



OPEN ACCESS

EDITED BY
Peter Robin Childs,
Imperial College London,
United Kingdom

REVIEWED BY
Rohan Prabhu,
Lafayette College, United States
Qiunan Meng,
Dalian University of Technology, China
Liang Guo,
Southwest Petroleum University, China

*CORRESPONDENCE
Corey Schimpf,
schimpf2@buffalo.edu

SPECIALTY SECTION
This article was submitted to Software
Technologies,
a section of the journal
Frontiers in Manufacturing Technology

RECEIVED 17 June 2022
ACCEPTED 06 October 2022
PUBLISHED 20 October 2022

CITATION
Schimpf C and Goldstein MH (2022),
Large data for design research: An
educational technology framework for
studying design activity using a big
data approach.
Front. Manuf. Technol. 2:971410.
doi: 10.3389/fmtec.2022.971410

COPYRIGHT
© 2022 Schimpf and Goldstein. This is
an open-access article distributed
under the terms of the [Creative
Commons Attribution License \(CC BY\)](#).
The use, distribution or reproduction in
other forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which does
not comply with these terms.

Large data for design research: An educational technology framework for studying design activity using a big data approach

Corey Schimpf^{1*} and Molly H. Goldstein²

¹Department of Engineering Education, SUNY, University at Buffalo, Buffalo, NY, United States,

²Industrial and Enterprise Systems Engineering, University of Illinois Urbana-Champaign, Champaign, IL, United States

The complexity of design problems compels the collection of rich process data to understand designers. While some methods exist for capturing detailed process data (e.g., protocol studies), design research focused on design activities still faces challenges, including the scalability of these methods and technology transformations in industry that require new training. This work proposes the Large Data for Design Research (LaDDR) framework, which seeks to integrate big data properties into platforms dedicated to studying design practice and design learning to offer a new approach for capturing process data. This technological framework has three design principles for transforming design platforms: broad simulation scope, unobtrusive logging and support for creation and analysis actions. The case is made that LaDDR platforms will lead to three affordances for research and education: capturing design activities, context setting and operationalization, and research design scalability. Big data and design expertise are reviewed to show how this approach builds on past work. Next, the framework and affordances are presented. Three previously published studies are presented as cases to illustrate the ways in which a LaDDR platform's affordances manifest. The discussion covers how LaDDR platforms can address the aforementioned challenges, including advancing human-technology collaboration and how this approach can be extended to other design platforms.

KEYWORDS

design research methods, big data, design platforms, supporting learning, sociotechnical systems

1 Introduction

Research in engineering design has established that design problems are simultaneously complex in structure (Summers and Shah, 2010), ill-defined (Simon, 1996), under-determined (Dorst, 2004) and allow for multiple viable solutions (Crismond and Adams, 2012; Yilmaz et al., 2016a). Consequently, it is often necessary to collect detailed design process data to understand designers more

fully. This need is particularly salient in the study of design activities or strategies used by designers which involve an iterative and intentional process to transform an initially ill-defined problem into a communicable design concept (Dym et al., 2005; Atman et al., 2007; Razzouk and Shute, 2012). Additionally, this need impacts the study and practice of engineering design education, which itself is inextricably linked to design research (Dym et al., 2005, pp. 10). The need for detailed process data raises several challenges for research, including: 1) adequately collecting comprehensive and relevant data, 2) time-costs for data preparation and analysis, and 3) subsequent constraints on advancing substantive and pedagogical understanding and applied efforts. More concretely, this need is tightly coupled with two major ongoing challenges in design research. First, a methodological challenge stems from the fact that many of the procedures used for collecting detailed process data, such as protocol studies where a designer thinks “outloud” while completing a task (Coley et al., 2007), require data to be collected serially from isolated designers and therefore greatly limits the number of participants (Chiu and Shu, 2010). This places constraints on what types of analytical approaches are usable (i.e., those requiring more participants) for studying design activities and may limit some design research and education progress. Second, a technology and training challenge stems from rapid and widespread advances in smart technologies and their impact on industry and training new engineers (National Science Foundation, 2020; Jiao et al., 2021). As Jiao et al. (2021) explain, we are moving into industry 4.0 where advances in AI, cyber-physical systems, and human-technology collaboration are transforming our industry, manufacturing, and design, making it imperative to incorporate these technologies into training future engineers. Moreover, this change opens new opportunities to deploy these smart technologies to study and support practicing designers and support the learning of new generations of engineers. In this manuscript, a technological framework is proposed that aims to fuse techniques from big data methods into design platforms (e.g., computer-aided design or CAD) that are used to study designers or design learning. In so doing the framework aims to create an approach that can capture detailed design process data in a highly scalable manner.

Big data approaches typically leverage information systems to collect and subsequently analyze extensive and detailed datasets to enable discovery (Chen et al., 2014, pp. 4), modeling (Pietsch, 2016) and associated applications. In educational fields, big data methods have been used to investigate similarly complex learning processes (Siemens, 2013), such as meta-cognition (Sonnenberg and Bannert, 2016). However, big data approaches have seen slower adoption in design research that studies designers and design learning, with the majority of existing work focusing on design

platforms that support the automatic logging of design actions for constrained tasks (Jin and Ishino 2006; Ritchie et al., 2008; Sung et al., 2012; Alelyani et al., 2017; McComb et al., 2017; Song et al., 2020; Phadnis, Wallace, and Olechowski, 2021).

We call our proposed framework for fusing big data methods and design platforms for studying designers the Large Data for Design Research (LaDDR) framework. At its core, the LaDDR framework is a set of software design principles for reimagining these platforms. There are three major principles: 1) a broad scope of simulated actions, 2) unobtrusive logging of designers’ actions, and 3) support for creation, transformation, and analysis actions. Following these principles, a LaDDR design platform records a detailed log of designers’ actions within the system, including their transformations to an artifact(s) and the production of information about the artifact(s). This record of sequential design actions forms a design process action or activity stream from which design activities may be inferred. In other words, this action stream provides a trace of a designer’s ongoing activities as they explore the problem and generate solutions. This work follows past research that uses observable design actions to understand how people think as they design (Rahman et al., 2019a; Hay et al., 2017, pp. 8-9; Boyle et al., 2009; Kruger and Cross 2006; Sim and Duffy 2003).

The central contribution of this work, the LaDDR framework, aims to afford researchers and educators a new approach to platforms that fuses big data methods and design platforms to address the specific challenges raised by needing to collect design process data to study design activities. This framework is intentionally moderately general so that it can be applied and adapted across different technical and problem areas. In so doing, this work seeks to advance the community’s research capacity to study design activities and ability to support the human-technology partnership key to the future of work. This paper is forward-looking as it describes this framework, what it affords research and education, illustrates its use, and then envisions how it could help address the challenges presented above.

The remainder of the paper is structured as follows. Big data methods and design expertise research are reviewed. The review ends with the proposal of the LaDDR framework and what distinguishes it from past work. Next, the framework’s software design principles are presented as well as the affordances LaDDR platforms provide. After this, the platform that inspired the framework is outlined. This is followed by a description of three brief case studies (Schimpf and Xie, 2017; Schimpf et al., 2018a; Schimpf et al., 2018b) that used this approach to illustrate its affordances. The discussion first examines what the LaDDR cases demonstrate for research design, and then explores how LaDDR platforms can help in addressing the aforementioned challenges in studying and supporting designers, training a

new generation of engineers, and enhancing human-technology collaboration. Second, it presents considerations for how the LaDDR framework can be extended to other design platforms. Lastly, conclusions address limitations of the approach and broader implications for the field.

2 Literature review

2.1 Big data

A common definition of big data emphasizes three aspects of the data: volume, variety, and velocity (Laney, 2001). While big data is sometimes used to refer to data capturing and storage systems, many authors use big data to refer to both data capture/storage systems, and the analysis applied to the data (e.g., see Kitchin, 2014, pp. 2); for example, analysis through data mining (Miller 2010; Wu et al., 2014) or machine learning (Qiu et al., 2016).

In addition to the aspects of big data outlined above, Kitchin (2013), Kitchin (2014) synthesizes several definitions of big data to create a more comprehensive definition of the approach. Two attributes are of particular interest for engineering design research: exhaustive and fine-grained. Exhaustive means capturing all possible relevant data for a given problem (for the present study, this would include all the unique strategies or task actions different designers employ), whereas fine-grained means collecting data of the highest granularity possible. Thus, big data includes capturing and analyzing data, where the data is voluminous, exhibits variety, is generated at a steady rate, attempts to be exhaustive in scope, and is fine-grained in detail.

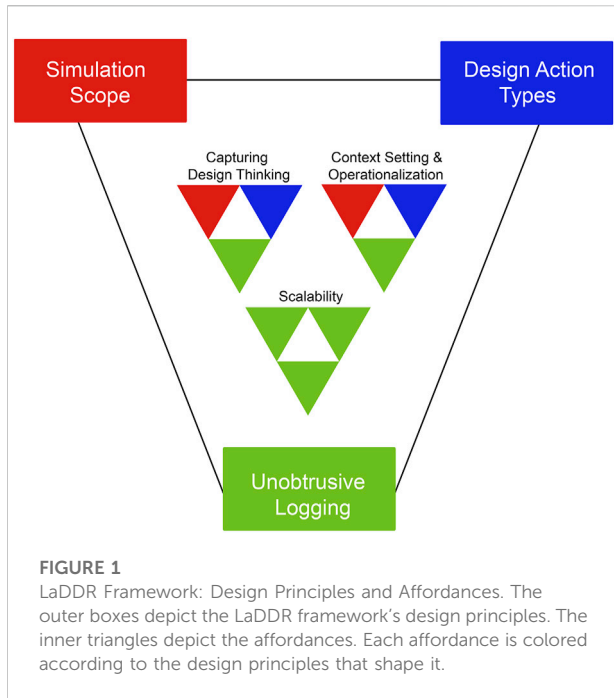
Design research has seen considerable attention to big data methods. Some areas in which big data has been applied or studied include analyzing customer requirements and preferences to guide new or revised product design (Shi and Peng, 2021; Chiu and Lin, 2018; Lin et al., 2016), using a fully integrated internet of things approach to new product design (Lee et al., 2022), assisting with planning and coordinating the manufacture of large-scale products (Bao et al., 2018), and assisting engineered system optimization or optimization of system components (Bostanabad et al., 2019; Xiong et al., 2019). For instance, Chiu and Lin (2018) used text mining and Kansei engineering (Nagamachi, 2002) on online customer reviews to extract key terms to predict consumer preferences for future products. These approaches are aligned with data-driven design, where big data, smart technologies and analytical innovations are applied to advance parts of the design process, decision-making, manufacturing, and associated areas (Kim et al., 2016; Jiao et al., 2021).

While there is a substantial and growing body of work on big data and data-driven methods for design research, these methods have not been fully applied in design research

focused on studying designers or how designers learn, in specific. In this area, researchers have employed automatic data-logging systems in engineering design software, such as CAD platforms, to unobtrusively collect designers' actions while they complete a task (Jin and Ishino 2006; Sung et al., 2012; Sivanathan et al., 2015; Alelyani et al., 2017; McComb et al., 2017; Rahman et al., 2019b; Rahman et al., 2020; Song et al., 2020; Phadnis et al., 2021; Deng et al., 2022). Studies in this area predominantly present designers with relatively constrained tasks and consequently capture design actions with minimal variety or fine-grained detail. Researchers have also used automatic data-logging to capture other types of data from designers, such as eye-tracking to capture where designers are focusing their attention (Li et al., 2019; Kwon et al., 2020) or web camera and emotion detection software to capture designers' emotional states (Phadnis et al., 2021). However, these data capture methods require extra physical equipment and software and may be more disruptive to designers, which impacts the scalability of these tools. Returning to design software that logs design actions, in terms of analysis, this work has employed machine learning approaches like artificial neural networks (e.g., Rahman, et al., 2020), data mining approaches like cluster analysis (Jin and Ishino, 2006) and more traditional statistical methods like regression (e.g., Alelyani et al., 2017; McComb et al., 2017). While this work represents a promising start for leveraging big data approaches to study designers and design learning, there remains considerable room for growth.

2.2 Design expertise

Design expertise is a core topic in design research. For example, past work has uncovered that design experts are solution-driven, using preliminary solutions to reason through problem and solution possibilities (Lloyd and Scott, 1994) or use abstract knowledge schemas to make analogies between the current problem and past problems encountered (Ball et al., 2004). In terms of expertise development, there is a considerable amount of research that analyzes novice and expert differences (Atman et al., 1999; Kavakli and Gero, 2002; Ho, 2001; Ball et al., 2004; Cross 2004; Ahmed et al., 2003; Atman et al., 2007; Bjorklund, 2013). Some work has distinguished several broad design expertise stages (Lawson and Dorst, 2009, pp. 99), however they do not provide clear indicators to identify when a designer has moved from one stage of expertise to another (e.g., in their terminology, from advanced beginning to competent). One study provides clearer indicators is Crismond and Adams (2012), whose comprehensive literature synthesis revealed nine critical design practices or activities (e.g., idea scarcity vs. idea fluency) that can act as indicators to distinguish novice, intermediate (what Crismond and Adams call the



informed stage) and expert stages. Data-driven platforms could prove particularly useful for developing quantitative measures for these indicators.

2.3 Developing the large data for design research framework

The present work seeks to build on past research and platforms that logged design actions as well as research that leveraged novel methods for analyzing big data sources. To do so, it draws on several attributes of big data covered above. First, following Laney (2001) and Kitchin (2014), logged design actions should exhibit high variety and fine-grained detail to enable a thorough record of any given designer's process. Most contemporary computer-aided design (CAD) or computer-aided engineering (CAE) simulate a wide variety of granular design actions but often do not record these. Second, following the same authors, the volume (or number) of designers whose actions are captured should be high and capture should seek to be exhaustive in recording all unique manifestations (see Pietsch, 2016, pp. 141) of designers' processes. Many current platforms used in scholarly engineering design research (e.g., Sivanathan, et al., 2015; McComb et al., 2017) support collection from large numbers of designers and their unique processes through automatic action logging, but these platforms are typically limited in the variety and detail of design actions captured. In brief, few existing platforms embody high variety and granular design actions with unobtrusive logging. In this space, the Large

Designer Data (LaDDR) framework is proposed as a set of software design principles to integrate these big data attributes more fully into platforms for studying designers and to expand the scope of what design actions and designers' unique processes are captured. This framework is intended to be adaptable across different technical and problem areas.

3 Large data for design research framework

This section describes the LaDDR framework in terms of its design principles and the affordances it offers researchers and educators. This section presents the design principles first and then its affordances (see Figure 2). These principles were derived from research and development with a design platform with extensive logging capability and research results from other design platforms.

3.1 Design principles for large data for design research platforms

This section explains three core design principles (Fu et al., 2016) for creating design platforms that enable the three affordances in Figure 1. It is important to note that the design principles covered here are not meant to be exhaustive; in future work these may evolve, or new principles may emerge. Instead, these principles aim to cover the minimal core conditions for creating a LaDDR design platform.

The first design principle concerns simulation scope. The design platform should simulate a wide array of recordable actions within a domain(s) where design can be undertaken. The design platform must simulate a domain in which design can be undertaken so that designers can interact with that domain. Simulations rely on models of their real-world counterparts (Landriscina 2013). Models can represent a broader or narrower set of elements and actions from a domain. To allow for a wide array of actions in a designable domain, the simulation should model a sufficiently large phenomenon, such as an elaborate system or system of systems (Boardman and Sauser 2006) like a vehicle or microgrid. More formally:

$$A = P\{subsys_1, subsys_2, \dots, subsys_n\}$$

Where A is the set of all recordable actions in the platform and P is a partition of A into nonoverlapping subsystems within the domain where actions are being recorded. It is important to note that the size of a simulated domain is highly correlated to its design space. As we move into industry 4.0, the scope of design problems continues to expand and may include actions or activities from other business operations that are integrated with design, such as finance and marketing. For example, designing an energy efficient house may necessitate striking a

balance between energy efficiency, cost, and aesthetics to create the “best” solution (e.g., see, [Goldstein, Adams, and Purzer, 2021](#)) Design platforms that provide interaction with small domains such as a system component (e.g., a truss) or constrained systems will not allow for the same depth of activity. A wide simulated design space also provides designers with the flexibility to reconfigure the space as desired or identify various criteria/constraints depending on their experience or objectives. In other words, a sufficiently large, simulated design space supports designers’ distinct design processes.

The second design principle concerns unobtrusive logging. The design platform should surreptitiously log designers’ actions at a granular level throughout their design process. Surreptitious data collection means that the design platform records designers’ actions without any additional user input such as verbalization or journaling, i.e., unobtrusively ([Jin and Ishino 2006](#); [Sung et al., 2011](#)). Design actions here refer to the smallest possible actions in the platform that either transform the design artifact or provide the designer with new information. For example, the first type could be the addition of or change to a subcomponent of the artifact, and the second type could be performance calculation after a single change to the artifact. Design actions should be highly granular and can be aggregated into higher-level practices, strategies, or other categories. Similar aggregation techniques have been used with smaller sets of actions in previously presented studies (e.g., see [Deng et al., 2022](#); [Jin and Ishino 2006](#)) and other fields (e.g., see [Hilbert and Redmiles, 2000](#)). Based on this data-logging approach, an explicit data schema for recordable actions should be constructed. Finally, the design platform should log data with timestamps to preserve the designer’s complete process.

The third design principle, concerns design action types. The design platform should consist of actions for creating/transforming design artifacts(s) and analyzing design artifacts(s) relevant attributes. More formally:

$$D = \{d_1, d_2, \dots, d_n\}$$

$$DA = \{da_1, da_2, \dots, da_n\}$$

$$D \cap DA = \{\emptyset\}$$

Where D is the set of design/creation actions and DA is the set of analysis actions and both are nonoverlapping. While this principle may appear self-evident, many design platforms as previously mentioned, such as CAD systems, primarily focus on creation actions. Incorporating analysis actions allows more stages of the design process to be simulated, such as optimization. Subsequently, a platform following this principle will enable engagement in a fuller scope of the design process. Analysis is also part of several design practices, including iteration ([Schimpf and Xie, 2017](#)), design experiments ([Vieira et al., 2016](#)) and trade-off decisions ([Goldstein 2018](#); [Goldstein et al., 2018](#)). Thus, the inclusion of analysis actions further enables designers to undertake a broader array of design activities in a platform.

3.2 Affordances of large data for design research platforms

This section covers the affordances the LaDDR framework offers design research. Each affordance is also related to the big data attributes, reviewed in [Section 2.1](#) (see, [Kitchin, 2014](#)).

3.2.1 Capturing design activity

Observing and capturing design activities is the central affordance of a LaDDR platform and pertains to the ability to infer designers’ strategies and reasoning from their interaction with a platform. Said differently, the action stream data collected by a LaDDR platform leave a trace of designers’ cognitive processing used to navigate the problem and solution space.

This affordance is shaped by all three design principles. In terms of the first design principle, capturing design activity is enabled by supporting a wide array of actions and thus providing a sizeable and flexible design space in which to work. The second principle emphasizes capturing a thorough record of designers’ processes. In regard to the third design principle, the broad action types of creating, modifying, and assessing the design allow for greater inference about designers’ activity throughout their process, for example from conceptual design to testing. Finally, this affordance embodies the fine-grained, velocity, and variety aspects of big data. Collecting designers’ processes through the platform produces a detailed, real-time rendering of their numerous actions for studying design activity.

3.2.2 Context setting and operationalization of design activity constructs

The context setting and operationalization affordance is the ability for researchers/educators to adaptively scope the bounds of the design space (context) within a LaDDR platform and define metrics for measuring the central design constructs (operationalization). Setting the bounds of the design space establishes a context for creating design tasks; for example, by focusing on different substantive areas of the space or simplifying a challenge for novices by constraining the design space/stages. Thus, setting the context can support different research or education goals. Care needs to be taken to avoid unduly constricting the design space as this may reduce the strategies/practices designers need to use and diminish their creativity in addressing the challenge. After setting the context, granular actions can then be aggregated into categories that operationalize different design activity constructs (e.g., several building actions may be categorized as modeling).

This affordance emerges from all three design principles. In terms of context setting, the first principle (i.e., a large designable space) and the third design principle (i.e., support for creation and analysis action-types to support different design activities), combine to allow for creating different design tasks. In terms of operationalization, defining metrics for the central design activity constructs(s) is enabled through the near real-time capture of

designers' processes from the second principle, in conjunction with different action types emphasized in the third principle. This affordance embodies the variety and fine-grained big data attributes. Context setting and operationalization rely on various actions across creation and analysis types, which need to be fine-grained to scope different design challenges and serve as building blocks for design activity constructs, respectively.

3.2.3 Research design scalability

The research design scalability affordance of LaDDR platforms enhances and extends the capability of research studies by augmenting data collection, data processing, and data comparability efforts. In this framework, scalability refers to specific aspects of research design listed above. Data collection is widened, supporting collection from a large number of designers, including those with the same skills/experience and those with different skills/experience. Data processing is facilitated through the automation of retrieval, preparation, and/or formatting of captured data. Greater data comparability is facilitated through a shared design ontology (Štorga et al., 2010), enabling more direct comparisons within and between designers' processes.

The scalability affordance emerges primarily from the second design principle. Unobtrusive data collection simplifies data collection and supports collection from multiple designers simultaneously. It may also be deployed remotely. The second principle also covers developing a data schema for logging actions. Typically, these data schemas use a generic machine-readable format, like JSON or XML, which can be run through a script or program to automate data cleaning, processing, and some analyses to arrive at results more quickly. In light of the wide simulation scope discussed in the first design principle, the data schema provides a consistent but broad and adaptable set of design actions that promote comparability across designers' processes. In other words, the data schema provides a design ontology for what designers can do within a particular LaDDR platform. Finally, this affordance embodies the volume and exhaustiveness of big data attributes, as it supports collection from a larger sample of designers and seeks to capture all unique manifestations of designers' processes.

4 An example large data for design research platform: Aladdin

Aladdin is a CAD platform that supports architectural, photovoltaic (PV), and concentrated solar power (CSP) system design (Xie et al., 2018). The LaDDR framework evolved out of research and development with this platform. Here Aladdin is presented, and discussion focuses on how it demonstrates the LaDDR principles. The platform originally entailed energy-efficient home (e.g., Purzer et al., 2015) and some PV system design (e.g., Goldstein et al., 2015). See also Figure 2A. Subsequently a larger suite of PV systems, such as

solar farms (Figure 2B) were added. Moreover, concentrated solar power systems (see Figure 2C) were added. These systems may be designed separately or integrated within a task. The simulation scope or design space is therefore sizeable.

Early in its development, Aladdin incorporated an unobtrusive logging system to capture designers' actions in JSON, a widely used machine-readable data format. Each logged entry includes a timestamp, filename, the design action taken, and relevant metadata. For example, Figure 3 displays two logged entries, showing what is logged when a user adds a solar panel rack and when they run a system annual energy analysis. The "Add Rack" metadata includes a unique ID and the coordinates of the rack within the design platform. The full data schema contains over 200 unique actions relating to the design of buildings, PV, and CSP systems (Xie 2016). The log for a single designer "may exceed" 2,000 actions depending on the nature of the design challenge.

Finally, Aladdin has a built-in physics engine that emulates solar energy and heat transfer processes to estimate building heating/cooling as well as energy production for PV and CSP systems. Aladdin's energy estimates have been validated against the building energy simulation test (Gajewski and Pieniżek 2017). The platform likewise includes cost estimation for all designed systems. Thus, both creation and analysis actions are supported.

While Aladdin simulates a wide design space, researchers using it are constrained to work within its simulation scope. The LaDDR framework seeks to abstract from this platform so that other platforms may be able to capture rich design actions for inferring design activity.

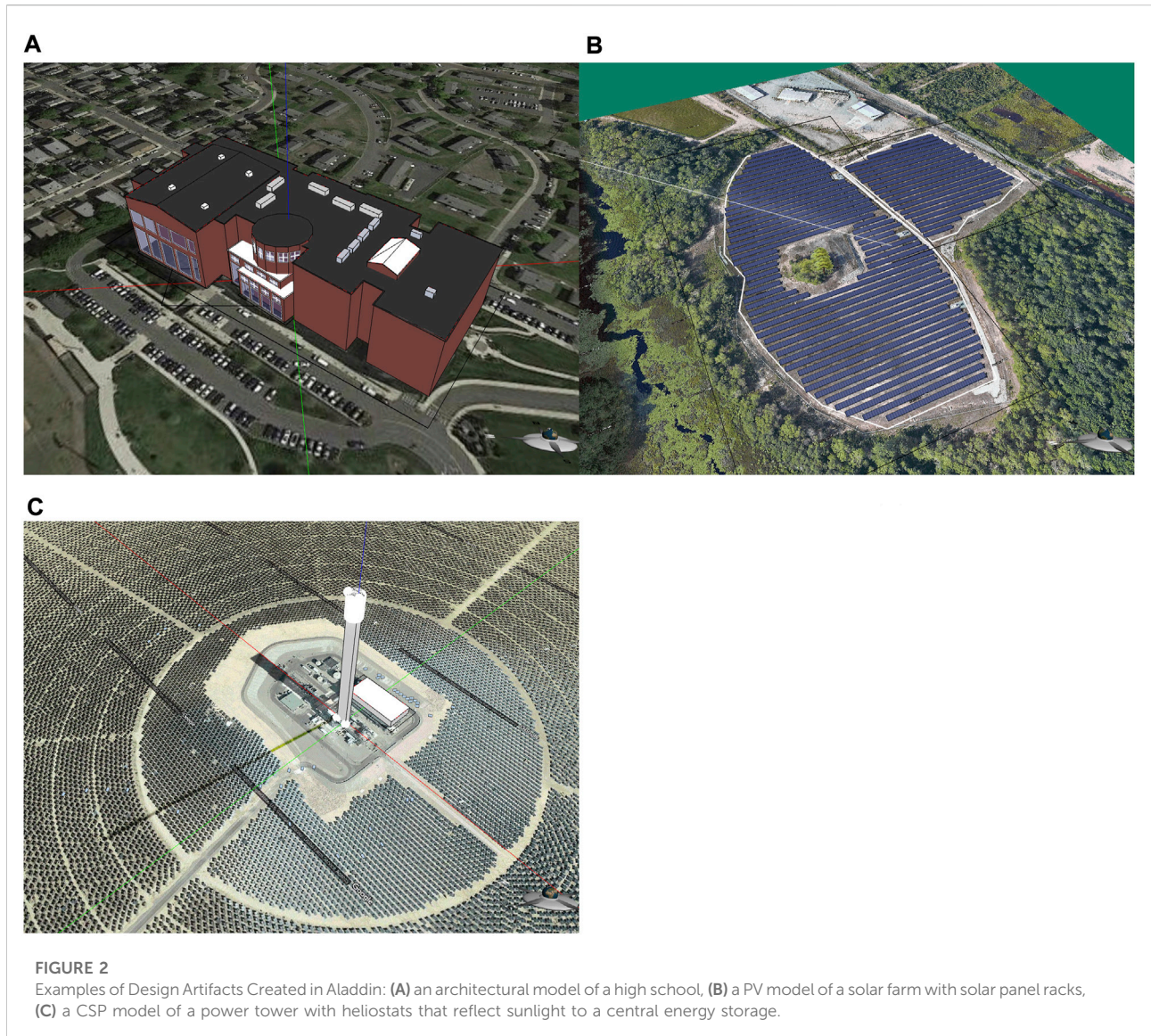
5 Results: Three illustrative cases

Aladdin is presented as an exemplar LaDDR platform for investigating design activities. Several studies have been conducted with Aladdin. Three studies are presented here: 1) micro-iterations, 2) design team analytics, and 3) design action sequences (summarized in Table 1). Each study has been previously published, and as such only the key details relevant for this manuscript are provided. These studies were selected to illustrate some of the ways a LaDDR platform can exhibit their three affordances, namely capturing design activities, context setting and operationalization of concepts of interest, and research scalability.

These cases were intentionally selected to emphasize the variety of ways a LaDDR platform may be applied to design research.

5.1 Case 1: Micro-iterations

This study sought to identify the micro-iterations of novice designers (Schimpf and Xie, 2017). On scalability,



data was collected from three-ninth grade classes for a total of 60 students, with 27 students from one section used in this study. These students were tasked with modeling their home and designing an adjoining solar array that generates sufficient energy to meet their household demand while also balancing budgetary constraints. This study used sequence mining, a data-mining approach for discovering nontrivial patterns and relationships in sequential data (Han et al., 2012, pp 588-589). The analysis was exploratory and sought to identify and characterize micro-iteration sequences. The central phenomenon of interest, iteration, can be defined as an intentional, goal-directed strategy (Adams and Atman, 2000) where designers reapply part of the design process to advance their design's development Wynn and Eckert (2017) a more recent piece.

Iteration has been analyzed as design stage cycles (meso-scale, e.g., see Chusilp and Jin, 2006; Adams and Atman, 2000) and as full artifact cycles (macro-scale, e.g., see Smith and Tjandra (1998). This study narrows the scale of iteration to a relatively unexplored area of micro-iteration (Schimpf and Xie, 2017) or cycles consisting of small design actions or operations. Several rules were established, discussed more below, for identifying iterations. Micro-iterations patterns were clustered into groups based on similarities in iteration patterns. Clustered iteration sequences were shared with an independent design expert to evaluate their soundness and consistency; differences were discussed and resolved.

Briefly, the results uncovered 20 micro-iterations, which were clustered into four major types: 1) solar panel system capacity testing, 2) solar panel location analysis, 3) solar



TABLE 1 Empirical case studies at a glance. This table highlights key aspects of each case.

	Micro-iterations	Design team analytics	Design action sequences
Designers	High School students (N = 27)	Engineering juniors and seniors (N = 17)	High school students (N = 33). Junior high students (N = 152)
Working alone or in a team	Alone	Team	Alone
Design challenge	Solarize your home	Solarize your community	Net-zero energy home
Study period	2 weeks of in-class time	2 days	2 weeks of in-class time

simulations with panel placements, and 4) investigating the Sun’s path across seasons. Around 41% (11) of the students used at least one micro-iteration and about 22% (6) engaged in two or more. While not commonly exhibited by these novices, the results verified the presence of micro-iterations and some forms they may take.

In terms of context setting for this challenge, the design space was primarily scoped to include photovoltaic system creation and analysis. The logs revealed 98 different actions, including home modeling, photovoltaic design, and other platform actions. Of these, 15 of the actions related to the micro-iterations, specifically actions for creating photovoltaic systems (e.g., adjusting solar panel efficiency) and both quantitative (e.g., annual kWh generation) and qualitative (e.g., tracing the sun’s path) analysis tools. The majority of these actions were related to home modeling and not design as the goal was a fixed model outcome. As such these modeling actions and some system-control actions (e.g., save) were not part of the analysis.

To operationalize micro-iterations, a system of constraints was defined to demarcate valid iteration sequences. Given:

$$Sol = \{sol_1, sol_2, \dots, sol_n\}$$

$$NSol = \{nsol_1, nsol_2, \dots, nsol_n\}$$

$$S_i = (s_{i1}, s_{i2}, \dots, s_{in})$$

Where Sol is the set of solar actions (both analysis and PV creation), $NSol$ is the set of non-solar actions and S_i is Student i ’s sequence of design actions, a micro-iteration MI_{ij} is identified when the following constraints are met:

$$MI_{ij} \geq 3 \text{ and } s_{i1}, s_{i2} \in Sol \text{ and } s_{i1} \neq s_{i2}$$

$$\forall x_{ik} \in MI_{ij}, x_{ik} \in NSol < 3$$

$$\exists s_{ik} \in MI_{ij} \text{ and } s_{ik} \geq 2 \text{ and } s_{ik} = s_{ik}$$

Where MI_{ij} is the j th micro-iteration and a subsequence of S_i ’s design action sequence, where this subsequence contained at least two distinct solar actions and one repeating solar action (as outlined in the first and third constraint, respectively). The

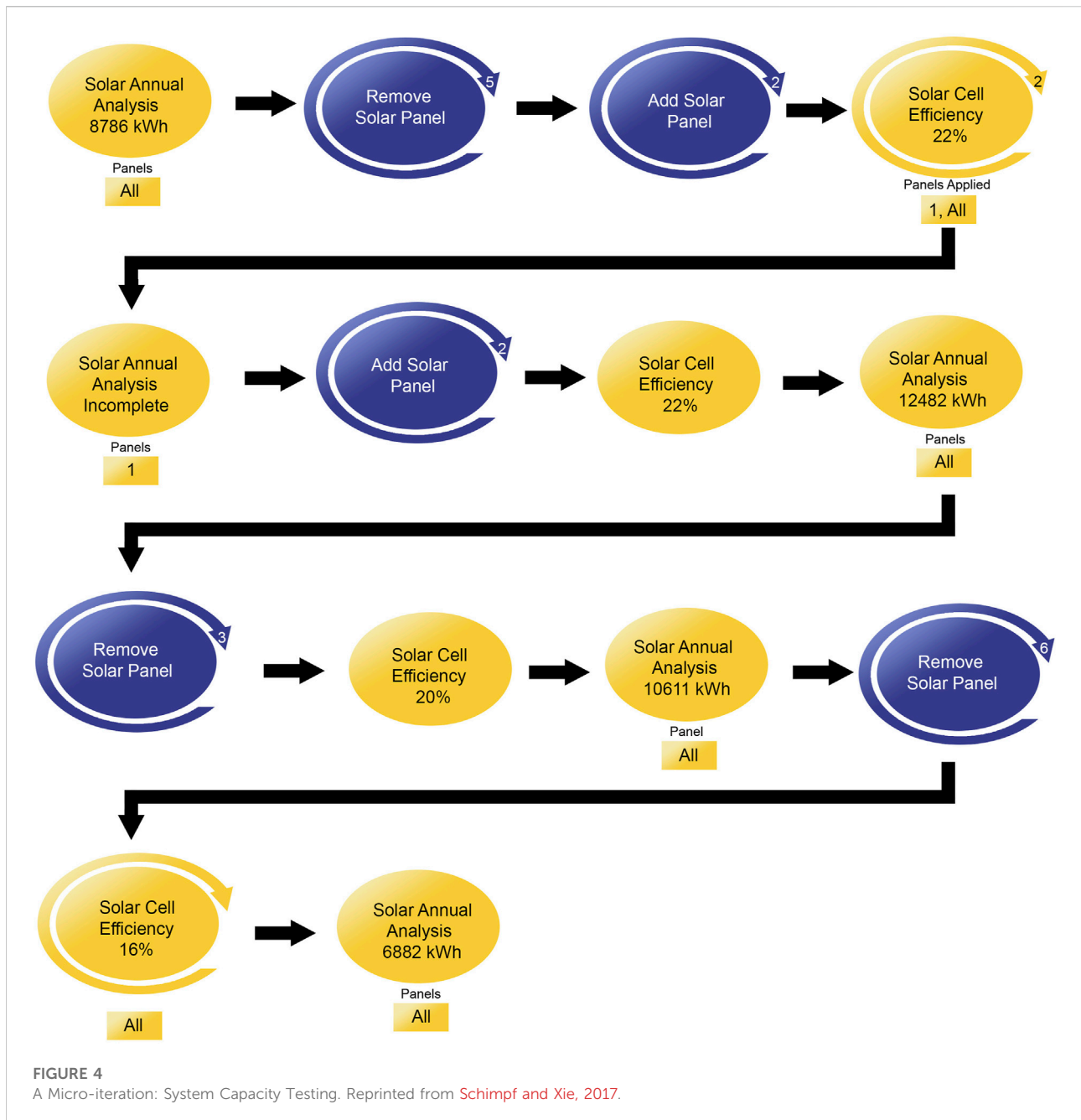


FIGURE 4
A Micro-iteration: System Capacity Testing. Reprinted from Schimpf and Xie, 2017.

second constraint allows for a small number of non-solar actions within a micro-iteration subsequence to mitigate instances of user error (e.g., clicking on the wrong part of the model) from disqualifying an otherwise valid iteration. From this system of constraints, a set of micro-iterations subsequences were identified.

Having covered context setting and operationalization, an example micro-iteration is presented in greater depth to illustrate the captured design activities. In Figure 4, yellow nodes represent solar analysis and blue nodes represent

photovoltaic system construction. Arrows around a node indicate repetition and additional information is presented below nodes. The micro-iteration in Figure 4 exemplified several iterations by designers where they cyclically tested, modified, and tested again. In this particular example, the performance of the system is tested, panels are added, removed, or adjusted leading to greater system performance. Two more cycles lead to progressively lower energy-generation and cost. More generally, this and related patterns were reflected efforts to balance cost and energy

production through evaluating alternative design configurations.

5.2 Case 2: Design team analytics

This study compared design teams' processes with particular emphasis on modeling, evaluation, and optimization, and these practices relationship with teams' artifact performance (Schimpf et al., 2018b). On scalability a group of 28 undergraduate engineering students from a medium-sized university were tasked to meet the energy demands of their campus by designing photovoltaic systems on a series of community sites (e.g., the local high school). Designs were assessed by four criteria, energy production, cost-effectiveness, cost, and aesthetics; each had weights to promote trade-off opportunities.

Methodologically this work draws on data visualization. Data visualization is a data mining approach (Han et al., 2012, pp 602) where visualizations are used to represent complex, multi-dimensional data for analysis. Several visualization techniques were applied (e.g., see Gleicher et al., 2011) such as overlaying modeling and optimization to show their collective trajectories and juxtaposing evaluation as a subplot to enable comparisons and identify transitions. Modeling, evaluation, and optimization are later stages of design, which designers must navigate to realize and refine design alternatives (Atman et al., 1999; Dym and Little, 2003). Past research has focused heavily on early-stage conceptual design (Dinar et al., 2015; Hay et al., 2017), with fewer studies about these later stages.

Briefly, after removing teams with non-consenting members, four teams were analyzed ($N = 17$). Two teams, A and D, submitted multiple sites whereas other teams submitted single sites. Each team's design process was visualized. The analysis revealed two connections between teams' activities and the design performance. First, a regular pattern of alternating between modeling and optimization was associated with greater cost-effectiveness in the teams' designs. Second, partial or absent evaluation of design sites was associated with lower cost-effectiveness. This work demonstrated the potential of this analysis for generating holistic and multivariate visualizations that can reveal complex interrelationships between design activities and final design performance.

In terms of context setting, the design space was scoped to photovoltaic system creation, manipulation, and analysis. The logs showed 40 different actions, including date/time changes, photovoltaic system, and other system actions. Design action logs were analyzed to partition actions into subsets *E*, *M*, and *O*, representing evaluation, modeling, and optimization. While evaluation actions were straightforward to identify, a distinction was made between actions involving the initial setup of a system and system parameter adjustments, as modeling and optimization, respectively. In total these accounted for 17 of the recorded actions. Due to

advancements in Aladdin to support solar racks, there were more PV actions than the micro-iterations case.

Design teams' modeling, optimization and evaluation actions were operationalized following Figure 5. Note that this figure displays fewer than 17 actions from above, as several actions can be applied at multiple artifact levels resulting in different action labels. For simplicity and readability, these are collapsed in Figure 5. Modeling included basic setting, sizing, and editing of teams' PV systems and optimization included adjustments to the types of panels used, rack tilt, and heading (azimuth). Evaluation focused primarily on annual analysis of their system including full system or group analysis, where a subsection of panels was analyzed.

A select team is now presented to illustrate the captured design activities. Team C's design process is visualized in Figure 6. Their process is divided into sessions based on visible breaks in activity from their log. The lower graph plots modeling and optimization and the top displays evaluation. The *y*-axis displays the number of actions and the kWh production, respectively. Evaluation may be for the full system or a subset, all year or partial year. The dashed vertical line indicates shifts between design sites. For the first site, an elementary school, the team spent time alternating between modeling and optimizing the system. After evaluating its energy performance, the team switched to a high school with a much larger roof-space. Here they created a more extensive model and used their last session to iteratively modify and evaluate their system, resulting in a large-scale optimization cycle. Ultimately Team C submitted this single design, which performed the highest in energy production and aesthetics. By plotting these three sets of activities on a shared timeline, this example reveals interconnections between activities and how they relate to the design performance.

5.3 Case 3: Design action sequences.

For this study, Markov chains were used to model students' design processes as a series of design sequences to facilitate comparisons and reveal design dynamics (Schimpf et al., 2018a). On scalability a 152 junior high and 33 high school students were tasked with designing an energy-efficient home which has annual net-zero energy use and stayed within budget. Additional structural constraints included area specifications and window-to-wall ratios.

Markov chains are used for modeling systems with multiple states, where future states depend on the current state; this method has received growing attention in design (Gero and Peng 2009; McComb et al., 2017; Rahman et al., 2019a). Markov chains are another type of sequence analysis (Cornwell 2015). Here states are designers' actions, and the full model contains all transitions between design actions. Transitions are displayed in a transition matrix, with entry α_{ij} of the matrix representing the

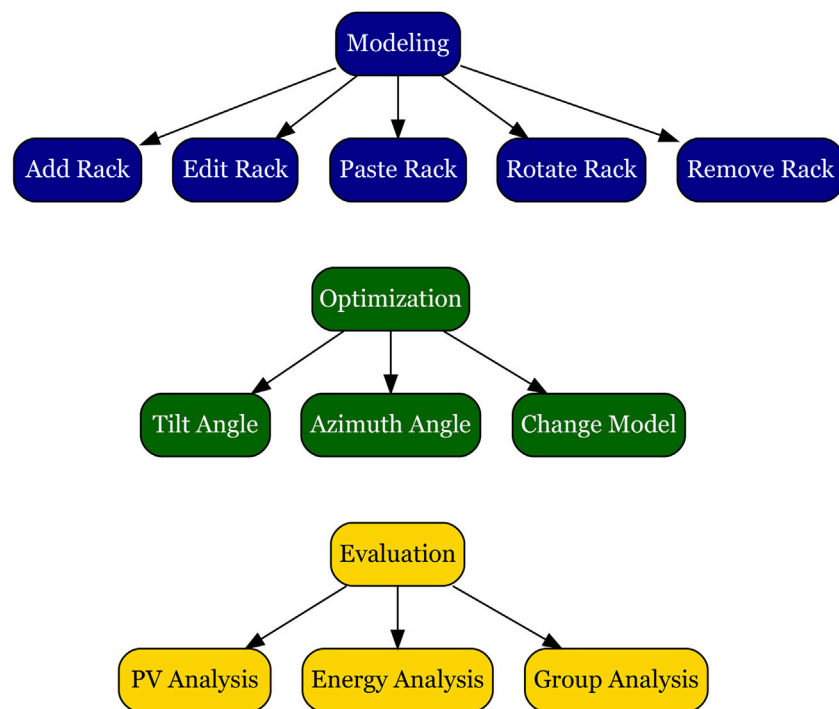


FIGURE 5
Design team analytics schema. Each higher order category subsumes the actions beneath it.

probability of transitioning from design action i to j . This is calculated by $\alpha_{ij} = a_{ij}/a_i$, where a_{ij} is the number of times design action j follows action i , and a_i is the total number of times action i occurs. Design sequences can reveal designers' strategies, such as what activities are associated with evaluation. Psychology researchers argue that sequences are central to many skills (Destrebecqz, 2005; Clegg et al., 1998) and some design research has found that learning sequences improves design performance (McComb et al., 2017). Thus, design sequences can give insight into how designers navigate problems.

Briefly, this preliminary study analyzed ten of the most engaged novice designers; five designers were sampled from each group. It was found that both groups had high levels of within-action sequences for modeling actions, e.g., windows, suggesting they focused on individual home subsystems sequentially. The two groups of students exhibited differences in what components of the building they interacted with; junior high students ignored insulation whereas high school students ignored foundations. Both groups had design sequences that concentrated on energy production or energy production and passive solar design. The high school students had more sequences connecting evaluation to other actions, for example note-taking. This work gave evidence that design sequences can help characterize similarities and differences between groups of designers.

In terms of context setting, the design space was scoped to include house construction, landscape layout, PV systems, energy analysis, and note-taking. The log revealed 95 unique actions from high school designers and 121 unique actions for junior high designers. Unlike the previous two examples, analysis for this challenge involved estimating solar panel generation and household energy-use. System controls and actions unrelated to the challenge were discarded from analysis. This resulted in a set of 54 unique design actions.

Due to the large number of design actions for this challenge, it was necessary to consider how to best operationalize key design activities. Given houses several components, actions were partitioned into subsets reflecting house subsystems as well as different analysis tools. For example, the "Building" subset $B = \{\text{"Move Building," "Resize Building," "Rotate Building," "Add Components."}\}$ The data schema is presented in Figure 7. The enclosed boxes contain actions subsets, with the category title atop the box.

Next, a detailed example of one student's transition matrix is presented to illustrate captured design activities. As shown in Figure 8, this junior high student had high levels of within-subsystem transitions, particularly with notetaking, solar panels, and windows. This represents a logical decomposition strategy to focus on individual subsystems. Furthermore, this student showed a weaker across-action sequence (probability = 0.33)

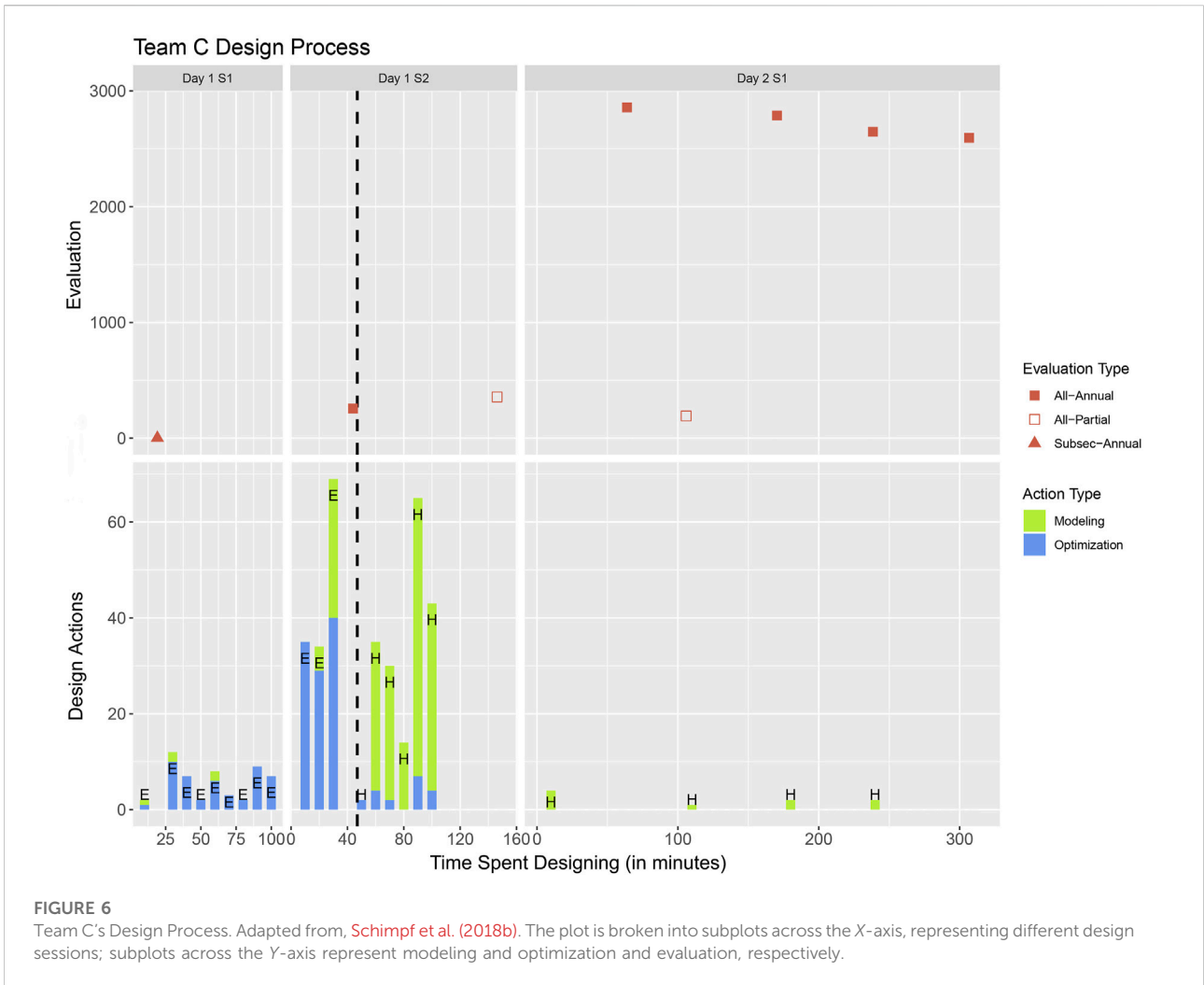


FIGURE 6 Team C's Design Process. Adapted from, Schimpf et al. (2018b). The plot is broken into subplots across the X-axis, representing different design sessions; subplots across the Y-axis represent modeling and optimization and evaluation, respectively.

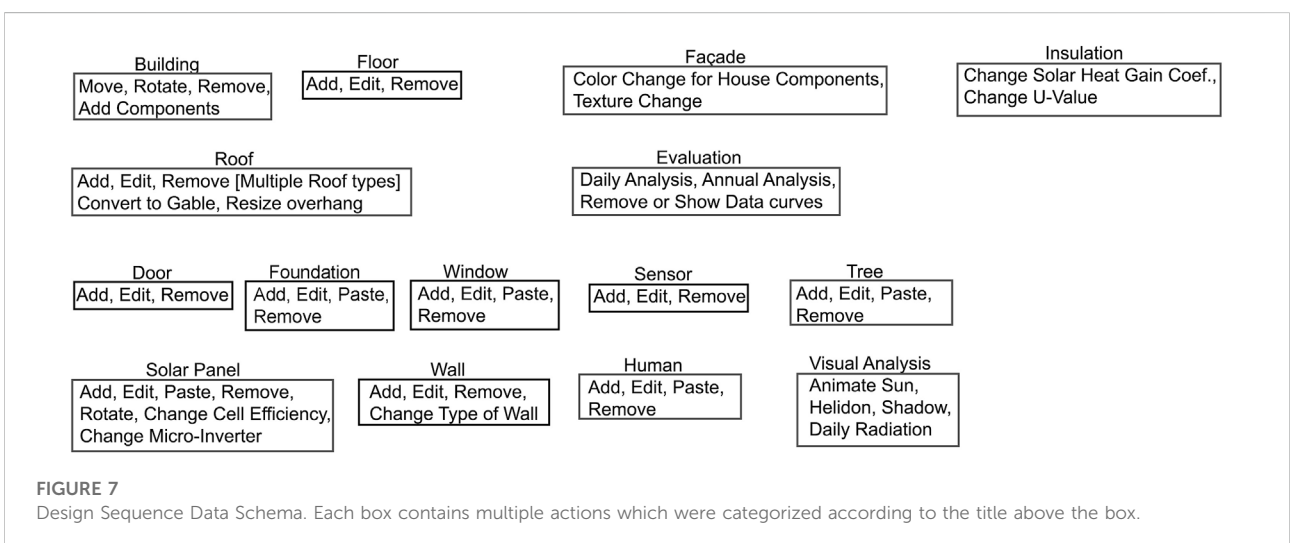


FIGURE 7 Design Sequence Data Schema. Each box contains multiple actions which were categorized according to the title above the box.

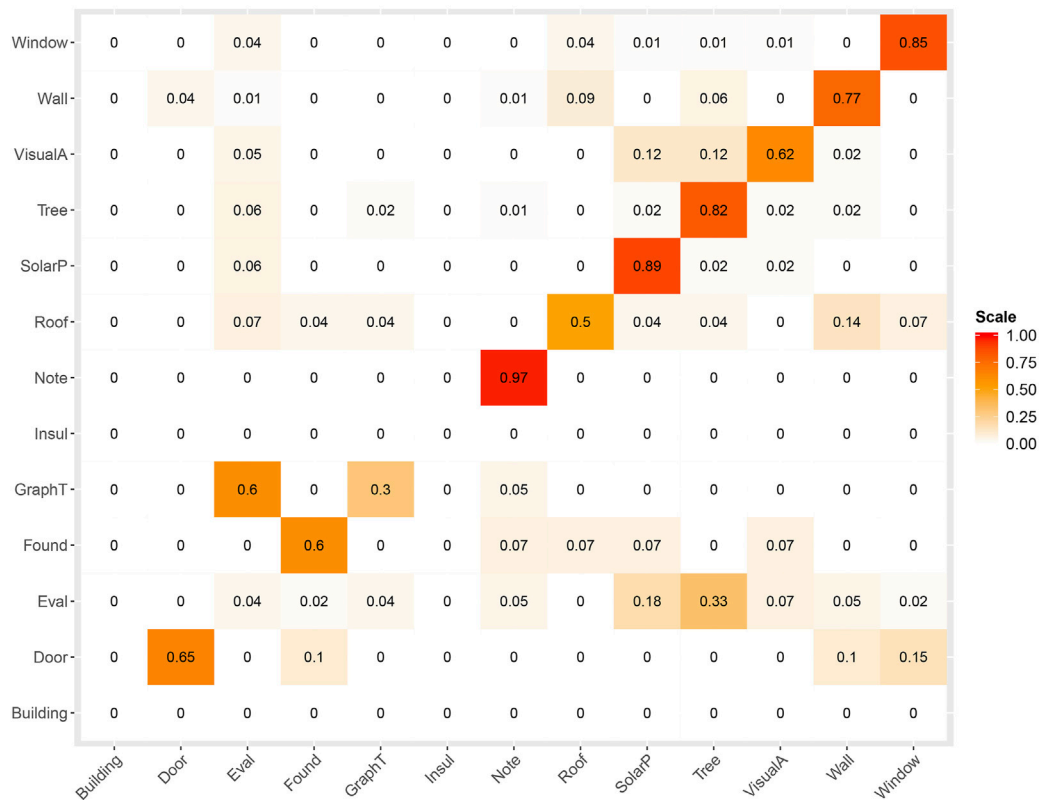


FIGURE 8

Novice Designer's Design Action Markov Chain Transition Matrix. The plot should be read from the Y-axis representing the first action, and entries on the X-axis representing the section action.

between evaluation to tree placement, suggesting this student was trying to use tree placement to improve the efficiency of their home. There is also a weak transition between evaluation and solar panels, reflecting PV system testing. The graph tab, which shows information about a design's cost and attributes, was often reviewed before evaluation (probability = 0.60). As this example illustrates, design sequences can give a window into designer's activities and can serve as a general model for comparing designers.

6 Discussion

In the first section of the discussion, we synthesize how the affordances operated across cases to examine the range of possibilities LaDDR platforms offer research design. Three dimensions emerge: Setting the design problem complexity, selecting design activities of interest, and recruiting study subjects.

The second section revisits the two challenges facing design research. The challenges are briefly restated and the discussion explores how LaDDR platforms can address or leverage

opportunities associated with these challenges. In particular, discussion of the second challenge emphasizes the promise LaDDR platforms hold for enhancing human-technology collaboration and details why these platforms are well positioned to advance work in this area. The third section presents considerations for extending and scaling the software framework to other design platforms.

6.1 Large data for design research for research design

6.1.1 Setting the design problem complexity

Summers and Shah (2010) proposed three metrics for measuring design problem complexity: the number of variables, the number of connections between variables and the solvability or difficulty of finding a solution. Setting design problem complexity is primarily shaped by the design context and operationalization affordance. First, as shown in the cases, it is possible to set a design problem's complexity from a greater to fewer number of design variables or actions, thereby decreasing or increasing the solvability of the challenge designers face. For

example, by including more variables by which solar panels can be manipulated, more connections between variables may emerge and solvability will require more domain knowledge, as seen with panel type and rack angle in the design team analytics case. In contrast, the micro-iterations case only had panel type and therefore increased solvability. The ability to set the problem complexity can also vary by changing the number of artifacts designers are responsible for, as seen in the design team analytics case where designers could select several sites for creating PV systems.

6.1.2 Selecting design activities of interest

Selecting a design activity of interest is shaped by both the capturing design activity and design context and operationalization affordances. In selecting a design activity to study, the case show support for defining and measuring well-established design activities, such as evaluation, modeling, and optimization processes covered in the design team analytics review. Moreover, new or understudied design activities may be uncovered, as was demonstrated in the micro-iterations case.

The cases also demonstrate how these platforms can be used to study design activities at multiple scale levels, from small two-action design sequences to middle range micro-iterations to full design process analysis in the design team analytics case.

6.1.3 Recruiting study subjects

This dimension of recruiting study subjects' is shaped both by the scalability and design context affordances. Turning first to scalability, a LaDDR platform can be run with a small or a large number of participants, with the design sequence study highlighting the relative ease of collecting logs from nearly 200 participants. In particular, the relative ease of data collection stems from the unobtrusive nature of LaDDR, allowing for simultaneous data collection across many participants without concern for cross-contamination or interference between designers engaged in a design challenge. This aspect lends itself well to university, school or workshop settings where there may be many designers available to participate at the same time. Finally, in regarding design context, different design challenges can be created for different levels of design experience and expertise, as reflected in the distinct problems in the cases.

6.2 Addressing design research challenges

This section returns to the two major challenges facing design research and elaborates how the LaDDR framework, when applied to platforms, provides an innovative toolset that opens new research directions for design. Recall, there was a methodological challenge centered around methods that collect rich design process data (e.g., protocol studies) only being able to collect data from a relatively small number of

designers. This constrains the research questions studied in design research and subsequently stymies advances in said research. There was also a technology and training challenge centered around smart technologies transformations on modern industry and design, and the need to train a new generation of engineers for these changes and support current practicing engineers. This challenge also raises opportunities to deploy these same technologies to revolutionize how designers are studied and supported and how students learn.

This leaves the community with a partial view on several fundamental topics, including the variety of ways design expertise may manifest, the different forms of greater or lesser mastery, and how practices may evolve as designers gain experience.

6.2.1 Methodological challenge

For the methodological challenge, a LaDDR platform opens new opportunities by enabling data collection from many designers simultaneously through unobtrusive logging. This expanded data collection could support the use of a wider set of research methods, particularly data mining and machine learning methods, to answer new and underexplored questions. For example, data mining methods such as sequence analysis or decision trees (see Han, Kamber, and Pei, pp 331-336) could be used to uncover new design activities or practices or draw out distinctions in practices, as was demonstrated in the micro-iterations case. Alternatively, work could use cluster analysis (e.g., k-means or hierarchical clustering, see Jain, 2010) to discover or characterize distinct subgroups of designers, at varying or similar levels of experience, who employ different practices or strategies while designing. In terms of machine learning, a new direction may involve building models of designers' activities for a task using a neural network (see Rahman et al., 2019a). After training the model with an initial group of designers, researchers could predict new designers' relative design outcome success or forewarn of the difficulties they may encounter.

6.2.2 Technology and training challenge

As industry 4.0 (Jiao et al., 2021), the future of work and the potential for human-technology collaboration has received growing attention, there has been increased interest in how AI and associated smart technologies can be used to support design, both in terms of practitioners (Zhang et al., 2021; Song et al., 2020; Khan and Awan, 2018) and learners (Schimpf et al., 2019; Chen et al., 2020). LaDDR platforms open up new opportunities for integrating AI support into design platforms. Much of the existing work that incorporate some form of AI assistance into a design platform have AI that operate on artifact permutations or the state space of design artifacts (e.g., Khan and Awan, 2018; Schimpf et al., 2019; Chen et al., 2020). In contrast, a LaDDR platform could inform AI systems by providing designers' action stream, representing their design activities. Operating on

designers' action stream, AI techniques, such as deep learning (Raina et al., 2019), could be built to recognize creative or effective design strategies designers use and provide targeted feedback to struggling designers, similar to design heuristics (Yilmaz et al., 2016b). This approach could be leveraged in professional settings as well, for example by extracting strategies from a broad range of engineers and providing these as alternatives to a target designers current strategy. These strategies could also be incorporated into computational design agents that support novice designers with difficult parts of the design process or increase the efficiency of practitioner teams by adding computational agents to the team trained on specialized strategies from technical domains not represented on the team. Some work has begun to use actions logged by smaller-scope design platforms to inform or create AI support systems (Zhang et al., 2021; Raina et al., 2019; Egan et al., 2015), creating AI systems from LaDDR action logs could build upon this work.

LaDDR platforms also hold promise for expanding human-technology collaboration as these platforms are versatile across different levels of designer experience and age. There are two major factors that make LaDDR platforms widely accessible. First, they use a consistently shared design action schema or what can be considered a design ontology (Štorga et al., 2010) for describing all possible actions in the simulated space. This eliminates any inconsistencies that may emerge when data collections rely on designers' idiosyncratic verbalizations, self-reporting, or subjective impressions. Second, LaDDR platforms' context-setting affordance allows researchers or instructors to shape the nature of the design task assigned to be appropriate for the skill-level and experience of the tested population (e.g., simplified tasks for novices or more open-ended tasks for experienced designers). While many studies have looked at undergraduate trainees or professionals (Kavalki and Gero, 2002; Cross, 2004; Atman et al., 2007; Björklund, 2013), there has been increased interest in targeting young population in K-12 with engineering design activities because of the broad utility of engineering design skills (NRC 2012; NAE and NRC, 2014) and need to reach students early to fully support their development as designers (NAE and NRC, 2009; Cunningham and Lachapelle, 2014). In short, LaDDR platforms offer a more complete pipeline for supporting development, ongoing professional training and active practice in engineering design skills and strategies.

6.3 Extending the large data for design research framework to other design platforms

The LaDDR framework and its design principles were written to be sufficiently general to be applied to a large selection of design platforms. These design principles could be incorporated into new design platforms, such as those

created by researchers, or into existing platforms such as computer-aided design (CAD), computer-aided engineering (CAE) or computer-aided manufacturing (CAM), or hybrid platforms. For existing platforms with large simulation scopes, to avoid extensive or costly changes, the unobtrusive principle may be incorporated through auxiliary scripts or applications that capture actions users take (Gopsill et al., 2016; Hu and Taylor, 2015; Jin and Ishino, 2006). For instance, Jin and Ishino (2006) presented the DAKA framework as a method for: monitoring CAD activity, capturing design events, and processing the events into design activities for analysis.

A critical decision when applying the LaDDR framework is deciding how action stream data will be stored. Two possible storage formats include relational databases or document storage. Document storage formats such as JSON and XML offer adaptable and flexible means to capture action streams into documents that automatically retain the sequence in which actions happened. Many programming languages support formats like JSON and XML allowing them to be used across most existing or new platforms. These formats are less structured, however. Relational databases offer a more structured way to store data, facilitating querying and analysis, but these formats require maintenance as the program evolves and new actions are added or removed. Moreover, if each entry in the database is an action taken by designers, this will result in a very sparse database and inefficient use of storage resources. These points should be taken into consideration when implementing the framework's principles. Extending the number of LaDDR platforms would expand the benefits of this approach to more design research and education efforts.

7 Conclusion and limitations

7.1 Limitations

While research using LaDDR platforms is an emerging approach with considerable potential for studying design activity, two major limitations exist. First, LaDDR platforms only capture actions within their environment. Actions, such as running cost estimates on a calculator or sketching ideas on scrap paper will not be captured. Given this limitation, effort should be made to limit the number of actions designers take outside of the platform or data quality may be damaged. Second, regarding LaDDR platforms themselves, developing a LaDDR platform or modifying an existing platform to implement the LaDDR framework will require additional resources for research. There are existing LaDDR platforms and other platforms, such as the open-source FreeCAD (Falck and Colette 2012), which can be modified to operate like a LaDDR platform but employing any of these will require time investment.

7.2 Conclusion

The primary innovation proposed by this work is the Large Data for Design Research (LaDDR) framework, a set of design principles that can be applied to a design platform to allow for the study of design activities. The design principles are: 1) broad simulation scope, 2) unobtrusive logging of designers' actions, and 3) enabling both creation and analysis design actions types. These principles lead to three affordances for research and learning: 1) capture of design activity, 2) context setting and operationalization, and 3) research design scalability. Collectively, these create a big data-inspired approach to design and designer research.

We used three case studies to illustrate the variety of ways research with a LaDDR platform may manifest. The discussion focused on what LaDDR platforms can bring 1) research design, 2) the study of human-technology collaboration, and 3) how this approach can help address two key methodological and technology and training challenges present in design research.

The framework's design principles are intentionally generic to make them highly transportable to other platforms used to study designers and design learning and broaden the impact of this methodological innovation. Moreover, LaDDR platforms provide a toolset for design researchers and educators that opens new directions for addressing the aforementioned challenges and enhancing how design is taught, studied, and supported in practice. Access to more designers' processes with different levels of experience through detailed action streams, facilitated by this approach, can help lead to more reliable quantitative indicators of design expertise and inform tools for supporting student's development as designers or practitioner's engagement in design, impacting design research, design education, and design practice.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Purdue University IRB. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

CS and MG worked collaboratively to develop the initial framework and its components. For writing CS was responsible for the introduction and literature review as related to big data and data-driven design. He was also responsible for the LaDDR framework section and description of Aladdin. For the cases, he wrote the Iteration and Design team section. He also lead the discussion and conclusion. MG was responsible for the design expertise section, the design sequence case and the summary/overview of the cases and table introducing them. She was also responsible for reviewing and revising all other sections.

Funding

This work was supported by the National Science Foundation (Grant Nos. DUE-1348530 and DUE-1348547: Large-Scale Research on Engineering Design Based on Big Learner Data Logged by a CAD Tool). Any opinions, findings, and conclusions or recommendations expressed in this paper, however, are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Acknowledgments

We would like to thank Senay Purzer, Robin Adams, and Charles Xie for their support and assistance throughout the larger project in which this manuscript is one outcome.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Adams, R. S., and Atman, C. J. (2000). "Characterizing engineering student design processes: An illustration of iteration," in *107th ASEE annual conference and exposition* (St. Louis, MI).
- Ahmed, S., Wallace, K. M., and Blessing, L. (2003). Understanding the differences between how novice and experienced designers approach design tasks. *Res. Eng. Des.* 14, 1–11.
- Alelyani, T., Yang, Y., and Grogan, P. T. (2017). "Understanding designers behavior in parameter design activities," in *29th international conference on design theory and methodology* (Cleveland, Ohio, USA: ASME), Vol. 7, V007T06A030. doi:10.1115/DETC2017-68335
- Atman, C. J., Adams, R. S., Cardella, M. E., Turns, J., Mosborg, S., and Saleem, J. (2007). Engineering design processes: A comparison of students and expert practitioners. *J. Eng. Educ.* 96, 359–379. doi:10.1002/j.2168-9830.2007.tb00945.x
- Atman, C. J., Chimka, J. R., Bursic, K. M., and Nachtmann, H. L. (1999). A comparison of freshman and senior engineering design processes. *Des. Stud.* 20, 131–152. doi:10.1016/S0142-694X(98)00031-3
- Ball, L. J., Ormerod, T. C., and Morley, N. J. (2004). Spontaneous analogising in engineering design: A comparative analysis of experts and novices. *Des. Stud.* 25, 495–508. doi:10.1016/j.destud.2004.05.004
- Bao, J., Zheng, X., Zhang, J., Ji, X., and Zhang, J. (2018). Data-driven process planning for shipbuilding. *Artif. Intell. Eng. Des. Anal. Manuf.* 32, 122–130. doi:10.1017/S089006041600055X
- Björklund, T. A. (2013). Initial mental representations of design problems: Differences between experts and novices. *Des. Stud.* 34, 135–160. doi:10.1016/j.destud.2012.08.005
- Boardman, J., and Sauser, B. (2006). "System of systems - the meaning of of," in *2006 IEEE/SMC international conference on system of systems engineering* (Los Angeles, California, USA: IEEE), 118–123. doi:10.1109/SYSOSE.2006.1652284
- Bostanabad, R., Chan, Y.-C., Wang, L., Zhu, P., and Chen, W. (2019). Globally approximate Gaussian processes for big data with application to data-driven metamaterials design. *J. Mech. Des.* N. Y. 141, 111402. doi:10.1115/1.4044257
- Boyle, I., Duffy, A. H. B., Whitfield, R. I., and Liu, S. (2009). "Towards an understanding of the impact of resources on the design process," in *ICED'09* (Stanford, CA, 323–334).
- Chen, C., Olajoyegbe, T., and Morkos, B. (2020). "The imminent educational paradigm shift: How artificial intelligence will reframe how we educate the next generation of engineering designers," in *2020 ASEE virtual annual conference content access proceedings (virtual on line: ASEE conferences)*, 35326. doi:10.18260/1-2-35326
- Chen, S. M., Zhang, Y., and Leung, V. (2014). *Big data related technologies, challenges and future prospects*. New York, NY: Springer. Available at: <http://link.springer.com/10.1007/s40558-015-0027-y> (Accessed February 23, 2019).
- Chiu, I., and Shu, L. H. (2010). "Potential limitations of verbal protocols in design experiments," in *22nd international conference on design theory and methodology; special conference on mechanical vibration and noise (montreal, quebec, Canada: ASME)*, Vol. 5, 287–296. doi:10.1115/DETC2010-28675
- Chiu, M.-C., and Lin, K.-Z. (2018). Utilizing text mining and Kansei Engineering to support data-driven design automation at conceptual design stage. *Adv. Eng. Inf.* 38, 826–839. doi:10.1016/j.aei.2018.11.002
- Chusilp, P., and Jin, Y. (2006). Impact of mental iteration on concept generation. *J. Mech. Des.* N. Y. 128, 14–25. doi:10.1115/1.2118707
- Clegg, B. A., DiGirolamo, G. J., and Keele, S. W. (1998). Sequence learning. *Trends Cogn. Sci.* 2, 275–281.
- Coley, F., Houseman, O., and Roy, R. (2007). An introduction to capturing and understanding the cognitive behaviour of design engineers. *J. Eng. Des.* 18, 311–325. doi:10.1080/09544820600963412
- Cornwell, B. (2015). *Social sequence analysis: Methods and applications*. New York, NY: Cambridge University Press.
- Crismond, D. P., and Adams, R. S. (2012). The informed design teaching and learning matrix. *J. Eng. Educ.* 101, 738–797. doi:10.1002/j.2168-9830.2012.tb01127.x
- Cross, N. (2004). Expertise in design: An overview. *Des. Stud.* 25, 427–441. doi:10.1016/j.destud.2004.06.002
- Cunningham, C., M., and Lachapelle, C. P. (2014). "Designing engineering experiences to engage all students," in *Engineering in pre-college settings: Synthesizing research, policy, and practices*. Editors S. Purzer, J. Strobel, and M. Cardella (West Lafayette, IN: Purdue University Press), 117–142.
- Deng, Y., Mueller, M., Rogers, C., and Olechowski, A. (2022). The multi-user computer-aided design collaborative learning framework. *Adv. Eng. Inf.* 51, 101446. doi:10.1016/j.aei.2021.101446
- Destrebecqz, A., Peigneux, P., Laureys, S., Degueldre, C., Del Fiore, G., Aerts, J., et al. (2005). The neural correlates of implicit and explicit sequence learning: Interacting networks revealed by the process dissociation procedure. *Learn. Mem.* 12, 480–490. doi:10.1101/lm.95605
- Dinar, M., Shah, J. J., Cagan, J., Leifer, L., Linsey, J., Smith, S. M., et al. (2015). Empirical studies of designer thinking: Past, present, and future. *J. Mech. Des.* N. Y. 137, 021101. doi:10.1115/1.4029025
- Dorst, K. (2004). On the problem of Design Problems - problem solving and design expertise. *J. Des. Res.* 4, 0. doi:10.1504/jdr.2004.009841
- Dym, C. L., Agogino, A. M., Eris, O., Frey, D. D., and Leifer, L. J. (2005). Engineering design thinking, teaching, and learning. *J. Eng. Educ.* 94, 103–120. doi:10.1002/j.2168-9830.2005.tb00832.x
- Dym, C. L., and Little, P. (2003). *Engineering design: A project-based introduction*. 2nd ed. New York, NY: John Wiley.
- Egan, P., Cagan, J., Schunn, C., and LeDuc, P. (2015). Synergistic human-agent methods for deriving effective search strategies: The case of nanoscale design. *Res. Eng. Des.* 26, 145–169. doi:10.1007/s00163-015-0190-3
- Falck, D., and Colette, B. (2012). *FreeCAD: Solid modeling with the power of Python*. Birmingham, UK: Packt Publishing Lt.
- Fu, K. K., Yang, M. C., and Wood, K. L. (2016). Design principles: Literature review, analysis, and future directions. *J. Mech. Des.* N. Y. 138, 101103. doi:10.1115/1.4034105
- Gajewski, R., and Pieniżek, P. (2017). Building energy modelling and simulations: Qualitative and quantitative analysis. *MATEC Web Conf.* 117, 00051. doi:10.1051/mateconf/201717100051
- Gero, J. S., and Peng, W. (2009). Understanding behaviors of a constructive memory agent: A Markov chain analysis. *Knowledge-Based Syst.* 22, 610–621. doi:10.1016/j.knsys.2009.05.006
- Gleicher, M., Albers, D., Walker, R., Jusufi, I., Hansen, C. D., and Roberts, J. C. (2011). Visual comparison for information visualization. *Inf. Vis.* 10, 289–309. doi:10.1177/1473871611416549
- Goldstein, M. H., Adams, R. S., and Purzer, S. (2021). Understanding informed design through trade-off decisions with an empirically-based protocol for students and design educators. *J. Pre-College Eng. Educ. Res. (J-PEER)* 11 (2). doi:10.7771/2157-9288.1279
- Goldstein, M. H. (2018). *Characterizing trade-off decisions in student designer*. Unpublished doctoral dissertation.
- Goldstein, M. H., Omar, S. A., Purzer, S., and Adams, R. S. (2018). "Comparing two approaches to engineering design in the 7th grade science classroom," in *International journal of education in mathematics, science and technology*, 381–397. doi:10.18404/ijemst.440340
- Goldstein, M. H., Purzer, S., Zielinski, M., and Adams, R. S. (2015). "High school students' ability to balance benefits and tradeoffs while engineering green buildings," in *122nd ASEE annual conference and exposition* (Seattle, WA).
- Gopsill, J., Snider, C., Shi, L., and Ben, H. (2016). "Computer aided design user interaction as a sensor for monitoring engineers and the engineering design process," in *Proceedings of the DESIGN 2016 14th International Design Conference* 12.
- Han, J., Kamber, M., and Pei, J. (2012). *Data mining: Concepts and techniques*. 3rd ed. Waltham, MA: Morgan Kaufmann.
- Hay, L., Duffy, A. H. B., McTeague, C., Pidgeon, L. M., Vuletic, T., and Grealy, M. (2017). Towards a shared ontology: A generic classification of cognitive processes in conceptual design. *Des. Sci.* 3, e7. doi:10.1017/dsj.2017.6
- Hilbert, D. M., and Redmiles, D. F. (2000). Extracting usability information from user interface events. *ACM Comput. Surv.* 32, 384–421. doi:10.1145/371578.371593
- Ho, C. H. (2001). Some phenomena of problem decomposition strategy for design thinking: differences between novices and experts. *Design Studies* 22, 27–45.
- Hu, Y., and Taylor, M. (2016). "WORK IN PROGRESS: A computer-aided design intelligent tutoring system teaching strategic flexibility," in *2016 ASEE Annual Conference & Exposition Proceedings* (New Orleans, Louisiana: ASEE Conferences), 27208. doi:10.18260/p.27208
- Jain, A., K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognit. Lett.* 31, 651–666. doi:10.1016/j.patrec.2009.09.011

- Jiao, R., Commuri, S., Panchal, J., Milisavljevic-Syed, J., Allen, J. K., Mistree, F., et al. (2021). Design engineering in the age of industry 4.0. *J. Mech. Des. N. Y.* 143, 070801. doi:10.1115/1.4051041
- Jin, Y., and Ishino, Y. (2006). Daka: Design activity knowledge acquisition through data-mining. *Int. J. Prod. Res.* 44, 2813–2837. doi:10.1080/00207540600654533
- Kavakli, M., and Gero, J. S. (2002). The structure of concurrent cognitive actions: A case study on novice and expert designers. *Des. Stud.* 23, 25–40. doi:10.1016/S0142-694X(01)00021-7
- Khan, S., and Awan, M. J. (2018). A generative design technique for exploring shape variations. *Adv. Eng. Inf.* 38, 712–724. doi:10.1016/j.aei.2018.10.005
- Kim, H., Liu, Y., Wang, Y., and Wang, C. (2016). Special issue: Data-driven design (D3). *J. Mech. Des. N. Y.* 138. doi:10.1115/1.4035002
- Kitchin, R. (2013). Big data and human geography: Opportunities, challenges and risks. *Dialogues Hum. Geogr.* 3, 262–267. doi:10.1177/2043820613513388
- Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data Soc.* 1, 205395171452848. doi:10.1177/2053951714528481
- Kruger, C., and Cross, N. (2006). Solution driven versus problem driven design: Strategies and outcomes. *Des. Stud.* 27, 527–548. doi:10.1016/j.destud.2006.01.001
- Kwon, E., Ryan, J. D., Bazylak, A., and Shu, L. H. (2020). Does visual fixation affect idea fixation? *J. Mech. Des. N. Y.* 142, 031118. doi:10.1115/1.4045600
- Landriscina, F. (2013). *Simulation and learning: A model centered approach*. New York, NY: Springer.
- Laney, D. (2001). *3D data management: Controlling data volume, velocity, and variety*. Meta Group Inc. Available at: <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>.
- Lawson, B., and Dorst, K. (2009). *Design expertise*.
- Lee, B., Cooper, R., Hands, D., and Coulton, P. (2022). Continuous cycles of data-enabled design: Reimagining the IoT development process. *Artif. Intell. Eng. Des. Anal. Manuf.* 36, e11. doi:10.1017/S0890060421000299
- Li, X., Jiang, Z., Guan, Y., Li, G., and Wang, F. (2019). Fostering the transfer of empirical engineering knowledge under technological paradigm shift: An experimental study in conceptual design. *Adv. Eng. Inf.* 41, 100927. doi:10.1016/j.aei.2019.100927
- Lin, X., Chien, C.-F., and Kerh, R. (2016). UNISON framework of data-driven innovation for extracting user experience of product design of wearable devices. *Comput. Ind. Eng.* 99, 487–502. doi:10.1016/j.cie.2016.05.023
- Lloyd, P., and Scott, P. (1994). Discovering the design problem. *Des. Stud.* 15, 125–140. doi:10.1016/0142-694X(94)90020-5
- Nagamachi, M. (2002). Kansei engineering as a powerful consumer-oriented technology for product development. *Appl. Ergon.* 33, 289–294. doi:10.1016/S0003-6870(02)00019-4
- McComb, C., Cagan, J., and Kotovsky, K. (2017). Capturing human sequence-learning abilities in configuration design tasks through Markov chains. *J. Mech. Des. N. Y.* 139, 091101. doi:10.1115/1.4037185
- Miller, H. J. (2010). The Data Avalanche is here. Shouldn't we be Digging? *J. Reg. Sci.* 50, 181–201. doi:10.1111/j.1467-9787.2009.00641.x
- National Academy of Engineering and National Research Council (2009). *Engineering in K-12 education: Understanding the status and improving the prospects*. Washington, DC: The National Academies Press. doi:10.17226/12635
- National Academy of Engineering and National Research Council (2014). *STEM integration in K-12 education: Status, prospects, and an agenda for research*. Washington, DC: The National Academies Press. doi:10.17226/18612
- National Research Council (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: The National Academies Press. doi:10.17226/13165
- Phadnis, V., Wallace, D., and Olechowski, A. (2021). A multimodal experimental approach to study CAD collaboration. *Comput. Aided. Des. Appl.* 18, 328–342. doi:10.14733/cadaps.2021.328-342
- Pietsch, W. (2016). The causal nature of modeling with big data. *Philos. Technol.* 29, 137–171. doi:10.1007/s13347-015-0202-2
- Purzer, Ş., Goldstein, M. H., Adams, R. S., Xie, C., and Nourian, S. (2015). An exploratory study of informed engineering design behaviors associated with scientific explanations. *Int. J. STEM Educ.* 2, 9. doi:10.1186/s40594-015-0019-7
- Qiu, J., Wu, Q., Ding, G., Xu, Y., and Feng, S. (2016). A survey of machine learning for big data processing. *EURASIP J. Adv. Signal Process.* 67, 67. doi:10.1186/s13634-016-0355-x
- Rahman, M. H., Schimpf, C., Xie, C., and Sha, Z. (2019a). A computer-aided design based research platform for design thinking studies. *J. Mech. Des. N. Y.* 141, 121102. doi:10.1115/1.4044395
- Rahman, M. H., Xie, C., and Sha, Z. (2019b). “A deep learning based approach to predict sequential design decisions,” in *39th computers and information in engineering conference* (Anaheim, California, USA: American Society of Mechanical Engineers (ASME)), Vol. 1. V001T02A029. doi:10.1115/DETC2019-97625
- Rahman, M. H., Yuan, S., Xie, C., and Sha, Z. (2020). Predicting human design decisions with deep recurrent neural network combining static and dynamic data. *Des. Sci.* 6, e15. doi:10.1017/dsj.2020.12
- Raina, A., McComb, C., and Cagan, J. (2019). Learning to design from humans: Imitating human designers through deep learning. *J. Mech. Des. N. Y.* 141, 111102. doi:10.1115/1.4044256
- Razzouk, R., and Shute, V. (2012). What is design thinking and why is it important? *Rev. Educ. Res.* 82, 330–348. doi:10.3102/0034654312457429
- Ritchie, J. M., Sung, R. C. W., Rea, H., Lim, T., Corney, J. R., and Howley, I. (2008). “The use of non-intrusive user logging to capture engineering rationale, knowledge and intent during the product life cycle,” in *Picmet '08 - 2008 portland international conference on management of engineering and technology* (Cape Town, South Africa: IEEE), 981–989. doi:10.1109/PICMET.2008.4599707
- Schimpf, C., Goldstein, M., Adams, R., Chao, J., Purzer, S., and Xie, C. (2018a). “Work in progress: A Markov chain method for modeling student behaviors,” in *2018 ASEE annual conference and exposition proceedings* (Salt Lake City, Utah: ASEE Conferences), 31262. doi:10.18260/1-2-31262
- Schimpf, C., Huang, X., Xie, C., Sha, Z., and Massicotte, J. (2019). “Developing instructional design agents to support novice and K-12 design education,” in *2019 ASEE annual conference and exposition proceedings* (Tampa, Florida: ASEE Conferences), 32640. doi:10.18260/1-2-32640
- Schimpf, C., Sleezer, R., and Xie, C. (2018b). “Work in progress: Visualizing design team Analytics for representing and understanding design teams’ process,” in *2018 ASEE annual conference and exposition proceedings* (Salt Lake City, Utah: ASEE Conferences), 31316. doi:10.18260/1-2-31316
- Schimpf, C., and Xie, C. (2017). “Characterizing students’ micro-iterations strategies through data-logged design actions,” in *2017 ASEE annual conference and exposition proceedings* (Columbus, Ohio: ASEE Conferences), 28027. doi:10.18260/1-2-28027
- Shi, Y., and Peng, Q. (2021). Enhanced customer requirement classification for product design using big data and improved Kano model. *Adv. Eng. Inf.* 49, 101340. doi:10.1016/j.aei.2021.101340
- Siemens, G. (2013). Learning analytics: The emergence of a discipline. *Am. Behav. Sci.* 57, 1380–1400. doi:10.1177/0002764213498851
- Sim, S. K., and Duffy, A. H. B. (2003). Towards an ontology of generic engineering design activities. *Res. Eng. Des.* 14, 200–223. doi:10.1007/s00163-003-0037-1
- Simon, H. (1996). *The sciences of the artificial*. 3rd ed. Cambridge, MA: The MIT Press.
- Sivanathan, A., Lim, T., Ritchie, J., Sung, R., Kosmadoudi, Z., and Liu, Y. (2015). The application of ubiquitous multimodal synchronous data capture in CAD. *Computer-Aided Des.* 59, 176–191. doi:10.1016/j.cad.2013.10.001
- Smith, R. P., and Tjandra, P. (1998). Experimental observation of iteration in engineering design. *Res. Eng. Des.* 10, 107–117. doi:10.1007/BF01616691
- Song, B., Soria Zurita, N. F., Zhang, G., Stump, G., Balon, C., Miller, S. W., et al. (2020). Toward hybrid teams: A platform to understand human-computer collaboration during the design of complex engineered systems. *Proc. Des. Soc. Des. Conf.* 1, 1551–1560. doi:10.1017/dsd.2020.68
- Sonnenberg, C., and Bannert, M. (2016). *Evaluating the impact of instructional support using data mining and process mining: A micro-level analysis of the effectiveness of metacognitive prompts*, 33.
- Štorga, M., Andreassen, M. M., and Marjanović, D. (2010). The design ontology: Foundation for the design knowledge exchange and management. *J. Eng. Des.* 21, 427–454. doi:10.1080/09544820802322557
- Summers, J. D., and Shah, J. J. (2010). Mechanical engineering design complexity metrics: Size, coupling, and solvability. *J. Mech. Des. N. Y.* 132, 021004. doi:10.1115/1.4000759
- Sung, R. C. W., Ritchie, J. M., Lim, T., and Kosmadoudi, Z. (2012). Automated generation of engineering rationale, knowledge and intent representations during the product life cycle. *Virtual Real.* 16, 69–85. doi:10.1007/s10055-011-0196-8
- Sung, R., Ritchie, J. M., Rea, H. J., and Corney, J. (2011). Automated design knowledge capture and representation in single-user CAD environments. *J. Eng. Des.* 22, 487–503. doi:10.1080/09544820903527187
- The National Science Foundation (2020). *Future of work at the human-technology frontier*.
- Vieira, C., Hathaway Goldstein, M., Purzer, Ş., and Magana, A. J. (2016). Using learning analytics to characterize student experimentation strategies in the context of engineering design. *Learn. Anal.* 3, 291–317. doi:10.18608/jla.2016.33.14

- Wu, X., Zhu, X., Wu, G.-Q., and Ding, W. (2014). Data mining with big data. *IEEE Trans. Knowl. Data Eng.* 26, 97–107. doi:10.1109/TKDE.2013.109
- Wynn, D. C., and Eckert, C. M. (2017). Perspectives on iteration in design and development. *Res. Eng. Des.* 28, 153–184. doi:10.1007/s00163-016-0226-3
- Xie, C., Schimpf, C., Chao, J., Nourian, S., and Massicotte, J. (2018). Learning and teaching engineering design through modeling and simulation on a CAD platform. *Comput. Appl. Eng. Educ.* 26, 824–840. doi:10.1002/cae.21920
- Xie, C. (2016). The JSON data schema that encodes Energy3D design processes. Available at: <http://energy.concord.org/energy3d/schema/schema.pdf>.
- Xiong, Y., Duong, P. L. T., Wang, D., Park, S.-I., Ge, Q., Raghavan, N., et al. (2019). Data-driven design space exploration and exploitation for design for additive manufacturing. *J. Mech. Des. N. Y.* 141, 101101. doi:10.1115/1.4043587
- Yilmaz, S., Daly, S. R., Seifert, C. M., and Gonzalez, R. (2016a). Evidence-based design heuristics for idea generation. *Des. Stud.* 46, 95–124. doi:10.1016/j.destud.2016.05.001
- Yilmaz, S., Seifert, C., Daly, S. R., and Gonzalez, R. (2016b). Design heuristics in innovative products. *J. Mech. Des. N. Y.* 138, 071102. doi:10.1115/1.4032219
- Zhang, G., Raina, A., Cagan, J., and McComb, C. (2021). A cautionary tale about the impact of AI on human design teams. *Design Studies* 72, 100990. doi:10.1016/j.destud.2021.100990