Supporting Theory Building in Design-Based Research through Large Scale Data-Based Models

Golnaz Arastoopour Irgens
Clemson University
garasto@clemson.edu

Ibrahim Oluwajoba Adisa
Clemson University
iadisa@clemson.edu

Deepika Sistla
Clemson University
dsistla@clemson.edu

Tolulope Famaye
Clemson University
tfamaye@clemson.edu

Cinamon Bailey
Clemson University
cinamob@clemson.edu

Atefeh Behboudi
Clemson University
abehbou@clemson.edu

Adenike Omalara Adefisayo
Clemson University
aadesfis@clemson.edu

ABSTRACT

Although the fields of educational data mining and learning analytics have grown significantly in terms of analytical sophistication and the breadth of applications, the impact on theory-building has been limited. To move these fields forward, studies should not only be driven by learning theories, but should also use analytics to inform and enrich theories. In this paper, we present an approach for integrating educational data mining models with design-based research approaches to promote theory-building that is informed by data-based models. This approach aligns theory, design of the learning environment, data collection, and analytic methods through iterations that focus on the refinement and improvement of these components. We provide an example from our own work: the design and development of a digital learning environment for elementary-school (ages 8 to 13) children to learn about artificial intelligence within sociopolitical contexts. The project is driven by a critical constructionist learning framework and uses epistemic network analysis as a tool for modeling learning. We conclude with how this approach can be reciprocally beneficial in that educational data miners can use their models to inform theory and learning scientists can augment their theory-building practices through big data models.

Keywords

Learning theory, theory building, design-based research, design framework, epistemic network analysis, constructionism

1. INTRODUCTION

The development of theory is needed in any scientific field. In educational research, framing studies through theories contributes to a systematic understanding of teaching, learning, and the design of learning environments [9]. Specifically in the fields of educational data mining (EDM) and learning analytics (LA), aligning studies with theory reduces the risk of finding results based on chance and

allows for the exploration of more nuanced metrics of learning as opposed to time-on-task or simple counts of activities [44].

More recently, the focus has shifted to not just grounding analytics in theory but using analytics to *develop or refine theory* [14]. The abundance of sophisticated technical approaches in EDM and LA have great potential for building and improving upon existing learning theory in a fast growing and changing technical world. Moreover, theory building and theory refining drive the growth of scientific knowledge [34] and help establish emerging fields like EDM and LA.

However, the impact of EDM and LA research on practice, theory, and frameworks has been limited. Dawson and colleagues [15] found that few studies suggested actions to interrogate data-based models and provide feedback to theory. The authors argue that researchers must extend beyond individual analysis and provide rigor in evaluating assumptions, interventions and actions to refine established or generate new theories of learning. Similarly, Romero and Ventura [37] argue for increasing the impact on practice and to move from exploratory models to more holistic and integrative systems-level research through developing theory and conceptual frameworks. Thus, there is a need for EDM, LA, and related educational research communities to explore approaches that allow for theory-building to move the field forward and have broader impact for learners and teachers.

One approach for theory-building in the learning sciences is through design-based research. The goals of design-based research are to test theory in naturalistic settings by designing and implementing interventions and to generate new theory through cycles of testing and refinement [7]. In initial cycles, the theories may be piecemeal and applicable to local contexts. As researchers progress through the cycles, they will develop more broadly applicable theories. In addition, design-based research provides guidance for aligning theory, design, and practice to more tightly connect findings to practice and in turn, directly improve educational settings [16]. These characteristics of design-based research can support the need for EDM researchers to engage in theory grounding and building.

In this paper, we propose an approach for integrating EDM and design-based research to inform theory building. We provide an example from our own work, designing and developing a digital learning environment for elementary-school-aged children to learn

about artificial intelligence within sociopolitical contexts. This project is grounded in a critical constructionist learning framework, and we used epistemic network analysis as a tool for modeling learning data collected from this environment.

2. BACKGROUND

2.1 Learning Theories

In learning analytics and the learning sciences, there is no clear definition of what constitutes a learning theory [29]. For example, Doroudi [18] identified four core learning theories: behaviorism, cognitivism, constructivism, and situativism/sociocultural. Doroudi argues that such theories can be viewed as worldviews for how to think about learning. These worldviews, or lenses, are developed through researcher's collection of scientific evidence, their philosophical ideas about learning, and their values regarding ways of learning, doing, and being. However, within these four broadly applicable big "T" Theories exist small "t" theories. For example, sociocultural theory views learning as inherently social and situated within a particular context. Within sociocultural theory, there are several narrower lenses, such as distributed cognition, embodied cognition, and communities of practice, which each focus on a particular aspect of social and situated learning. Researchers can synthesize several small "t" theories into a conceptual framework used to guide research questions, methods, and design. Often, the conceptual framework is represented in a visual diagram that depicts the relationships among theories and ideas in the literature [31]. Conceptual frameworks change over time as research studies are conducted and scientific evidence (re)informs theory.

In EDM and LA large-scale datasets are analyzed to make claims about learning. For example, clickstream data can be analyzed to make predictions about student performance. However, as Kitto and colleagues [30] note, clickstream data is typically collected to aid developers in debugging software and not well-structured for describing educational phenomena. They argue that some studies that investigate clickstream data can lead to insightful findings; however, these studies often yield results that are already well understood or difficult to interpret and test. Referred to as the "clicks to constructs" problem, researchers may rely on existing low-level descriptive data on learning platforms to make meaningful claims about learning but are difficult to link to theoretically grounded concepts [12].

While some EDM and LA studies have grounded their work in learning theories, it is still not a common practice in EDM research. According to a recent literature review by Baek and Doleck [6], 66% of published EDM research did not identify a theoretical orientation. Of those studies that identified a theory, about half used the same theory: self-regulated learning. This review underscores the need for processes that guide EDM researchers towards grounding their studies in wider varieties of learning sciences theories to contribute to advancing learning theory.

Consequently, there have been calls for EDM and LA researchers to embed educational theory into advanced analytics tools and methods [43, 44]. One way to achieve this is to design theory-driven studies that allow for driving actionable change in learning environments and ecosystems [43]. In large datasets with almost endless models and significant correlations, theoretical and conceptual frameworks justify variables to choose in a model, how to interpret results, and, importantly, how to guide stakeholders to make decisions about education and learning [30].

2.2 Design-Based Research

Theory building in the learning sciences goes hand in hand with design-based research, which is a series of approaches for producing theory, artifacts, and practices that impact teaching and learning. In such studies, researchers systematically build and test new designs in learning settings and adjust the design context through cycles of implementations [11, 42]. Design-based research aims to develop theory-based learning interventions and generate new theories of learning [7]. Using this theory-based approach, learning theories are developed through reflective cycles of experimentation within the context in which they will be used [24]. Such cycles of design involve design conjectures that claim if learners engage in this certain activity, then a particular mediating process will emerge that produces desired learning outcomes [38]. The outcomes then inform novel design processes or principles [20], new hypotheses [11], or the development of new small "t" theories or ontological innovations [17] which are explanatory constructs, categories, or taxonomies that explain how things work.

In design-based research, the outcomes of a learning environment that drive theory development are informed by data collection and analysis. In digital learning environments that collect and track multiple data streams, the datasets are large and require data analytic tools for building and interpreting models. This type of analysis is common in EDM and LA, and thus, such approaches are an apt fit for design-based research outcome analyses. However, few studies in EDM and LA have explored theory building through design-based research [15] although scholars have argued for this integration of fields to advance learning theory and the fields of EDM and LA [36]. A few recent studies have integrated EDM and design-based research for the purposes of improving the designed intervention itself or for developing design principles [28, 43]. For example, Quigley and colleagues [35] used sequencing and Naïve Bayes classifiers to model high school biology students' science practices. The findings informed which scientific practices teachers should focus on in their instruction and use of the intervention and learning tools. However, studies that integrate EDM approaches and design-based research for the purposes of theory building are rare [15].

In this paper, we propose a process framework for how design-based research can be integrated with large data-based models to inform learning theory development.

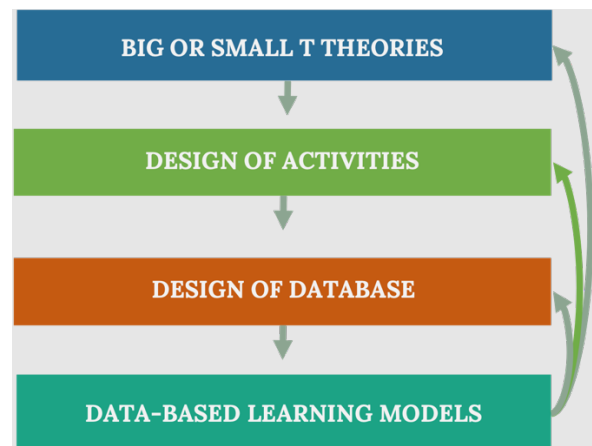## 3. PROPOSED DATA-BASED MODEL DESIGN PROCESS FRAMEWORK

Figure 1. Proposed model-based design process framework

In our proposed framework (figure 1), the process begins with a learning theory that guides the remainder of the study. This theory is the lens through which the design of the learning environment activities and tools will be designed. Through the theory's lens, the activities' design drives the design of database and choices for data collection. After the implementation of the digital learning environment, the large amount of data collected is modeled and analyzed to answer research questions. These large data-based models then inform the next iteration of database design and activities and reflection on advancing learning theory.

4. APPLIED EXAMPLE: SPOT DIGITAL LEARNING ENVIRONMENT

To concretize the proposed framework, we offer an example from our recent research project: the SPOT digital learning environment. The goal of the project is to develop theory-based learning activities for elementary-school-aged children to engage in computational thinking practices [3, 10, 23] by developing machine learning image classification algorithms and programing robots to respond to these algorithms. Children engage in these computational thinking practices through a critical lens [22]: interrogating, dismantling, and reimaging technologies that oppress marginalized populations. This work is motivated by recent abundance of artificial intelligence applications, concerns for biased and harmful machine learning algorithms [13, 32], and calls for an early education around computational thinking and computer science for children [46]. The overarching research question in this project asks how children develop computational practices through a critical lens? The long-term goals are to inform the design of critical computational thinking learning environments and to develop learning theory at the intersection of critical pedagogies and computational thinking for elementary-school-aged children.

4.1 Cycle 1

The big "T" Theory framing this project is sociocultural theory. Within this worldview, we draw on constructionism [33], critical pedagogies [22], and rely on a newly proposed small "t" theory of critical computational thinking [26, 27] and critical constructionism [25]. Based on our theoretical orientation, our assumptions and values around learning are: 1) construction of and reflection with an object facilitates children's meaning-making and 2) learning is inherently political and children should engage in questioning, dismantling, and reimagining systemic oppression.

Through this lens, in the first iteration of this research project, we designed activities that allowed children to engage in machine learning practices and reflect on their learning through a sociopolitical lens that involve questioning machine learning outputs and creating interest-based machine learning objects. [2, 21]. One key activity was developing a "robot story" to draw and write a story about a robot for social good. Children were encouraged to create fantastical stories (Figure 2). Then, with the assistance of adults, children used 1) their computer webcam to train an image classifier to recognize classes of images (Google Teachable Machine), and 2) a block-based programming language (Scratch) to train a robot prototype inspired by their stories to respond to their classification system, and 3) investigate and mitigate bias in their training datasets such that groups of users were not excluded or discriminated against. A full description of activities are available in [2, 5, 21]
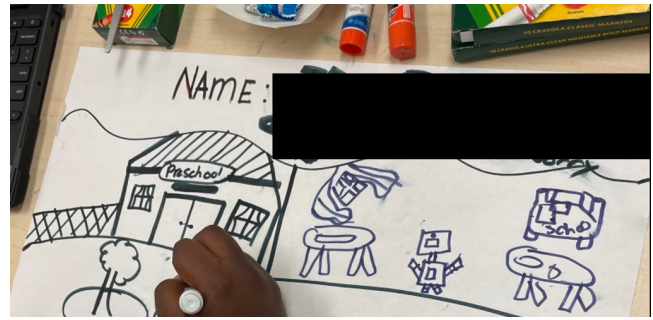

Figure 2. One example of a robot story drawing of a robot helping children learn their colors. An adult and three children programming a robot together.

In this example, the theoretical orientation is *critical computational thinking and constructionism*, which is still a developing orientation. This small t "theory" drove the design of the activities that focused on creating objects for learning computational thinking practices in a sociopolitical context.

The data collected in this first cycle consisted of 44 children's block-based programming code files, children's machine learning classification algorithms, recorded video of activities, and other written or drawn artifacts from children. The database choices were driven by identifying the forms of evidence for learning that needed to be extracted within the scope of the designed activities. At times, we reconfigured the structure of the activity to better collect evidence of children's learning and engagement. For example, we asked children at specific timepoints during the programming activities to deposit their code and projects into shared folders for data collection purposes.

This database then informed the type of analytic models we constructed to answer our research questions about children's critical computational practices and the design of a critical machine learning education program. In this first iteration, we relied on *epistemic network analysis* (ENA) to create data-based models from children's artifacts and discourse data. Before building the model, the data must be organized in rows and columns in a format that is machine readable by the ENA webtool. The data are segmented into meaningful divisions of discourse [4]. For example, in our study, each row in the data represented one student, one column represented their discourse or activity at a point in time, and each column was a meta-data variable. After segmentation, each line of discourse or activity is codified numerically. In our studies, we added columns to the dataset for each code that represent children's knowledge of critical computational thinking practices. The data were coded for three categories: how humans develop machine learning applications (5 codes), harmful machine learning applications (3 codes), and helpful machine learning applications (2 codes)

(full coding description available in cml paper). From here, we used ENA to build network models of children's discourse and activity.

ENA measures the connections between discourse elements, or codes, by quantifying the *co-occurrence* of those elements within a defined segmentation [39–41]. In our studies, co-occurrences of codes were calculated if they occurred within a child's response. Each child's co-occurrences for each of the nine activities were totaled and each activity was visualized as a single weighted node-link network representation. Each network represented a summation of all the children's co-occurrences within an activity.

To analyze several networks at one time, we used an alternative ENA representation in which the centroid (center of mass) of each network was calculated and plotted in a fixed two-dimensional space that was mathematically created by conducting a multi-dimensional scaling routine (a singular value decomposition) and a sphere normalization. The space is interpreted by examining the location of the nodes in the two-dimensional space and evaluating the goodness of fit. In this analysis, the Spearman goodness of fit was 0 for both the x and y axis and the Pearson goodness of fit was 1.0 for both the x and y axis, indicating the location placement of the nodes was reliable. For more detailed mathematical explanations of ENA, see work by Bowman and colleagues [8], authors and colleagues [3], and Shaffer and Ruis [41].
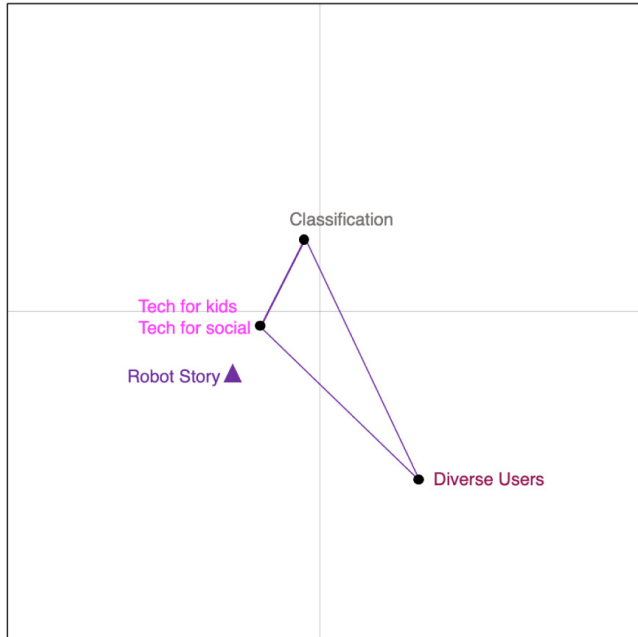

Figure 3. Epistemic network visualization of all children's discourse for Robot Superhero Story activity. Triangle represents the center of mass of the weighted network and is plotted in a two-dimensional mathematical space. Node labels for Helpful ML Applications are pink, for Harmful ML Applications codes are red, and for How People Develop ML Applications are grey. Figure from [5].

Figure 3 displays one of the network visualizations that was created. This network represents the data collected and modeled from children's robot stories. In this activity, children connected among TECHNOLOGY FOR KIDS, TECHNOLOGY TO ADDRESS SOCIAL ISSUES, CLASSIFICATION ALGORITHMS, and DIVERSE USERS suggesting an understanding of how machine learning classification algorithms can be used to design technology for kids, for social good, and for a diverse range of users. These connections reflect the goal of the designed activity for children to imagine and design their own machine learning based robot that could benefit a potentially marginalized population. For example, LaToya, an African American child, designed a robot that relied on a color classification algorithm that would identify the colors of real-world objects and teach colors to young children. She was inspired by her younger cousin's lack of access to learning tools. The remainder of the ENA models and interpretations can be seen in authors and colleagues' publication [5].

The collection of ENA networks in our models advanced characterizations of the small t "theory" of *critical computational thinking* by showing varies engagements of children's application of a critical lens to designing and discussing machine learning technologies. The models showed evidence of youth creating products using existing tools and designing more equitable and socially beneficial alternatives. Based on these models, we proposed a framework of critical machine learning education that involves 1) posing and answering questions about the roles of producers and consumers of AI technologies, specifically who designs technologies, for what purposes, who benefits, who is harmed, and what are the histories embedded in the data being used, and 2) identifying how people design AI technologies and applying this knowledge to build applications for marginalized populations. This is a first step towards advancing ideas around critical computational thinking.

Our models and supplementary observational data collection suggested that children engaged deeply in activities when provided with a narrative and allowed to create their own stories and technologies. Our models also suggested that children intermittently took a critical lens to their work, but *they did not fully integrate a critical perspective into their investigations*. This missing link was problematic because our main goal was for children to integrate a critical lens into their computational thinking practices throughout the program.

We hypothesized that this lack of an integrated critical lens was because some activities included computational thinking practices without focusing on sociopolitical issues, while other activities involved discussion of sociopolitical issues without integrating computational thinking practices. Moreover, students were not invited to explore oppressive histories of marginalized populations, which is important for situating students' understanding of *systemic* oppression. Finally, we realized that we adopted existing programming tools that were not *specifically* designed for students to explore socio-political issues in AI.

Thus, in this first design-based iteration cycle, our data-based ENA models informed the characterization of a newly developing theoretical orientation of critical computational thinking, the refinement of our programmatic activities to integrate sociopolitical issues and computational thinking practices within every activity, to include more narrative-based activities to increase engagement, and to develop our own programming tools that center critical lenses in computing.

4.2 Cycle 2
In the second iteration of this research project, we redesigned the activities to be in the form of a digital game driven by a cohesive narrative while integrating a sociopolitical and constructionism lens within every activity [1]. We hypothesized that through gameplay and immersion, students will actively construct their own understanding of the world and gain computational thinking skills.

In addition, we involved children from the after-school centers in the redesign of the activities based on our findings from our models.

Using cooperative inquiry techniques [19, 45] we provided opportunities for children to gather in groups and generate ideas for the game design and activities. One re-occurring theme was the desire to customize their space and objects. For example, children wanted to customize other characters' appearance and capabilities. In another example, one child expressed their desire to have a personal space to "hang up and look at" earned badges and wanted to decorate this space.

In our research lab, we ideated, designed, and developed a hybrid physical/digital role-playing game in which students role-play as agents-in-training for a top-secret agency called Solving Problems of Tomorrow (S.P.O.T). In the game, children play the role of the protagonist as they work as secret agents for the S.P.O.T. agency. They undertake the task of traveling into the future to investigate and solve technology-related problems that they encounter and then bring back the knowledge and skills they acquire to solve present problems. To sign up for their mission, children receive signup instructions from their classroom teacher, who role-plays as a senior S.P.O.T agent providing the child agent with the resources they will need as they proceed to login and signup as a young S.P.O.T agent.

The game experience is divided into progressive levels, and child agents must complete a series of activities to earn badges and powerups that take them to the next level. When they travel to the future, the agents discover that the main mode of transportation are self-driving cars that use facial recognition for access. In this futuristic scenario, children are having trouble accessing the cars because the facial recognition technology does not consistently classify children's faces correctly. The purpose of this activity is for children to learn about machine learning classification algorithms while at the same time exploring algorithm bias and the consequences on vulnerable populations, such as children. We chose children as the initial vulnerable population because this is a shared identity of all the game players. The frontend of the web-based game is developed using the Svelte framework which compiles components into efficient JavaScript and HTML at the same time x(figure 4).

Figure 4. Svelte interface for the game-based environment with typing and speech-to-text options for responses.

Finally, we developed an adapted version of block-based programming language (inspired by Scratch) and a user interface for an image classification trainer using tensorflow.js that is developmentally appropriate for children (inspired by Google Teachable Machine). Users will be able to train an image classifier, import the trained machine learning algorithm into the block-based programming tool, and program a robot to respond to the classification algorithm (Figure 5).
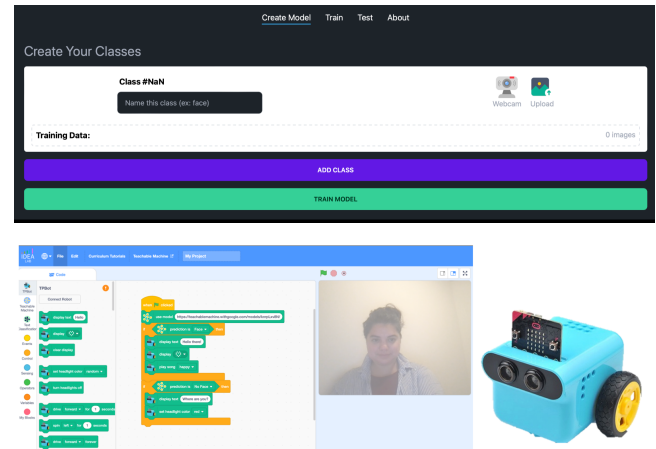

Figure 5. Image learning classification model using tensorflow.js and adapted version of Scratch block-based programming with custom blocks to program a TPBot robot with a micro:bit processor to respond to the classification algorithm through the computer webcam.

These designed activities influenced our data collection decisions. Data collection in this iteration include students' personal and demographic data, timestamped preliminary information about their existing knowledge and perception about AI concepts and technologies, timestamped text-based reflections and answers to question prompts, students' digital sketches using a drawing application, short video recordings asking children to reflect on their work, drag and drop icon responses, decorative customization choices, and saved files from their classification algorithms and block-based programming.

In turn, the design of the database was influenced by the activities and choices for data collection. The application uses NodeJS for backend operations, complemented by MongoDB and Amazon web services (AWS) cloud for data storage. MongoDB stores data in collections comprising documents that are analogous to rows in conventional databases. Text data are posted and stored directly to the MongoDB database while multimedia data like videos, audios, and visuals, are stored in an Amazon Web Service (AWS) S3 bucket through an API that allows us to send media data to the S3 bucket and returns a link that is eventually stored and linked to each user in the MongoDB database. The data in the MongoDB database and the S3 bucket are securely saved and can be virtually accessed by researchers for data access and analysis.

Currently, this project is ongoing, and the implementation of the game and the model creation has not been completed. S.P.O.T. will be implemented in classrooms in 2025 and ENA models will be created and analyzed to further inform theory building, design activities, and design of the database.

Figure 6 provides a summary of the model-based design process framework applied to the S.P.O.T. digital learning environment research project. In short, in each design-based research cycle, our

worldview is framed by the small "t" theories of critical constructionism and critical computational thinking. This worldview guides the design of the activities such that learners are designing and reflecting on computational objects in the context of oppressive power structures and equity. These activities influence the data collection choices and design of the digital database. Finally, the student discourse and activity data are analyzed by coding data and then building and interpreting ENA models. These data-based models further help us to refine our theoretical orientations by providing evidence that align with or contradict tenets of the theories.
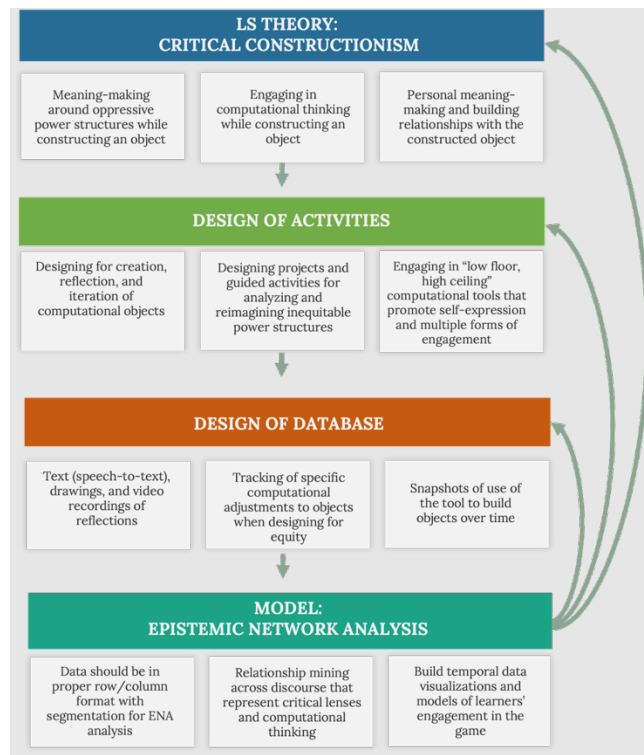

Figure 6. Data-based model design process framework applied to the SPOT digital learning environment research project

5. CONCLUSION

With the recent advancements of sophisticated tools and analyses in EDM and LA, theory matters more than ever [44]. Theories elucidate the mechanisms behind student learning and are instrumental in shaping the design of digital learning environments and interpreting the data derived from these settings [29, 43]. The EDM and LA fields are making strides towards theory driven studies. However, using EDM models to inform and build theory is still rare. Thus, in this paper we propose a model-based design process framework to guide theory grounding and building in educational research.

Our approach emphasizes the importance of anchoring data collection and analytics in a broad range of educational theories or frameworks—what we refer to as small "t" theories. This foundational perspective enables us to gather comprehensive multimodal data essential for discourse ENA models, moving beyond the reliance on superficial metrics such as clickstream data.

By embedding our analytics within a theoretical framework, we capture richer data and gain insights that prompt us to revisit and refine our initial educational theories and the design of learning activities. The discrepancies and alignments between our theoretical assumptions and the empirical evidence from the data inform this reflective process. Such an iterative interaction between theory and data analytics fosters continuous improvement, enhancing our understanding of learning behaviors and our methodologies in digital environments.

Such iterative data-based model design frameworks that inform data collection and analytics through theory and at the same time inform theory through analytics is crucial for moving the field of EDM and the learning sciences forward. Design-based research that relies on large data-based models can be reciprocally beneficial in that educational data miners can use their models to inform theory and learning scientists can augment their theory-building practices through big data models.

6. ACKNOWLEDGEMENTS

7. REFERENCES

[1] Adisa, I.O., Thompson, I., Famaye, T., Sistla, D., Bailey, C.S., Mulholland, K., Fecher, A., Lancaster, C.M. and Arastoopour Irgens, G. 2023. S.P.O.T: A Game-Based Application for Fostering Critical Machine Learning Literacy Among Children. *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference* (Chicago IL USA, Jun. 2023), 507–511.

[2] Arastoopour Irgens, G., Adisa, I.O., Bailey, C.S. and Vega Quesada, H. 2022. Designing with and for Youth: A Participatory Design Research Approach for Critical Machine Learning Education. *Education Technology & Society*. 25, 4 (2022), 126–141.

[3] Arastoopour Irgens, G., Dabholkar, S., Bain, C., Woods, P., Hall, K., Swanson, H., Horn, M. and Wilensky, U. 2020. Modeling and Measuring High School Students' Computational Thinking Practices in Science. *Journal of Science Education and Technology*. 29, 1 (Feb. 2020), 137–161. DOI:https://doi.org/10.1007/s10956-020-09811-1.

[4] Arastoopour Irgens, G. and Eagan, B. 2023. The Foundations and Fundamentals of Quantitative Ethnography. *Fourth International Conference of Quantitative Ethnography* (Copenhagen, Denmark, 2023).

[5] Arastoopour Irgens, G., Vega, H., Adisa, I. and Bailey, C. 2022. Characterizing children's conceptual knowledge and computational practices in a critical machine learning educational program. *International Journal of Child-Computer Interaction*. 34, (Dec. 2022), 100541. DOI:https://doi.org/10.1016/j.ijcci.2022.100541.

[6] Baek, C. and Doleck, T. 2023. Educational Data Mining versus Learning Analytics: A Review of Publications From 2015 to 2019. *Interactive Learning Environments*. 31, 6 (Aug. 2023), 3828–3850. DOI:https://doi.org/10.1080/10494820.2021.1943689.

[7] Barab, S. and Squire, K. 2004. Design-Based Research: Putting a Stake in the Ground. *Journal of the Learning Sciences*. 13, 1 (Jan. 2004), 1–14. DOI:https://doi.org/10.1207/s15327809jls1301_1.

[8] Bowman, D., Swiecki, Z., Cai, Z., Wang, Y., Eagan, B., Linderoth, J. and Shaffer, D.W. 2021. The mathematical

foundations of epistemic network analysis. *International Conference on Quantitative Ethnography* (2021), 91–105.

[9] Bransford, J.D., Brown, A.L. and Cocking, R.R. 2000. *How people learn*. Washington, DC: National academy press.

[10] Brennan, K. and Resnick, M. 2012. New frameworks for studying and assessing the development of computational thinking. (2012), 25.

[11] Brown, A.L. 1992. Design Experiments: Theoretical and Methodological Challenges in Creating Complex Interventions in Classroom Settings. *The Journal of the Learning Sciences*. 2, 2 (1992), 141–178.

[12] Buckingham Shum, S. and Crick, R.D. 2016. Learning Analytics for 21st Century Competencies. *Journal of Learning Analytics*. 3, 2 (2016), 6–21.

[13] Buolamwini, J. and Gebru, T. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Conference on Fairness, Accountability and Transparency* (Jan. 2018), 77–91.

[14] Chen, B. 2015. From Theory Use to Theory Building in Learning Analytics: A Commentary on "Learning Analytics to Support Teachers during Synchronous CSCL." *Journal of Learning Analytics*. 2, 2 (Dec. 2015), 163–168. DOI:https://doi.org/10.18608/jla.2015.22.12.

[15] Dawson, S., Joksimovic, S., Poquet, O. and Siemens, G. 2019. Increasing the Impact of Learning Analytics. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge* (Tempe AZ USA, Mar. 2019), 446–455.

[16] Design Based Research Collective 2003. Design-based research: An emerging paradigm for educational inquiry. *Educational researcher*. 32, 1 (2003), 5–8.

[17] diSessa, A.A. and Cobb, P. 2004. Ontological Innovation and the Role of Theory in Design Experiments. *Journal of the Learning Sciences*. 13, 1 (Jan. 2004), 77–103. DOI:https://doi.org/10.1207/s15327809jls1301_4.

[18] Doroudi, S. 2021. *A Primer on Learning Theories*. EdArXiv.

[19] Druin, A. 2002. The Role of Children in the Design of New Technology. *Behaviour and Information Technology*. 21, 1 (2002), 38.

[20] Edelson, D.C. 2002. Design Research: What We Learn When We Engage in Design. *Journal of the Learning Sciences*. 11, 1 (Jan. 2002), 105–121. DOI:https://doi.org/10.1207/S15327809JLS1101_4.

[21] Famaye, T., Irgens, G.A. and Adisa, I. 2024. Shifting roles and slow research: children's roles in participatory co-design of critical machine learning activities and technologies. *Behaviour & Information Technology*. (Feb. 2024), 1–22. DOI:https://doi.org/10.1080/0144929X.2024.2313147.

[22] Freire, P. 1970. *Pedagogy of the oppressed*. Continuum.

[23] Grover, S. and Pea, R. 2013. Computational thinking in K–12: A review of the state of the field. *Educational researcher*. 42, 1 (2013), 38–43.

[24] Hoadley, C. and Campos, F.C. 2022. Design-based research: What it is and why it matters to studying online learning. *Educational Psychologist*. 57, 3 (Jul. 2022), 207–220. DOI:https://doi.org/10.1080/00461520.2022.2079128.

[25] Holbert, N., Dando, M. and Correa, I. 2020. Afrofuturism as critical constructionist design: building futures from the past and present. *Learning, Media and Technology*. 45, 4 (Oct. 2020), 328–344. DOI:https://doi.org/10.1080/17439884.2020.1754237.

[26] Kafai, Y., Proctor, C. and Lui, D. 2020. From theory bias to theory dialogue: embracing cognitive, situated, and critical framings of computational thinking in K-12 CS education. *ACM Inroads*. 11, 1 (Feb. 2020), 44–53. DOI:https://doi.org/10.1145/3381887.

[27] Kafai, Y., Searle, K., Martinez, C. and Brayboy, B. 2014. Ethnocomputing with electronic textiles: culturally responsive open design to broaden participation in computing in American indian youth and communities. *Proceedings of the 45th ACM technical symposium on Computer science education* (Atlanta Georgia USA, Mar. 2014), 241–246.

[28] Kaliisa, R. 2022. *Designing Learning Analytics Tools for Teachers with Teachers - A Design-Based Research Study in a Blended Higher Education Context*.

[29] Khalil, M., Prinsloo, P. and Slade, S. 2023. The use and application of learning theory in learning analytics: a scoping review. *Journal of Computing in Higher Education*. 35, 3 (Dec. 2023), 573–594. DOI:https://doi.org/10.1007/s12528-022-09340-3.

[30] Kitto, K., Hicks, B. and Buckingham Shum, S. 2023. Using causal models to bridge the divide between big data and educational theory. *British Journal of Educational Technology*. 54, 5 (Sep. 2023), 1095–1124. DOI:https://doi.org/10.1111/bjet.13321.

[31] Miles, M.B. and Huberman, A.M. 1994. *Qualitative data analysis: An expanded sourcebook, 2nd ed*. Sage Publications, Inc.

[32] Noble, S. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press.

[33] Papert, S. 1980. *Mindstorms: children, computers, and powerful ideas*.

[34] Popper, K. 1992. *Conjectures and Refutations: The Growth of Scientific Knowledge*. Routledge.

[35] Quigley, D., Ostwald, J. and Sumner, T. 2017. Scientific modeling: using learning analytics to examine student practices and classroom variation. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (Vancouver British Columbia Canada, Mar. 2017), 329–338.

[36] Reimann, P. 2016. Connecting learning analytics with learning research: the role of design-based research. *Learning: Research and Practice*. 2, 2 (Jul. 2016), 130–142. DOI:https://doi.org/10.1080/23735082.2016.1210198.

[37] Romero, C. and Ventura, S. 2020. Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery*. 10, 3 (May 2020), e1355. DOI:https://doi.org/10.1002/widm.1355.

[38] Sandoval, W. 2014. Conjecture Mapping: An Approach to Systematic Educational Design Research. *Journal of the Learning Sciences*. 23, 1 (Jan. 2014), 18–36. DOI:https://doi.org/10.1080/10508406.2013.778204.

[39] Shaffer, D.W. 2018. Epistemic Network Analysis. *International Handbook of the Learning Sciences*. (2018), 520.

[40] Shaffer, D.W., Collier, W. and Ruis, A.R. 2016. A Tutorial on Epistemic Network Analysis: Analyzing the Structure of Connections in Cognitive, Social, and Interaction Data. *Journal of Learning Analytics*. 3, 3 (2016), 9–45.

[41] Shaffer, D.W. and Ruis, A.R. 2017. Epistemic network analysis: A worked example of theory-based learning analytics. *Handbook of learning analytics*. (2017).

[42] The Design-Based Research Collective 2003. Design-Based Research: An Emerging Paradigm for Educational Inquiry. *Educational Researcher*. 32, 1 (Jan. 2003), 5–8. DOI:https://doi.org/10.3102/0013189X032001005.

[43] Wiley, K.J., Dimitriadis, Y., Bradford, A. and Linn, M.C. 2020. From theory to action: developing and evaluating

learning analytics for learning design. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge* (Frankfurt Germany, Mar. 2020), 569–578.

[44] Wise, A.F. and Shaffer, D.W. 2015. Why Theory Matters More than Ever in the Age of Big Data. *Journal of Learning Analytics*. 2, 2 (Dec. 2015), 5–13. DOI:https://doi.org/10.18608/jla.2015.22.2.

[45] Yip, J., Clegg, T., Bonsignore, E., Gelderblom, H., Rhodes, E. and Druin, A. 2013. Brownies or bags-of-stuff?: domain expertise in cooperative inquiry with children. *Proceedings of the 12th International Conference on Interaction Design and Children* (New York New York USA, Jun. 2013), 201–210.

[46] 2010. *Report of a Workshop on The Scope and Nature of Computational Thinking*. National Academies Press.