Educational Technology Research and Development*, 53(3), 15-24.

Dixon, T., Connaughton, J., and Green, S. (2018). “Understanding and Shaping Sustainable Futures in the Built Environment to 2050.” *Sustainable Futures in the Built Environment to 2050: A Foresight Approach to Construction and Development*, 337.

Edelson, D. C. (2001). “Learning-for-use: A framework for the design of technology-supported inquiry activities.” *Journal of Research in Science teaching*, 38(3), 355-385.

Forum, W. E. *Shaping the Future of Construction: A Breakthrough in Mindset and Technology*. World Economic Forum Cologny, Switzerland.

Greeno, J. G. (1998). “The situativity of knowing, learning, and research.” *American psychologist*, 53(1), 5.

Greeno, J. G., and Engeström, Y. (2006). *Learning in activity*, Cambridge University Press, Cambridge, England.

Ko, A. J., Abraham, R., Beckwith, L., Blackwell, A., Burnett, M., Erwig, M., Scaffidi, C., Lawrance, J., Lieberman, H., and Myers, B. (2011). “The state of the art in end-user software engineering.” *ACM Computing Surveys (CSUR)*, 43(3), 21.

Lave, J., and Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*, Cambridge university press.

Letondal, C. (2006). “Participatory programming: Developing programmable bioinformatics tools for end-users.” *End user development*, Springer, 207-242.

Lieberman, H. (2001). *Your wish is my command: Programming by example*, Morgan Kaufmann.

Omar, T., and Nehdi, M. L. (2016). “Data acquisition technologies for construction progress tracking.” *Automation in Construction*, 70, 143-155.

Scaffidi, C., Shaw, M., and Myers, B. “Estimating the numbers of end users and end user programmers.” *Proc., 2005 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC'05)*, IEEE, 207-214.

Shyamala, C., Velayutham, C. S., and Parameswaran, L. “Teaching computational thinking to entry-level undergraduate engineering students at amrita university.” *Proc., 2017 IEEE Global Engineering Education Conference (EDUCON)*, IEEE, 1731-1734.

Wang, A. Y., Mitts, R., Guo, P. J., and Chilana, P. K. “Mismatch of expectations: How modern learning resources fail conversational programmers.” *Proc., Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1-13.