HOW TO SUPPORT SITUATED DESIGN EDUCATION THROUGH AI-BASED ANALYTICS

A Dissertation

by

AJIT JAIN

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Co-Chairs of Committee,	Andruid Kerne
	Ruihong Huang
Committee Members,	John Keyser
	James Caverlee
	Steven M. Smith
	Jinsil Seo
Head of Department,	Scott Schaefer

May 2021

Major Subject: Computer Science

ABSTRACT

Design education is essential to pedagogy across fields. It helps develop creative skills and abilities. To meet growing demand for design education, researchers are investigating novel means for assessing and providing feedback on students' design projects. To complement instructors' work, we investigate the potential of artificial intelligence (AI), which has matched humans in performing complex tasks and activities.

We build on prior research by Suchman and Dourish in showing how developing useful AI support requires understanding *situated practice*. We engage instructors in co-design. In theorizing, we twice invoke creative cognition's family resemblance principle, first, to contribute new understandings of uses and limitations of design rubrics, and then to identify an analogous role for *AI-based design creativity analytics*: assessing no particular characteristic is necessary or sufficient; each only tends to indicate good design work. We contribute *situating analytics*, a paradigm for conveying the meaning of measures that align with design rubrics, by contextually integrating the presentation of measures with associated design work.

We develop results across fields. We integrate the measurement and presentation of multiscale design characteristics that provide insights into students' use of space and scale. We find that situating analytics supports instructors in understanding what the measures mean. Through quantitative analysis, we establish the baseline performance for AI recognition of multiscale design characteristics. Through qualitative analysis of instructors' experiences in situated course contexts, we derive implications for conveying the meaning of measures.

DEDICATION

To my wife and parents, for their endless love and all the sacrifices they made.

ACKNOWLEDGMENTS

I express my immense gratitude to my advisor, Dr. Andruid Kerne, for his continuous guidance and support. He instills in his students a passion for exploring new territories through interdisciplinary research, as well as the values of diversity and inclusion. I am certain that the knowledge, skills, and values I have gained through working with him will serve me throughout my career. I thank my co-advisor, Dr. Ruihong Haung, and committee members Drs. Steven M. Smith, John Keyser, James Caveralee, and Jinsil Seo, for the invaluable learnings and feedback I received from them. I would also like to thank Dr. Yoonsuck Choe for his great advice and feedback on various aspects. A special thanks to all design instructors who participated in this research for being supportive of my endeavors and sharing valuable perspectives; I cannot name you all due to study protocols, but without your support, this work would not have been possible.

Thanks to my labmate, Nic Lupfer, for providing useful suggestions and help on multiple aspects; our graduate school journeys ran almost parallel, and I hope our collaborations continue in the coming time. Thanks to undergraduate researchers in the lab—Hannah Fowler, Gabriel Britain, and Aaron Perrine—for bringing extraordinary skills and helping immensely with this work. Thanks to my labmates Rhema Linder, Yin Qu, William Hamilton, and Andrew Webb for the enjoyable time I spent with them both at work and outside.

Thanks to my entire family and all my friends for their continuous love, encouragement, and support. Finally, thanks to the Almighty for making this work possible.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professors Andruid Kerne, Ruihong Huang, John Keyser, and James Caverlee of the Department of Computer Science and Engineering, Professor Steven M. Smith of the Department of Psychology, and Professor Jinsil Seo of the Department of Visualization.

The data analysis in Chapter 3 was in part conducted by undergraduate researcher Hannah Fowler. The analytics dashboard and its integration with the multiscale design environment for supporting student submission of design work, in Chapter 4, were in part developed by undergraduate researchers Gabriel Britain and Aaron Perrine and graduate researcher Nic Lupfer. The multiscale design environment was developed under the leadership of Nic Lupfer, previously working in collaboration with graduate researcher William Hamilton. The ground truth labeling and qualitative analysis of data from instructors' interviews in Chapter 4 were in part performed by Nic Lupfer. The quantitative data analysis in Chapter 4 was in part supervised by Dr. Yoonsuck Choe of the Department of Computer Science and Engineering.

Funding Sources

The doctoral study was primarily supported by teaching assistantships from the Department of Computer Science and Engineering.

This work was made possible in part by the National Science Foundation under grants IIS-1247126 and IIS-1528044. Any opinions, findings, and conclusions or recommendations expressed in this work are those of the author and do not necessarily reflect the views of the NSF.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

1. INTRODUCTION

Design education has become an essential part of diverse fields—such as engineering, architecture, medicine, and business—as it helps develop creative skills and capabilities that can drive innovation [4]. However, due to high demands, the practice of teaching creative innovation through design has become increasingly difficult to manage. Design instructors face challenges to provide students with frequent assessment and feedback, which are critical to students' progress [5]. Alternative forms of human assistance that have proven useful in supporting instructor efforts include peer [6, 7] and crowd [8, 9, 10]. The present research investigates artificial intelligence (AI)—which has the potential to complement human assistance by providing the ability to process big data at speed [11]—as a means for assessment and feedback.

We draw on Suchman's seminal work on the need for an AI system's transparency, which allows conveying its intended purpose to the users and establishing its accountability [12]. Suchman articulates that developing such AI support requires understanding users' *situated practice*, i.e., how their actions develop purpose and intelligibility within particular circumstances. Dourish further breaks this principle down to focus on how translating ideas between intellectually different domains of situated practice (social) and technology (computational) "can be both exceptionally valuable and unexpectedly difficult" [13]. Dourish argues that making contextual properties transparent is vital in addressing the difficulty, and as context acquires "meaning or relevance through their relationship to forms of practice", studying work practices becomes fundamentally important.

The present research takes a two-pronged approach to develop understandings of design instructors' situated assessment and feedback work practices. One, we survey prior work to develop understandings of assessment criteria, feedback processes, and sources—e.g., instructors, peer, crowd, and computational—of assessment and feedback (Chapter 2). Simultaneously, the present research engages co-design methods with 11 instructors from 5 fields—Landscape Architecture & Urban Planning, Architecture, Interactive Art & Design, Mechanical Engineering, and Computer Science and Engineering—to understand their situated assessment and feedback practices and needs (Chapter 3).

Based on the understandings from co-design engagements, to complement instructors' efforts, we develop ideas for *AI-based analytics* that align with design course rubrics. Like recent AI-based analytics research [14, 15], we adopt a broad view of analytics, which includes not only facts but also inferences. We contribute *situating analytics*, a paradigm for conveying the meaning of measures that align with design rubrics, by contextually integrating the presentation of measures with associated design work. Following this paradigm, we develop approaches for measuring and presenting Fluency, Flexibility, Visual Consistency, Multiscale Organization, and Legible Contrast in students' design work. To render the measures meaningful and actionable in instructors' and students' experiences, we develop ideas for situating their interaction with the analytics via dashboards integrated with actual design environments in which students perform their work.

We put the situating analytics paradigm into practice (Chapter 4). We develop support for *multiscale design*—a visual design characteristic—which refers to "the use of space and scale to explore and articulate relationships, [through] the juxtaposition and synthesis of diverse design elements" [16] (See examples in Chapter 4). We deploy a technology probe [17]—comprised by AI-based analytics that measure the use of space and scale, presented via a dashboard integrated with a multiscale design environment—in situated design course contexts, across fields. By integrating the dashboard and design environment, we connect the dashboard presentation of analytics with the particular sets of spatially related elements they measure. We contribute understandings of how situating analytics affects design instructors' assessment and feedback practices. We derive implications for conveying the meaning of analytics to users, in situated contexts of practice, so as to support teaching and learning.

In this chapter, we present sensitizing concepts that guided our investigation. We follow this with our approaches for developing understandings of design instructors' situated assessment and feedback practices and investigating the use of AI-based analytics in design course contexts. Finally, we present the main contributions of this research.

1.1 Sensitizing Concepts

Sensitizing concepts are background ideas that provide framing and guidance to researchers in an investigation [18, 19]. Our investigation draws and connects sensitizing concepts from research in design education, creative cognition, computational design, learning analytics, co-design, and multiscale design. These concepts guide our research methods, system design, interpretation of findings, and implications.

1.1.1 Design Pedagogy

Design education, across fields, is based on common pedagogical beliefs and practices [20]. As Dym et al. describe, design students engage in project-based learning, where they develop multidisciplinary ideas to address challenging problems [21]. To facilitate design process, instructors engage students in prototyping [22], divergent thinking [23, 20], extensive critique [24, 25], and working closely with end-users [23, 26]. In essence, design pedagogy inducts students into a *community of practice* [27], where they learn and develop skills, often through socially situated understandings of work that practitioners perform [28].

As Sawyer identifies, a creative process is central in design pedagogy, as students engage in developing novel solutions to challenging problems while drawing on precedents [20]. To support the creative nature of projects, instructors define learning outcomes, project assignments, and rubrics, but keep them open-ended and flexible [29]. However, students unfamiliar with the process struggle and need structure [30]. They need frequent assessment and feedback to make progress [25]. Alas, instructor feedback is not able to keep pace with increasing design education demands [5]. Further, instructors may not be available at critical times, e.g., late the night before an assignment is due. The present research develops novel, AI-based measures of design creativity to augment instructors' efforts and provide on-demand assessment and feedback.

1.1.2 Creative Cognition

Ideation is the creative process of developing and generating ideas. Design is a general activity oriented toward creative ideation [31]. Hence, understanding how to measure creativity can prove beneficial in developing approaches for assessing students' design work. At the same time, as creativity researchers discuss, "the exact question of what is creativity is often ignored or answered in too many different ways" [32]. Frich et al.'s recent survey recommends "looking to the well-established tradition of psychology-based creativity research" [31].

In cognitive psychology, creative cognition is a field of investigating processes and structures that contribute to creative thinking. Some creative processes are exploratory—e.g., attribute finding, conceptual interpretation, and functional inference—while others are generative—e.g., memory retrieval, association, and mental synthesis [33]. Creative structures are representations, such as novel visual patterns, verbal combinations, and mental blends. Building on creative cognition theory, researchers have derived and applied ideation metrics for measuring design creativity in particular contexts, including Fluency, Flexibility/Variety, Novelty, Emergence and Visual Presentation [34, 35].

Creative cognition avoids the epistemological trap of defining creativity in any absolute way. Alternatively, it presents the *family resemblance* principle, which states: while cognitive processes and structures are indicators, *no particular process or structure is necessary or sufficient for creative ideation*. The present research applies this principle to develop new understanding of how design instructors use assessment rubrics and theorize the validity of conceptually aligned AI-based analytics.

1.1.3 Computational Design

Computational approaches have proven beneficial for a supporting a variety of design processes, including idea generation [36, 37, 38], re-design [39, 40], and assessment [41, 42]. We focus on the use of computational approaches for design assessment and feedback.

Reinecke et al. assessed website aesthetics by developing a model based on features such as colors, number of words, and number of images [41]. Mackeprang et al. assess diversity by computing semantic similarity among ideas, using external knowledge graphs such as Wikidata and DBpedia [43]. Oulasvirta et al. assess a GUI design against a range of characteristics, such as the number of unique colors, color clustering, edge density, and whitespace [42]. However, only a limited number of systems make assessment transparent to the user. For example, Oulasvirta et al. inform the user the rationale behind the assessed characteristics, by including references to theoretical bases for ideas, and in some cases, visualizing the characteristic. However, these systems have not been investigated in design course contexts. As we argue in the introduction, developing computational support that provides transparent assessment and feedback in design course contexts requires understanding situated practice. The present research meets the gap at the intersection of two bodies of research: one that studies situated practice in design course contexts, and other that develops computational support for assessment of design characteristics.

Computational support for providing feedback on design has been investigated as well. Krause et al. crowdsourced labeled examples of feedback on student designs, and then used a natural language processing model to provide suggestions and improve crowdsourcing feedback on new designs [5]. Ngoon et al. developed a text classifier to categorize feedback as actionable, specific, and/or justified. They used the categorization to suggest examples to the reviewers, which improved the quality of feedback [44]. Unlike assessment systems mentioned further above, both these feedback systems were investigated in design course contexts. The present research makes complementary contributions by developing ideas for feedback sourced from AI assessment, as an alternative to the crowd, as well as by identifying and addressing challenges in tracking students' incorporation of feedback.

1.1.4 Learning Analytics and Dashboards

Learning analytics have proven useful in providing instructors means for scaling feedback [45, 46] and informing pedagogical action [47, 48]. They have supported students' self-reflection [49, 50] and skill development [51, 52]. Example learning analytics in lecture-based contexts include the number of times a student accessed a resource, time spent, and length of textual annotations, which have been found to assist instructors in evaluating student understanding [53]. Likewise, dashboards have proven effective in lecture-based contexts, providing a quick understanding of student progress through representations such as tables and graphs [54, 55].

As we described above, design education involves project-based learning. As Blikstein discusses, for analytics to be useful in project-based contexts, they need to measure complex characteristics, such as creativity and the ability to address ill-defined problems [56]. In design course contexts, Britain et al.'s study surfaced this need, as they presented Fluency analytics—i.e., the number of elements, words, and images—to design instructors [57]. While the instructors found Fluency useful in gaining insight into students' efforts across various dimensions, they desired more sophisticated analytics. The present research identifies the underlying challenges and develops ideas for building analytics and dashboards suited for project-based design education.

1.1.5 Co-Design

Co-Design refers to "the creativity of designers and people not trained in design working together in the design development process" [58]. Sanders and Stappers map a repertoire of co-design approaches across two intersecting dimensions: one defined by participatory vs expert mindset and the other by design-led vs research-led approach. User-centered design—including techniques such as usability testing [59] and ergonomics [60]—forms the majority of research-led approaches with an expert mindset. Participatory design includes both design-led—e.g., enactments [61] and generative [62] tools—and research-led—e.g., Scandinavian [63]—approaches, which together form the majority of participatory mindset dimension. While probes [64] started out as a part of critical design, which is a design-led approach with expert mindset, they have evolved to take participatory mindset forms. As Mattelmäki notes, probes serve as tools for provoking discussions, gathering data, and directing further research [65].

We use a variety of co-design approaches for involving users at different stages of research. To understand users' practices and needs, we use approaches similar to the AT project [66], including individual discussions, followed by a workshop. Workshops provide a relatively neutral space, in which designers and users openly exchange ideas and opinions [67]. This facilitates co-creation of shared meanings and goals. We use the technology probes methodology [17] to understand how novel AI-based assessment of students' design work affects instructors' situated practices. The technology probes methodology allows simultaneously: (a) understanding users' needs and desires in a real-world setting, (b) collecting data through field-testing the technology, and (c) stimulate users' and researchers' thinking about technological possibilities.

1.1.6 Multiscale Design

As we introduced above, multiscale design refers to the use of space and scale to develop relationships among diverse design elements [16]. It facilitates a hierarchical or nested organization of ideas, which has key importance in design. As Alexander describes, designers employ hierarchies to organize each design component both as a unit and as a pattern consisting of other units, and thus address the important challenge of building up a form from components [68]. Multiscale design allows people to shift cognitive point of view up and down hierarchies, which supports a holistic analysis and development of ideas [69].

Multiscale design environments, such as Photoshop, Illustrator, and IdeaMâché [70], provide a free-form zoomable space, which allows designers to assemble visual elements, as ideas, into nested clusters. The *visual chunking* of this clustering has been shown to help overcome limits of human working memory [71]. As designers engage in thinking about and working with space and scale, they invoke multiscale design strategies, such as *map* and *shift perspective* [72, 16]. Map refers to defining spatial relationships among design elements through scale transitions: zooming out shows "encompassing contexts" and zooming in shows "nested details". Shift perspective refers to using "views from different zoom levels and angles to gain a wider understanding of a context" [16].

Prior work has assessed multiscale design characteristics, such as visual chunking and levels of nesting [70, 35]. However, human raters performed these assessments, which therefore are hard to scale. The present research develops computational approaches to measure multiscale design characteristics.

1.2 Understanding Situated Assessment and Feedback Practices

Supporting design education with AI is challenging, as it requires translating ideas between intellectually different social and computational domains [13]. This interface border zone [73], where the domains intersect, is fundamentally sociotechnical. Drawing on the fundamental work of

Suchman [12] and Dourish [13], we find that traversing the social / AI border requires identifying and building with contextual properties, derived through understanding situated practice. The situated practices that we focus on take place in this border zone, involving instructors' design assessment and feedback. We survey prior work and engage co-design methods with instructors across diverse fields to develop understandings of situated practice.

1.2.1 Prior Work Studying Practice

Design education, across fields, is found to share pedagogical beliefs and practices [20]. Through surveys of prior work studying assessment and feedback practices, researchers have developed understandings, such as instructor preferences for using various criteria [74] and how critique of design work takes place across settings, modalities, and design phases [25]. These surveys demonstrate how a systematic literature review can provide new understandings and analyses of practices across various dimensions.

Taking inspiration, we conducted a survey to develop understandings of specific approaches that instructors use to assess and provide feedback on a range of design characteristics, across fields. This, in turn, had the potential to help us identify "contextual properties" of situated assessment and feedback, which is vital for informing the design of computational support. To this end, we developed a keyword search based methodology, using which we formed a corpus of 60 papers. We performed a qualitative analysis of ideas in these papers, using Charmaz' approach to grounded theory [18], to develop codes and categories, aka facets.

Our analysis contributes facets of: (1) assessment criteria, (2) feedback modalities, (3) process vs. product focus, and (4) sources—e.g., instructors, peer, crowd, and computational—of assessment and feedback. We identify assessment approaches that have proven useful for criteria, such as variety and depth of solution space. We derive implications for computationally assessing and providing feedback on design work, including the need for going beyond correlation and regression analyses and deriving measures for a range of design characteristics. Implications also include the need for undertaking studies that provide insights into not only *what* criteria instructors use, but also *how* they apply these criteria, in practice. We present the approach, findings, and implications in detail in Chapter 2.

1.2.2 Co-Design with Design Instructors

Users are 'experts of their experience' [75]. Their expertise often involves forms of tacit and shared knowledge and communication [76]. Co-design methods support designers in the elicitation of users' tacit expertise [75]. This elicitation, i.e., making the tacit *visible*, is crucial to designing technology support for users' work practices [77]. As we noted, prior work lacks in making the tacit visible, in particular, *how* instructors apply assessment criteria, in practice. Hence, the present research engages co-design methods to make the "how" visible and derive implications for technology support.

Similar to the approach used in AT project [66], we first engaged individual discussions with design instructors from diverse fields. We began by contacting instructors whose work we were aware of from previous collaborations. Based on their recommendation, we performed snowball sampling [78], motivating new participants by identifying explicit, immediate value to them, such as opportunities to reflect on their own processes and share perspectives with peers. After that, we brought instructors together for a workshop for an interdisciplinary discussion of assessment and feedback practices, as well as how technology can support them. We witnessed periodic sighing and murmurs of enthusiasm; participants developed shared meanings, identifying common pain points and sharing ideas for addressing them. We analyzed qualitative data gathered from individual and workshop discussions using Charmaz' approach [18] to grounded theory.

In theorizing, we invoke creative cognition's family resemblance principle [33] to explain the uses and limitations of design rubrics: no particular characteristic is necessary or sufficient; together they tend to indicate quality design work. We again invoke the family resemblance principle to explain the analogous role for AI-based design creativity analytics—Fluency, Flexibility, Visual Consistency, Multiscale Organization, and Legible Contrast—that align with instructors' criteria, i.e., they have the potential to indicate good design work, without being necessary or sufficient.

In the situated contexts of design courses, AI-based analytics and instructors' and students' interaction with them become inseparable. Hence, we contribute situating analytics, a paradigm for explaining AI-based analytics to users by integrating their measurement and presentation. For this, we develop ideas for connecting the dashboard presentation of analytics with the particular design elements and assemblages they measure. Further, we develop ideas for making the AI integration controllable, providing affordances for instructors and students to make actionable changes, when AI outcomes do not make sense to them. We present the investigation approach, findings, and implications in detail in Chapter 3.

1.3 Situating Analytics Technology Probe

We use the situating analytics paradigm—developed based on the understandings from co-design engagements—to guide our development of AI-based analytics and explaining them to the users. We continue to take a co-design approach, this time in the form of a technology probe [17].

We focused on multiscale design, which involves the "use of space and scale to explore and articulate relationships" among diverse design elements [16]. We adapted a spatial clustering algorithm, AMOEBA [3], to derive novel AI-based analytics: the number of scales and clusters in a design. *Scale* refers to the zoom level: elements at the same scale are equally legible at the same viewport zoom [70]. *Cluster* refers to a set of elements that are spatially proximal to each other, in comparison to distances with other elements in the design. We presented the analytics to instructors via a dashboard integrated with a multiscale design environment in which students performed their work. The dashboard connects analytics with the particular sets of spatially related elements within the actual design work, thus showing the characteristics that analytics measure, in context.

We performed a mixed methods study. We quantitatively evaluated the AI performance, by deriving measures of precision, recall, and F-score. We also identified patterns where the AI does not match humans through a qualitative analysis based on visual inspection. We interviewed nine instructors about experiences with the integrated dashboard and design environment. We analyzed qualitative data from interviews using Charmaz' approach [18] to grounded theory.

We found that our approach to integrating measurement and presentation of analytics explains them to instructors. Instructors observed the analytics' potential as rubric elements and sources of reflection for students. We discuss how ambiguity in the interpretation of design work becomes a resource for explaining what analytics and AI mean. We derive implications for improving the AI performance, as well as explaining analytics through techniques, such as the animation of design characteristics within the actual design environment, the indexicality of dashboard representation of analytics, and exemplification to demonstrate a range of analytics measures. We present the investigation approach, findings, and implications in Chapter 4.

1.4 Research Contributions

We contribute new understandings of situated assessment and feedback practices, through our survey and co-design engagements. We develop implications for developing computational support, focusing on the use of AI, to support situated practice. Aligned with the implications, we develop novel AI-based multiscale design analytics and an approach to explaining them. We investigate the use of novel AI-based analytics in situated course contexts. The main contributions of the present research are:

- a new understanding of uses and limitations of project-based learning rubrics through creative cognition's family resemblance principle;
- situating analytics, as a paradigm for conveying the meaning of measures that align with design rubrics, by contextually integrating the presentation of measures with associated design work;
- a novel approach to measuring multiscale design, by adapting AMOEBA, a multilevel spatial clustering algorithm;
- establishing baseline performance for AI recognition of multiscale design characteristics, through a quantitative evaluation based on precision, recall, F-score measures; and
- implications for conveying the meaning of measures, based on a qualitative analysis of instructors' experiences with multicale design analytics in situated course contexts.

2. DESIGN EDUCATION ASSESSMENT AND FEEDBACK PRACTICES: A SURVEY

In this chapter, we perform a survey of prior work addressing practices—developing facets across multiple dimensions—so as to advance assessment and feedback in design education, involving new forms of computation. In design education, assessment and feedback are significant in helping students make progress [25]. Although design instructors often work independently, in differing disciplinary contexts, their practices share commonalities in assessment and feedback [79]. Assessment plays a pivotal role, as a territory in which the valuable creative idiosyncrasies of independent instructors become systematized as rubrics of design education accreditation [80, 81]. Feedback is a significant mechanism for advancing design students' development of expertise [25]. It transforms a student's point of view to be more inclusive and integrative [82].

As design education demands continue to grow, there is a need for scaling assessment and feedback [5]. A systematic understanding of practices has the potential to inform design of novel support. Prior surveys have compared criteria for assessing student work across the disciplines [74] and feedback practices across dimensions, such as critique settings, teacher-student relationships, communication modalities, and design phases [25]. However, these surveys do not detail approaches used for assessing and providing feedback on various design characteristics. We address the gap and focus on informing design of computational support. Our research questions are:
RQ1: What approaches are used for assessing and providing feedback on various design characteristics, across fields?
RQ2: How, if at all, can computational means support assessing and providing feedback on various design characteristics?

We form a corpus of 60 prior works addressing design assessment and feedback practice, then develop facets for systematic analysis. Among these facets, we investigate the diversity of design criteria, and the modalities and sources of feedback. This enables us to identify potential assessment and feedback opportunities, for sets of criteria, which can be useful for instructors in structuring their design courses. Further, it enables identifying gaps in current understandings of the usefulness of approaches—such as computation and crowd—for assessing and providing feedback on particular design characteristics.

We begin by examining prior surveys on assessment and feedback practices in design education. We then describe our methodology of how to identify prior work to survey. Next, we develop codes and categories through analysis of these papers, using a constant comparative method. We then use the codes and categories to define facets for conceptualizing and theorizing implications for the design of design education, focusing on assessment and feedback.

2.1 Prior Surveys

We discuss prior surveys that contribute toward developing understandings of assessment and feedback practices, as well as distinguish the present research from the prior surveys.

Sawyer's review [29], focusing on pedagogical practices in design studios, includes two themes on assessment: 'Assessment through feedback and critique' and 'Use of rubrics'. In the first theme, he discusses that assessments are designed toward guiding students on their own path rather than being performed for a course grade and that design instructors emphasize both process and product. Further, he notes a range of student perspectives on the critique environment, from encouraging and supportive to valuable yet stressful to frustrating and unsafe. In the second theme, he discusses complex viewpoints on the use of rubrics: "useful in demonstrating objectivity, but they don't capture the richness of a student's work." While the survey provides useful insights, assessment practices only received limited attention. 9 out of 11 themes focus on practices other than the assessment, e.g., open-ended and flexible pedagogy, tension between the open-ended nature and the need for structure, and students' confusion about the learning outcomes.

Two prior surveys specifically focus on design assessment practices. One is de la Harpe et al.'s systematic review of articles focused on assessing learning in design, architecture, and art studios [74]. The other is Oh et al.'s review of design critiquing in architecture studios [25].

We note, epistemologically, that de la Harpe et al. consider design, art, and architecture, as separate fields. Oh et al. investigate design as an aspect of architecture. The approach of the present research is to consider design as a sort of meta field, a transdisciplinary way of working, which spans constituent specific fields, such as mechanical engineering, computer science, interactive art, and architecture.

de la Harpe et al. formed a corpus of 118 papers related to design learning and assessment [74]. Twelve categories emerged from their comparative analysis of assessment focus in design, architecture, and art studios: product, process, person, knowledge/content, hard skills, soft skills, learning approach/style, technology, reflective practice, professional & innovative practice, inter-disciplinary collaboration, and participation/ attendance. While architecture had the highest number of papers focusing on product, design ranked the product as sixth and instead had process as the top focus. Other areas of focus in assessment in design include soft skills, hard skills, learning approach/style, and professional & innovative practice.

While the survey is a valuable resource for understanding comparative focus in different disciplines, as well as criteria that fall under different categories, it does not organize and analyze the papers by different criteria. Thus, it is not possible for a designer or researcher to understand assessment and feedback approaches used for a specific criterion, say aesthetics or originality. For example, it does not include sources or modalities used for assessment and feedback, which thus does not aid the process of finding tried-and-true methods with respect to a criterion.

Oh et al. reviewed literature from a number of sources, including journals and key conferences, for articles focusing on design critiquing in architecture [25]. However, they do not specify the number of articles reviewed. They identify 11 'fundamental factors' of design critiquing and organize them into two broad categories: methods and conditions. Methods refer to how studio instructors facilitate students' knowledge and skills development and includes 5 factors: critiquing settings, teacher-student relationships, communication modalities, delivery types, and delivery. Conditions refer to the context in which critique takes place and includes 6 factors: design phases, individual differences, knowledge/experiences, student response types, design artifacts, and learning goals.

Without going into the detail of each factor, we note that while we focus on some of the same aspects such as the modality of assessment and feedback, other aspects of the present research differ—e.g., focus on various criteria and sources used for their assessment—which makes our contribution complementary.

We note that prior surveys do not include a list of papers analyzed.

2.2 Survey Methodology

To form our corpus of 60 papers, we began by asking colleagues for their favorites that address assessment and feedback. We studied the favorites and performed citation chaining, forming an initial set of 19 seed papers. We expanded this set with keyword search.

We formulated a series of keyword based queries based on recurring ideas in the seed set. In this, our approach resembles de la Harpe's methodology of using a number of different keywords / descriptors [74]. We used Google Scholar for performing queries, as it allows searching across multiple peer-reviewed collections. Gehanno et al.'s study showed that Google Scholar searches provide 100% coverage, finding all the articles included in 29 systematic reviews, which were originally obtained via searching individual databases [83].

One of the major challenges we faced was largely irrelevant results produced by queries including keywords such as "design" and "education". This is because these keywords are quite general and occur in a variety of contexts. For example, for "design", results range from instructional design to study design to interface design. As we are interested in design education practices, we decided to use an exact phrase search by using quotes around "design education". We created additional keywords based on recurrent ideas in the seed papers and our focus on computational support.

To perform searches, we used "design education" as the first descriptor, followed by assessment OR evaluation OR feedback as the second, and zero or one of the following as third: 1) "artificial intelligence", 2) computational, 3) "iterative design", and 4) "creative design". As it can be noted, we used the phrase search method with other descriptors as well to minimize irrelevant results.

With the phrase queries, we obtained better results, but not all were focused on assessment and feedback. Many of them minimally referred to assessment and feedback, in the larger context of design pedagogy. For example, some of them mentioned descriptors such as "evaluation" and

Count	Publication Venue
10	Design Studies
4	ACM Conference on Human Factors in Computing Systems
4	International Journal of Technology and Design Education
3	International Journal of Art and Design Education
3	International Journal of Engineering Education
2	ACM Conference on Computer Supported Cooperative Work
2	ASEE Annual Conference and Exposition
1	ACM Conference on Designing Interactive Systems
1	ACM Interactions
1	ACM Transactions on Computer-Human Interaction
1	Annual Frontiers in Education Conference
1	Annual Meeting of the Cognitive Science Society
1	Architectural Science Review
1	ASME International Design Engineering Technical Conferences
1	Australasian Journal of Educational Technology
1	Automation in Construction
1	Bulletin of the IEEE Technical Committee on Learning Technology
1	Clothing and Textiles Research Journal
1	Communication Education
1	Computer Science Education
1	Computers
1	Computers & Education
1	Design and Technology Education
1	Design Issues
1	Educational Research Review
1	European Journal of Engineering Education
1	Higher Education Research & Development
1	IFIP International Conference on Human-Computer Interaction
1	International Conference on Computer-Aided Architectural Design Futures
1	International Journal for the Scholarship of Teaching and Learning
1	International Journal of Intelligent Systems Design and Computing
1	Journal of Architectural Education
1	Journal of Business and Technical Communication
1	Journal of Engineering Education
1	Journal of the Learning Sciences
1	Online Journal of Art and Design
1	Planning for Higher Education
1	Teaching in Higher Education
1	TechTrends

Table 2.1: We used Google Scholar to perform keyword search, which has been shown to have the same coverage as searching individual databases. Our corpus includes papers from diverse venues that publish design education research. The distribution simultaneously shows that there is a need for more focused publishing opportunities for design assessment and feedback research.

"feedback" only while pointing to practices in the introduction, and some included "design education" just in the discussion or conclusion when arguing for potential broader impact. Hence, we did a quick first pass over search results to identify whether each paper is relevant. We first looked at the abstract and, if needed, at the rest of the paper. We considered up to a maximum of 100 results for each query. We stopped early if 20 of them, i.e, 2 pages were consecutively irrelevant. We added 41 papers to the corpus through this process. Table 2.1 shows the list of publication venues for 60 papers in our corpus.

We used a constant comparative method to analyze the papers. We used Charmaz's approach [84], engaging in stages of initial coding, raising codes into categories, and then performing focused coding of the data. The first author inductively developed initial codes and categories from 10 papers. The first and second authors then met to revise the codes and categories to suitably represent the phenomena. The first author then engaged in focused coding of the rest of the papers, revising codes and categories, as needed, based on the emergence of phenomena.

2.3 Facets

The categories and codes, aka facets, represent dimensions, across which assessment and feedback practices can be structured, understood, and investigated. Our constant comparative analysis resulted in the following facets: Criteria, What/How, Modality, and Source. The Criteria facet focuses on the design characteristics on which emphasis is paid in design classrooms. The What/How facet focuses on the creative outcome and the process used to produce it. The Modality facet focuses on the modes or forms in which feedback is provided in design classrooms. The Source facet focuses on the channels that students receive feedback from.

2.3.1 Criteria

Design assessment encompasses a range of criteria. Firstly, design education pays emphasis on both process and product. Secondly, while there is a broad set of indicators, disciplinary differences exist in the focus and criteria of assessment [74]. Each design course is unique in its pedagogy and assessment.

In the initial coding phase, we organized criteria into three subcategories: conceptual, visual, and writing. However, as we continued the analysis, we found criteria that did not fall under these categories. Hence, we formed one new category: other. In Table 2.2, we indicate the category of

			Assessment and Feedback															
	Facet →		*Criteria*				*What/How*		*Modality*			*Source*						
Prior Work ↓	**Year**	**Field**	**Conceptual**	**Visual**	**Writing**	**Other**	**Process**	**Product**	**Physical**	**Digital**	**Verbal**	**Instructor**	**Peer**	**Jury**	**Client**	**Researcher**	**Crowd**	**Computation**
Watkins [85]	1988	Textile	●			●	●	●				●						
Barrett [86]	1990	Art and Design					●	●										
Sheppard et al. [87]	1997	Engr.,Graphics	●	●		●	●	●		●	●	●	●	●				
Adams & Atman [88] +	1999	Engineering				●												
Jung et al. [89]	1999	Architecture				●		●		●					●			
Ulusoy [90]	1999	Landscape						●	●		●		●					
Brusasco et al. [91]	2000	Architecture						●		●		●						
Ochsner [92]	2000	Architecture									●	●		●				
Cardella et al. [93] +	2002	Engineering				●	●											
Lewis & Bonollo [94] +	2002	Industrial	●			●	●	●							●			
Reimer & Douglas [95]	2003	HCI	●		●	●	●	●				●	●					
Shah et al. [34] +	2003	Engineering	●					●					●			●		
Zimmerman [96]	2003	HCI	●	●		●	●	●				●						
Blevis et al. [97]	2004	HCI				●	●											
Dym et al. [21]	2005	Engineering					●	●										
Cardella et al. [98]	2006	Engineering	●	●		●		●								●		
Adamczyk & Twidale [99]	2007	HCI	●				●	●		●			●					
Jeffries [100]	2007	Transport	●			●		●		●								●
Dannels et al. [24] +	2008	Several				●	●				●	●		●				
Dannels & Martin [101]	2008	Several					●	●				●	●	●				
Linsey et al. [102] +	2008	Engineering	●					●								●		
Harpe et al. [74]	2009	Art,Architecture	●	●		●	●	●										
Doppelt [103]	2009	Engineering	●			●	●	●						●				
Cai et al. [104] +	2010	Arch., Engr.	●					●						●				
Chance [105]	2010	Architecture				●	●											
Dow et al. [106] +	2010	Graphics	●	●		●								●				
Hanington [107]	2010	Industrial,HCI	●					●										
Lande & Leifer [108]	2010	Engineering													●			
Shreeve et al. [28]	2010	Several					●	●			●	●						
Dannels et al. [109]	2011	Several				●	●				●							
Ham & Schnabel [110]	2011	Architecture						●		●			●	●				
Kershaw et al. [111]	2011	Engineering	●					●			●					●		
Palmer & Hall [112]	2011	Engineering		●	●	●						●						
Park [113]	2011	Art and Design										●	●					
Crismond & Adams [23] +	2012	Engineering	●			●	●	●										
Kolko [82] +	2012	HCI		●				●			●	●						
McCarthy [114]	2012	Art,Architecture											●	●				
McLaren [115]	2012	Unspecified	●	●						●		●	●					
Schrand & Eliason [116]	2012	Several										●	●					
Wong & Siu [117]	2012	Design & Tech	●					●										
Dow et al. [8]	2013	HCI						●		●		●	●				●	
Oh et al. [25]	2013	Architecture	●				●	●	●	●	●	●	●	●				
Tinapple et al. [118]	2013	Art,Arch.,Engr.						●		●			●					
Yuan and Lee [119] +	2013	Industrial	●	●		●		●						●				●
Hui et al. [9] +	2014	Engineering				●											●	
Osmond & Tovey [30]	2015	Industrial									●	●						
Wei et al. [120]	2015	Creative	●					●								●		
Xu et al. [121] +	2015	Graphics		●				●		●				●			●	
Adams et al. [122]	2016	Several	●	●		●					●	●	●	●				
Christensen & Ball [123] +	2016	Industrial	●	●		●		●			●			●				
Yilmaz & Daly [79]	2016	Several	●	●		●	●	●			●	●	●					
Yuan et al. [10] +	2016	Graphics	●	●		●		●		●							●	
Arslan et al. [124]	2017	Architecture		●				●		●			●	●				●
Gray et al. [125]	2017	Art,Engineering	●			●						●						
Krause et al. [5] +	2017	Unspecified		●				●		●							●	
Sawyer [29]	2017	Art,Architecture					●	●				●						
Gray [126]	2018	HCI								●	●	●	●					
McDonald et al. [6]	2018	Instructional											●					
Sawyer [20] +	2018	Art and Design	●	●		●	●	●			●	●		●				
Soni et al. [127] +	2018	Industrial		●				●										●

Table 2.2: Each paper in our corpus mapped to the categories under the assessment and feedback facets. '+' indicates that the paper explains the assessment approaches for the criteria used. Criteria facet focuses on design characteristics. What/How facet focuses on design process vs outcome. Modality facet focuses on forms in which assessment and feedback are provided to students. Source facet focuses on who provides assessment and feedback.

Criteria	Process	Product	Instructor	Peer	Jury	Client	Researcher	Crowd	Computation	Unspecified
abstract thinking	•									[74]
addressing feedback +	•									[23]
adherence to client's theme		•	[79]	[79]	[106]					
aesthetics +		•	[122, 115]	[122, 115, 124]	[122, 123] [106, 119]		[98]		[127]	
analogical thinking +	•						[102]			
analytical thinking	•									[74]
autonomy	•		[125]							
class participation	•		[95]	[95]						
click-through +		•			[106]					
color, form, and composition +		•	[96]	[124]	[121]	[89]		[121, 10]		
complementary visuals and text +		•						[10]		
communication +	•		[24, 87]	[87]	[24, 87]	[94]				
computer aided drawing		•	[112]							
concept generation +	•		[87]	[87]	[87]	[94]				
delayed decision	•									[23]
design rationale		•	[25, 96]	[25]	[25]					
design reports		•	[112, 95]	[95]						
diversity of activities	•						[98]			
diversity of ideas +		•	[125, 25] [20, 87]	[25, 87, 124]	[104, 123, 103] [25, 20] [87, 119]		[34]		[100]	
elaboration		•	[79]	[79]			[120]			
exploratory searches +	•									[23, 74]
fluency or the no. of ideas +		•	[122]	[122]	[122]		[102, 34]			
functionality		•	[115]	[115, 124]	[123]					
idea selection	•	•	[20]		[103, 20]					
iterative design +	•									[88, 23]
negative space +		•	[20]		[20]					
novelty or originality +		•	[122]	[122]	[122, 106, 119]		[98, 120, 34]			
precedents		•	[25]	[25]	[25]					
presentation +	•	•	[112, 20, 85] [79, 96]	[79]	[20]			[9]	[124]	
problem framing		•	[122]	[122]	[122]					
process and creativity correlation +	•	•							[119]	
project management +	•					[94]				
purpose					[103]		[111]	[10]		
self-reflection	•		[87]	[87]	[103, 87]		[111]			
synthesis		•								[74, 120]
tastefulness		•			[106]					
teamwork	•		[87]	[87]	[87]					
technical feasibility	•		[115, 79]	[115, 79]			[111, 98]			
time management	•		[125]		[106]					
typeface +		•	[96]	[82]				[10]		
# of pages visited +		•			[106]					
# of visitors +		•			[106]					

Table 2.3: In order to facilitate gaining insights with respect to specific criteria, we map the criteria and publications that discuss them. In addition, we simultaneously map the What/How nature and the Source of assessment and feedback against each criterion. '+' indicates that for the given criterion, one or more papers explain assessment approaches. In order to make the representation concise, we merge a few criteria where they refer to same aspect, but use different name, for e.g., originality and novelty, variety and diversity, etc.

criteria. In Table 2.3, we develop a detailed representation, creating a grid of papers, with criteria on one axis and the source of feedback on the other. This serves as a ready resource that helps identify papers assessing a certain design criterion as well as how they assess it. If a paper explains the approach for assessing a criterion, then we indicate that with '+' in both the tables. Figure 2.1 shows relative distribution of papers that explain the approach vs. those that do not.

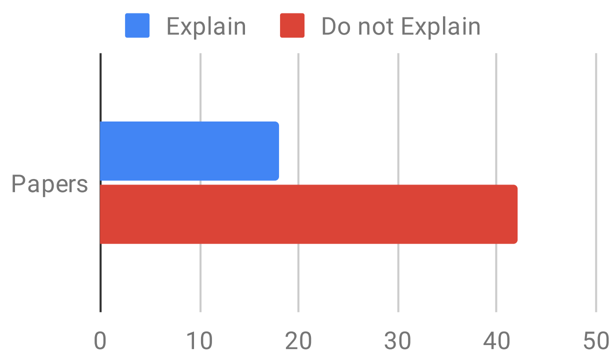

Figure 2.1: The chart shows the relative distribution of papers that explain the assessment approaches for the criteria used vs. those that do not. Only a small subset explains the assessment approaches.

Conceptual criteria focus on the development and generation of ideas. In our corpus, 28 out of 60 papers discuss conceptual criteria [85, 99, 100, 79, 125, 87, 102, 10, 74, 96, 25, 98, 23, 122, 115, 20, 117, 119, 120, 34, 123, 106, 95, 103, 94, 111, 107, 104]. Examples of conceptual criteria include variety or alternatives, originality, precedents, fluency or the number of ideas, idea selection, functionality, adherence to client's theme, problem framing, synthesis, analytical thinking, abstract thinking, analogical thinking, and technical feasibility.

A common problem we found is that students tend to fixate on a certain idea or set of ideas in the early stages of the design process [23, 111, 20]. Through feedback, from various sources, students gain new perspectives and start exploring new aspects or alternative solutions. This not only helps students in overcoming shortcomings of their initial solution, but also understand a variety of approaches that could be useful in different scenarios. Hence, the variety of ideas is a key assessment criterion [34, 104, 103, 20, 87, 125]. Likewise, originality has a major emphasis in design assessment [34, 106, 123, 120, 117, 122, 98]. Students work on "wicked" problems [128] that, by design, are constrained by requirements, time, and resources, and are ill-defined. Working on these problems require students to come up with creative solutions.

Visual criteria involve the effective presentation of ideas. In our corpus, 18 out of 60 papers address visual criteria [98, 122, 115, 119, 123, 127, 74, 96, 112, 20, 82, 5, 124, 121, 79, 106, 10, 87]. Examples of visual criteria include aesthetics, color, form, composition, typography, whitespace, negative space, visual representation and thinking, computer aided drawing, and facial details.

A common criterion is aesthetics, addressed by 8 papers [123, 115, 122, 98, 124, 106, 119, 127]. Aesthetic refers to the overall visual appeal of the page. While the evaluation of aesthetics can be subjective, adhering to various visual design principles—e.g., Gestalt laws of order, symmetry, and proximity—helps create an aesthetic design [129, 130]. In a similar direction, [74] finds in their survey that multiple design education contexts emphasize the use of visual elements and principles, such as color, form, and composition. In a study of 9 different design disciplines, [20] finds that instructors teach students to see negative space that gets created when putting together what is visible, such as the space between letters in typeface design or 3D space created by structures in architecture.

Writing criteria focus on the effective expression of ideas in the written form. Two papers discuss the writing aspects, mainly in the context of design reports [95, 112]. [95] discusses that students improve their written communication skills by engaging in documenting a report on their design solutions. [112] motivates the need of engaging students in a series of preliminary and final design reports, as project-based learning should provide opportunities for articulating the basis of design solutions.

Other criteria includes a number of aspects assessed in a design. We discuss a couple of criteria here. One is soft skills, which refer to the personality traits and interpersonal skills that enable effective interaction with others. Example criteria include communication, teamwork, autonomy, self-reflection, and time management. As [96] describes, students need to learn communicating the design, as well design rationale. As [74] states, design education has a major emphasis on soft skills of teamwork and collaboration. Another is a set of website access metrics, which include the number of clicks, visitors, and pages visited and the amount of time spent [106].

2.3.2 What/How

Design education puts emphasis on both the process and product [74]. "How" refers to the process through which a creative outcome is produced, and includes various aspects such as brainstorming, teamwork, reflection, and critique. "What" refers to the creative outcome, i.e., the product that the process has resulted in, and whose assessment is focused more on what is produced than how it was produced.

In the initial coding phase, we organized papers using these two subcategories: process and product. We did not form additional categories during the later stages of analysis.

Process refers to work practices, including soft skills that enable the effective development of a creative product. In our corpus, 22 papers discuss process criteria [74, 96, 20, 79, 87, 85, 99, 25, 23, 95, 103, 94, 86, 93, 97, 21, 24, 101, 105, 28, 109, 29]. Example criteria include presentations, self-reflection, iterative design, exploratory searches, delayed decision, incorporation of feedback, communication, and teamwork. As we briefly discussed soft skills, we focus on other aspects here. [95] discusses engaging students in weekly group design activities and assessing based on their presentation of artifacts, as well as participation during critiques and performance evaluation by peers. [101] breaks down communication, such as through presentations, into "competencies of interaction management, demonstration of design evolution, transparent advocacy of intent, explanation of visuals, and the staging of the performance".

Iterative design is emphasized, as working on the "wicked" problems require students to not only come up with creative solutions, but also co-evolve the problem and solution spaces [131]. Students often make use of prototypes that help them elicit feedback from stakeholders [26].

Product refers to the creative outcome. Example criteria include originality, functionality, adherence to client's theme, aesthetics, and typography. In our corpus, 40 papers discuss product criteria [74, 96, 20, 79, 87, 85, 99, 25, 23, 95, 103, 94, 86, 21, 101, 28, 29, 98, 119, 123, 127, 82, 5, 124, 121, 10, 100, 102, 117, 120, 34, 111, 107, 104, 89, 90, 91, 110, 8, 118]. A creative process is central to design education [20]. However, as many design educators argue the ultimate goal of the process is to develop a creative product [74]. Further, the product is usually designed with an

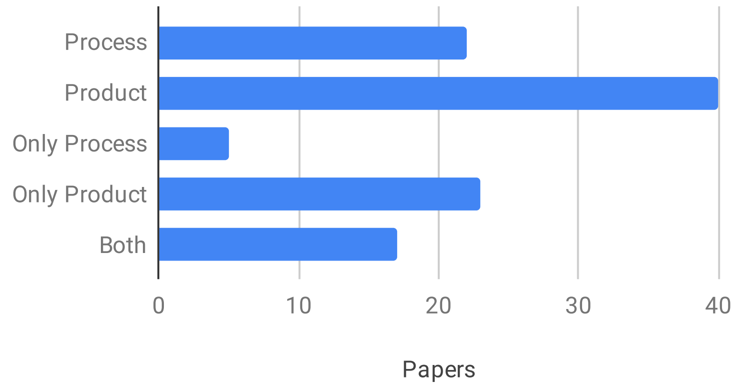

Figure 2.2: Distribution of papers based on their focus on process and product assessment. Process has received less attention.

end-user or client in mind, making the adherence to a client's theme an important criterion. We note Zimmerman's point of view here: "[Students] must convince the audience to want to possess their product or service" [96].

The what/how categorization (Tables 2.2 and 2.3; Figure 2.2) provides a different perspective that is more suited in certain cases, than the earlier organization by specific criteria. For example, a prior study found a need to focus on the process rather than the product when investigating self-regulated learning in design course contexts [132]. Such studies can benefit from identifying prior criteria and approaches using the what/how mapping. We note that the characteristics that fall under one category in 'Criteria' can fall under multiple categories in 'What/How'. For example, the conceptual criteria category includes both process and product aspects. Process aspects include analytical thinking, abstract thinking, analogical thinking, etc. Product aspects include variety or alternatives, originality, precedents, fluency or the number of ideas, etc.

2.3.3 Modality

Modality refers to the mode or form in which assessment and feedback are provided to students. Design critique takes place in a variety of settings, such as individual and group crits, interim and final review, and informal interaction [25]. Different modalities are suited in different types of interactions and for providing feedback on different types of design characteristics.

In the initial coding phase, we organized the modalities into three subcategories: verbal, digital, and physical. We did not form additional categories during the later analysis.

Verbal communication is the primary critique modality in design courses [25]. "Desk crit", which is the discussion between an instructor and a student at the student's desk, is noted to be primarily verbal in nature. Other settings such as interim review and final review also include significant verbal feedback. This is reflected in our corpus, with the highest number, i.e., 15 out of 60 papers discussing verbal modality as the form of feedback [20, 79, 87, 25, 28, 123, 82, 111, 90, 24, 109, 122, 92, 126, 30]. 10 of these 15 papers discuss instructor as the source, 6 jury, and 7 peer.

As Oh et al. point out, instructors often use multiple modalities together as that helps students in understanding their intentions [25]. Design studios are increasingly integrating *digital* support tools, including that for feedback processes [91]. 15 papers in our corpus use digital tools [87, 5, 124, 121, 10, 100, 118, 99, 115, 110, 126, 8, 89, 91, 25]. Jung et al. discuss a web-based 3D environment, which supports collaboration by allowing redlining and annotating of building designs [89]. Gray et al. discuss that instructors and peers engaging with students through document collaborations and backchannel chat increases the amount of feedback [126]. e-scape, a web-based authoring system, where students upload their work for instructor and peer review, is found to support collaboration and providing ongoing feedback [115]. Other examples include getting feedback from the crowd on recorded videos via MindSwarm [8], e-mail [110], and annotation using blogs & wikis [99].

Design studios are known for their purposely designed space that has tools and resources to help students perform design work, and facilitates students' collaboration with each other and instructors [22]. The physical space, for example, is designed such that students can pin up their drawings on the walls [95] or lay them out on desks [25] for review. Two papers in our corpus use physical modality [90, 25]. Instructors often draw or sketch over student work to help them develop ideas [90]. Gestures also facilitate communication and are considered 'invisible drawings'

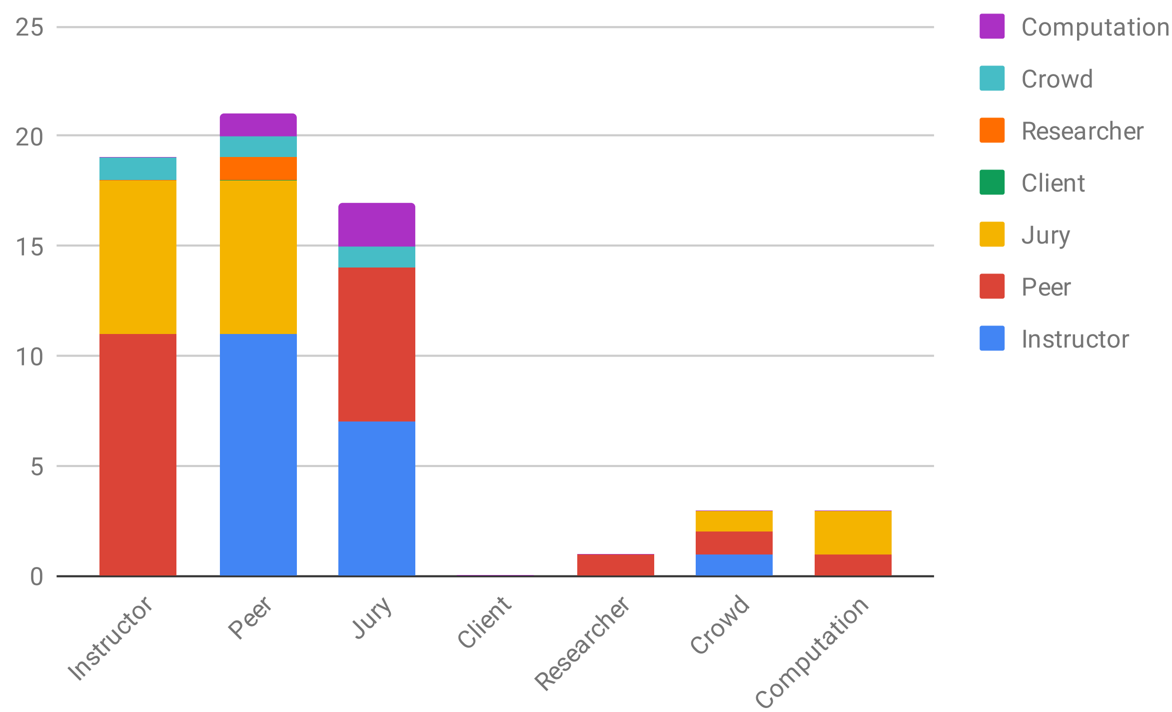

Figure 2.3: The chart shows the number of times papers in our corpus used two sources together for assessment and feedback. While traditional sources of assessment—e.g., instructor, peer, and jury—and their pairing are frequent, other sources have not yet been employed extensively. These include new sources, such as computation and crowd, which have the potential to scale assessment and feedback.

[25].

2.3.4 Source

As Krause et al. point out, design classrooms have changed, to employ various sources of feedback, as instructor-led feedback falls short in meeting the increasing scale of design education [5]. Further, multiple sources are valued in design education, as that allows students to obtain multiple perspectives on their design work. As Tinapple et al. [118] put, "Hearing feedback from multiple people allows the critique recipient to blend the feedback, discounting the outlying overly positive or negative comments, forming a better sense of the overall reception." Figure 2.3 shows the how frequent different sources are paired with each other in the papers in our corpus.

During the initial coding phase, we organized sources into three subcategories: instructor, peer, and crowd. During the later phases, as we discovered new sources through the analysis of remaining papers in the corpus, we revised the categories. As a result, we have six categories: instructor, peer, crowd, jury, client, and researcher.

Instructor assessing student work and providing feedback is a common approach in design education. In general, it can be considered as the de facto approach in education, as an instructor is well positioned to assess work performed in their course. This is reflected in our corpus as well, with the highest percentage, i.e., 23 out of 60 papers discussing instructors as the source [87, 115, 126, 8, 91, 25, 20, 79, 28, 82, 24, 122, 92, 30, 96, 85, 95, 101, 29, 112, 125, 113, 116]. The instructor feedback is noted to provide guidance, but at the same time, not stifle creativity. As [122] notes, instructors use a “suggest don’t tell” approach and help students develop conceptual connections. As [82] notes, instructors work as facilitators rather than lecturers.

Peer refers to students, who in most cases are at the same academic level, studying the same course, while sometimes, at different levels and/or in different courses. An example of different levels and courses is the interaction between two student cohorts: ten undergraduate students at Penn State and 12 postgraduate students at the University of Adelaide [114]. In our corpus, it is the second most common form of assessment, with 19 out of 60 papers discussing it as a source [87, 115, 126, 8, 25, 79, 122, 95, 101, 113, 116, 124, 118, 110, 90, 34, 114, 6, 99]. [118] notes that students sometimes develop a false sense of insecurity based on peer critique, and think they know less. But if students are able to see each other’s work, then that insecurity is alleviated.

Jury refers to a set of experts, who possess the knowledge, skills, and experience required for effective assessment of work in a specific domain. 16 papers discuss jury or experts as a source [87, 25, 122, 101, 124, 110, 20, 24, 92, 123, 103, 119, 104, 106, 121, 114]. As [25], discusses having a jury allows students to engage in inspiring conversations, learn from their expertise, and observe professional skills. At the same time, the critique can sometimes be harsh and interaction can become intimidating.

Client refers to the end-user who eventually uses or is intended to use the product or service being designed. Only 3 papers in the corpus discuss the client or user as the source [89, 94, 108]. [94] presents a study from an industrial design context where the client assessed student work, with their comments largely on the quality of product designed, students' interpersonal skills, management of resources, and level of maturity.

Researcher here refers to people who study and aim to develop or advance understandings of design education theory and practice, through experimenting with different tools, methods, and environments. We formulated this category because, in a number of studies [98, 120, 102, 111, 34], the researchers evaluate a system or environment they have designed by assessing the student work in controlled settings. However, these studies do not provide the assessment to the students themselves.

Crowd refers to large online communities of people, who make themselves available for undertaking paid or unpaid tasks published by requesters. As Dow et al. put, the modern web makes it possible to "leverage the scale, diversity, and immediacy of online crowds". There are different platforms that facilitate crowd work. In our analysis, we found 5 out of 60 papers discussing the crowd as a source [8, 5, 9, 121, 10]. 5 papers used Amazon Mechanical Turk [8, 9, 5, 10, 121], 2 MindSwarms [8, 9], 1 Reddit [9], 1 Facebook [8], 1 Upwork [10], and 1 CrowdCrit [10]. The purposes of use include needfinding [8, 9], ideating solutions [8, 9], testing prototypes [8, 9], and pitching [9]. Providing structure, such as guidelines or examples to the crowd has shown to improve the quality of feedback [5, 10].

Computation refers to the use of mathematical calculations to assess and provide feedback on characteristics of design work. As Jeffries point out, the use of computation has the potential to address the increasing scale of design education [100]. In our corpus, 4 out of 60 papers discuss using computational approaches [100, 119, 124, 127]. A couple of these papers use assessments from other sources—e.g., peer and jury—and then perform correlation or regression analysis [119, 124]. The other two papers assess specific characteristics, such as aesthetics [127] and consideration of alternatives [100].

2.4 Discussion and Implications

We connect the analysis with prior work, as well as use our observations during the bottom-up faceted analysis toward answering initial research questions. We discuss the potential role our analysis can play in supporting design instructors, by offering them a comparative understanding of assessment and feedback across facets. We follow with the potential value of our analysis for design researchers. We discuss and derive implications for: (1) studying design course contexts for developing understandings of instructors' assessment methods; (2) extending the use of alternative sources to meet growing demands, and (3) undertaking possess-focused investigations.

2.4.1 A Resource for Structuring Assessment and Feedback in Design Courses

Through our faceted analysis, we contribute a resource through which design instructors can refer to various aspects—i.e., criteria, what/how, modality, and source—and inform the structure of their courses. Making the present analysis available to instructors has the potential to catalyze evolution of their assessment and feedback practices. For example, an instructor—building a comparative understanding through facets—could incorporate additional criteria or sources used by their peers. This includes incorporating ideas from disciplinary contexts similar to theirs, as well as from other contexts in which similar criteria are employed. As prior research has shown, there are commonalities in assessment and feedback across disciplines [79]. Hence, while instructors can directly utilize ideas from similar contexts, perspectives from other contexts would also be useful.

2.4.2 Study Design Course Contexts for Developing Understandings of Instructors' Assessment Methods

One of our motivations for undertaking this survey was to develop understanding of approaches used for assessing and providing feedback on various design characteristics across fields (RQ1). We expected to gain insights into approaches for a range of characteristics that are a part of design instructors' rubrics. While we were able to develop understandings across the facets of criteria, what/how, modality, and source, there are opportunities for developing additional insights by undertaking studies of design course contexts.

Except for a few, the papers in our corpus do not explain the assessment approaches in detail (Table 2.2 and Figure 2.1). For example, for novelty, Shah et al. explain a normative measure, counting how many times an idea occurs in the entire collection of ideas [34]. They provide measures for fluency and variety as well. Cai et al. [104] also provide a methodology for assessing variety—based on extended linkography and distance graph—where they determine the breadth and depth of solution space. They quantify differences among multiple ideas generated by the designer, based on whether one idea is an alternative, or slightly different, or a more detailed version of another idea. Another example is Yuan et al.'s work, where they list principle statements that the crowd used as a basis for assessing student work against a set of criteria, such as choice of typeface, use of color, and the purpose of the project [10]. However, for a large number of criteria (Table 2.3), approaches are unknown.

Investigating design courses to make the details of assessment visible is vital. Such understandings can be beneficial to develop implications for how computational means can support practices (RQ2). As Dourish says, making contextual properties transparent is vital to developing computational support, and as context acquires "meaning or relevance through their relationship to forms of practice", study of work practices becomes fundamentally important [13].

2.4.3 Extend the Use of Alternative Sources to Meet Growing Demands

In the introduction, we pointed to the need for alternative sources for assessing and providing feedback, so as to meet the growing demands of design education. In our corpus, 23 out of 60 papers discuss instructors as the source [87, 115, 126, 8, 91, 25, 20, 79, 28, 82, 24, 122, 92, 30, 96, 85, 95, 101, 29, 112, 125, 113, 116]. A majority of these, i.e., 14 papers discuss one or more additional source of feedback [87, 34, 115, 122, 95, 25, 101, 8, 126, 113, 116, 20, 24, 92]. The observation corroborates the need and value of multiple sources in design education. Also, instructor and peer sources co-occur in 11 papers, the most when considering any two sources together (Table 2.2 and Figure 2.3). This may allude to the frequent critiques that are held in the classroom, typically in a studio setting [133], in which both the instructor and peer provide their feedback. Similarly, jury pairs with instructor or peer 10 times, pointing to the traditional settings where instructors invite people with valuable experience to critique student work. Hence, there is an opportunity for new, non-traditional sources to play a role in supporting instructor assessment and feedback efforts.

Computational: A distinct advantage of computational approaches is that they can process large amounts of data at speed. As we can see from Table 2.2, the investigations for computational approaches are fairly recent, as compared to that of traditional forms, which started two to three decades back. One paper develops a computational approach that can be used for assessing aesthetics through the use of genetic algorithms and artificial neural networks [127]. However, others are focused on correlation and regression analysis of human-assigned scores. For example, Arslan et al. develop a regression model to predict jury's scores for design presentation based on peer ratings of specific design characteristics of form, function, flexibility, aesthetics, and authenticity [124]. The dearth of work presents an opportunity for researchers to develop novel computational approaches. For example, a study could be performed by substituting peer ratings in Arslan et al.'s model with computational assessment of specific characteristics. As we pointed above, understanding practices is vital to developing computational support. Hence, it would be beneficial to begin with criteria for which prior work has made the assessment approaches visible. Simultaneously, computational approaches that have proven successful for design assessment, but have not yet been investigated in design education contexts—e.g., flexibility/diversity [43] and use of color [42]—could be considered.

Crowd: A distinct advantage of crowd techniques is that they have been found to come close to expert assessment and feedback on a variety of design characteristics [10]. In our corpus, we observe that crowd techniques have been used for criteria of complementary visuals and text, purpose, presentation, typeface, and color, form, and composition. Future investigations of crowd assessment have the potential to inform the community regarding its suitability and effectiveness for different types of design education criteria.

2.4.4 Undertake Process-Focused Investigations

We note the distribution of papers across process and/or product category. As [74] discusses, many design educators hold the view that product is the most important outcome in education. It quotes Lindström [134], "it is ultimately the products of the creative process that count in society." Lately, many educators have started emphasizing the process as an equally valued outcome. Watkins argues that while students may not work on the same product ever again, they can apply the learning from the process to different contexts [85]. Out of 40 papers categorized under Product, we categorized 17 as both Process and Product. While there are 23 papers that focus only on Product, there are only 5 papers that focus only on Process (Tables 2.2 and 2.3; Figure 2.2), thus indicating a potential opportunity for conducting process-focused investigations.

2.5 Conclusion

To advance design education assessment and feedback practices, we conducted a survey, forming and analyzing a corpus of 60 papers. We used a constant comparative method to develop facets: Criteria, What/How, Modality, and Source. The faceted analysis contributes a synthesized resource for design educators and has the potential to inform the assessment and feedback structure in design courses.

Our faceted analysis allowed a systematic identification of patterns and gaps in practices. To complement traditional sources—e.g., instructor, peer, and jury—there is an opportunity to develop computational means, which has the potential to scale design assessment and feedback. However, as we found, prior work does not make visible how assessment and feedback is performed in situated design course contexts. As Dourish discusses, a comprehensive understanding of socially situated work practices is fundamental to creating technologies that support users [13]. Developing a detailed understanding, across contexts, would be valuable for designing computational systems that support design assessment and feedback. In addition, our analysis reveals opportunities for establishing the suitability and usefulness of the crowd for a range of criteria, as well as undertaking process-focused investigations.

3. HOW COULD AI SUPPORT DESIGN EDUCATION? A STUDY ACROSS FIELDS FUELS *SITUATING ANALYTICS*

In this chapter, we use co-design methods to develop understandings of design instructors' situated practices, across fields, and derive implications for AI support for assessing and providing feedback to students studying design. As students learn to develop creative solutions for design problems, they need frequent assessment and feedback to make progress [25]. Design problems are known as *wicked*, that is, addressing the people and needs at hand; replete with confusing information and conflicting stakeholders' values; open to multiple explanations; and formulated and reformulated not absolutely, but in terms of a designer's conception [135, 136, 128, 137, 138]. Design students are appropriately challenged by this wickedness. They sometimes struggle to the extent that they drop out [30].

To help students learn in the face of wicked problems, design instructors structure courses to provide frequent, helpful feedback [109]. However, as design education demands continue to grow, instructors face challenges in providing timely assessment and feedback [5]. Further, instructors may not be available at critical times, e.g., late the night before an assignment is due. These conditions translate to needs for alternative channels of assessment and feedback. Human assistance channels include peer [7, 6] and crowdsourced [10, 8, 9] design feedback. The present research alternatively focuses on the potential of new forms of computation, which complement the human assistance by providing the ability to process big data at speed [11].

To supplement instructor efforts, under time pressure, our objective is to derive implications for AI-based design analytics and their presentation via dashboards. Like recent AI-based analytics research [14, 15], we adopt a broad view of analytics, which includes not only facts but also inferences. While learning analytics have been found useful in various courses, prior work lacks in investigating their efficacy in design education contexts. This does not mean that computational modeling of creative design, which can constitute a basis for AI and analytics, is impossible. Gero and Maher developed a computational model for supporting creative design based on the application of analogy and mutation processes to design representations [36]. Reinecke et al. demonstrated the potential of computation to assess aesthetic quality of website design [41]. Oulasvirta et al. provide a web service that assesses a graphical user interface design against a variety of metrics, ranging from symmetry to colorfulness to visual clutter [42]. This dissertation addresses the research gap of how to build AI support for design courses.

As we argued in Chapter 1, AI support for design education must provide transparent, on demand assessment and feedback. How can this work? Suchman seminally articulated that designing a useful AI system requires understanding users' *situated practice*, i.e., how their actions develop purpose and intelligibility within particular circumstances. Suchman emphasizes a system's transparency—to convey AI's intended purpose to the users and establish its accountability—as requisite for effectively supporting situated practice [12]. Dourish breaks this principle down to focus on how translating ideas between intellectually different domains of situated practice (social) and technology (computational) "can be both exceptionally valuable and unexpectedly difficult" [13]. While focused on user experience and HCI, we find that recent studies by Yang et al. [139] and Dove et al. [140] corroborate the difficulty of translating ideas between the domains of *design* practice and *AI* technology. Dourish argues that making contextual properties transparent is vital in addressing the difficulty, and as context acquires "meaning or relevance through their relationship to forms of practice", study of work practices becomes fundamentally important [13].

The present research develops a case study for human-centered AI, using qualitative, co-design methods to undertake the 'difficult' work of understanding practices in design education, in order to derive a basis of 'contextual properties' for AI support. While prior work has studied assessment contexts in design education, and identified a range of criteria [74], it has not detailed *how* instructors apply these criteria in practice. Hence, to derive implications for computational support, we investigated how design instructors, across fields—e.g., architecture, interactive art & design, mechanical engineering, and computer science—perform assessment. In our study contexts, these courses are taught by each field's faculty, in contrast with dedicated design programs—e.g., TU Eindhoven, TU Delft, and Stanford d.school—which teach students from multiple disciplines [4].

The present investigation began with the goal of addressing this research question:

RQ1: How, if at all, could analytics and AI support instructors in assessing and providing feedback on design?

Early in our investigation, we realized that co-design methods are imperative, to elicit and build on tacit and contextualized knowledge and practices of stakeholders [75, 58]. Our focus evolved, by performing the research, through co-design engagements. Working closely with design instructors, we discovered another, emergent and underlying research question:

RQ2: What challenges characterize how instructors teach, assess, and provide feedback on design (across fields)?

We present prior work relevant to various aspects of our investigation. Then, we describe our study methodology, including aspects of co-design for involving instructors, and a grounded theory approach for analyzing qualitative data acquired through co-design engagements. We follow this by presenting findings and discussing challenges, which characterize assessment and feedback across the diverse design courses of our study. The particulars of these situated understandings and needs for using AI to support design education lead us to next contribute a generalized methodology for situating AI-based analytics. We invoke this methodology to derive potential new forms of AI support for design education. We reflect on the value of and role for co-design methods for situating AI support. We conclude by reiterating theoretical contributions for human-centered AI, in general, and design education, as a particular situated domain.

3.1 Prior Work

We present prior work investigating design education assessment and feedback practices and identify gaps in current understandings of instructors' approaches. Further, we rely on prior work in creative cognition, computational design, and learning analytics and dashboards, which we introduced as sensitizing concepts in Chapter 1.

3.1.1 Design Education: Assessment and Feedback Practices

In Section 1.1.1, we introduced how design education, across fields, is based on common pedagogical beliefs and practices. Here, we focus specifically on assessment and feedback practices.

Design instructors assess student work against a range of *process*—e.g., problem solving, analytical thinking, abstract thinking, synthesis, creative thinking, and decision making—and *product*—e.g., concept, layout, color, shape, texture, and rhythm—characteristics [74]. Students need to develop both group—e.g., teamwork and communication—and individual—e.g., abilities to act independently and meta-cognitive skills to assess their own actions—competences [125]. While this prior work enumerates a range of characteristics, it does not surface instructors' approaches to assessing these characteristics. Without this contextual knowledge, as per Suchman and Dourish (above), it is not possible to derive implications for analytics and AI support. Hence, the present research focuses on addressing this gap.

We next draw on prior work addressing how instructors provide structure to elicit and facilitate students' learning and expertise through feedback. Feedback based on assessment of various design characteristics plays an important role, for advancing design students' development of expertise [25]. As students begin, they may be unfamiliar with the open-endedness of the design work [141]. Many students have trouble working [108] on wicked design problems [128] that require ongoing formulation, while dealing with confusing information and stakeholders' conflicting values. Instructors scaffold students' mastery of creative processes and development of key competencies [142]. They use a "suggest, don't tell" approach in their feedback, thus simultaneously directing students' attention to key aspects and empowering them to act independently [122].

To facilitate feedback, design courses schedule critique sessions, where instructors and peers help students reflect on their work [25, 109, 82]. Critiques usually take place in studio settings: "active sites where students are engaged intellectually and socially, shifting between analytic, synthetic, and evaluative modes of thinking in different sets of activities" [133]. Critiques range from frequent, informal "desk crits" and "pin ups" to more formal "jury" reviews at the end of a project [25, 109]. They provide useful feedback to students and help them develop communication competencies—e.g., demonstration of design evolution, interaction with audience, and credible staging of presentation performance—through understandings of disciplinary norms, expectations, and behaviors [24]. In addition to public critique, instructors and peers engage through document collaborations and backchannel chat [126].

As instructor and peer availability are limited, additional feedback channels have been investigated. For example, prior work focused on the potential of crowdsourcing [10, 8, 143, 9]. We alternatively focus on computational approaches, which have received little attention in this regard.

3.2 Methodology

We engaged aspects of co-design [75, 58] to develop understandings of assessment and feedback practices and needs in a range of design course contexts, across fields. "We" here refers to the resulting research team, comprised of the "initial research team" (IRT) members plus 2 design instructors who later chose to get more involved in the research, through the co-design process. The IRT is a group of HCI researchers, consisting of the PI, and graduate and undergraduate students. The PI additionally participated as a design instructor. The remainder of the section differentiates between this larger we, and the IRT, in describing processes and methods.

The IRT employed a co-design approach, with the goal of building on tacit, shared knowledge and communication involving users, as participants, so that outcomes support situated use. To initiate the co-design process, the IRT recruited other design instructors via email, starting with whose work they were aware of from past interdisciplinary collaborations. Based on instructors' recommendations, the IRT performed snowball sampling [78]. They motivated new participants by identifying explicit, immediate research goals of value to them, such as opportunities to reflect on their own processes and share perspectives with peers. IRT engaged instructors in dialogue about course learning objectives, assignment specifications, and assessment methods (See appendix A). Furthermore, they discussed whether and how computational means could support them, in particular, in design assessment.

The IRT initiated a *grounded theory* approach, based on the work of Charmaz [18], to data collection and analysis. In this approach, grounded theory is a rigorous qualitative method, wherein

ID	Gender	Field	Title	UG Course Level
D1	F	Architecture	Associate Professor	Freshman, Sophomore, Senior
D2	F	Interactive Art & Design	Associate Professor	Junior, Senior
D3	M	Architecture	Professor	Sophomore, Senior
D4	F	Interactive Art & Design	Assistant Professor	Junior, Senior
D5	M	Landscape Architecture & Urban Planning	Assistant Professor	Sophomore, Junior
D6	M	Landscape Architecture & Urban Planning	Associate Professor	Sophomore, Junior
D7	M	Architecture	Dean	Freshman
D8	M	Mechanical Engineering	Professor	Senior
D9	F	Mechanical Engineering	Lecturer	Sophomore, Senior
D10	F	Mechanical Engineering	Assistant Professor	Senior
D11	M	Computer Science and Engineering	Professor	Junior

Table 3.1: Our study participants consist of design instructors from diverse fields. We discussed course learning objectives, assignment specification, and assessment and feedback methods. With the exception of D7, D9, and D10, all teach graduate level courses, in addition to the undergraduate levels included above.

researchers perform constant comparisons among pieces of data, to develop analytical codes and categories. The method begins with sensitizing concepts—i.e., background ideas that provide researchers framing and guidance—and (cross) disciplinary perspectives, which inform the formulation of opening research questions. These, in turn, inform data collection and, the initial qualitative coding (labeling) of data elements. Initial coding is followed by focused coding, in which codes become merged and rearticulated. Conceptual categories get refined. Directions emerge through this continuous, comparative qualitative analysis. The sensitizing concepts and research questions continue to feed back, as researchers inductively interpret the codes and categories to derive theory that explains investigated phenomena [18].

Our discussions resulted in ongoing discourse, spread through multiple sessions, over a period of a year, as our understanding of teaching and assessment practices evolved by talking to different instructors. Emergent discussion topics include students' use of sketching, instructors' creativity assessment, and students' considering instructor feedback beyond a specific deliverable. We refer to these sessions as discussions, because the IRT shared perspectives gained from prior work and by talking with other instructors. In total, 11 instructors from 2 universities—from 5 fields—participated in these serial discussions (Table 3.1). During these discussions, the IRT asked instructors for example design assignments, including feedback provided to students, developing context.

Before proceeding further, we note that, in our research, like others [144, 145, 146], co-design methods and grounded theory are complementary approaches. Co-design supports designers and stakeholders not necessarily trained in design to work together in defining research products and designing solutions. This involvement generates data. However, co-design does not prescribe a particular data analysis approach. For data analysis, we use grounded theory methods, which enable discoveries, both expected and unexpected, to emerge through data.

The IRT organized a Design Assessment Workshop, bringing together 5 instructors, for an interdisciplinary discussion of situated practice. These included 2 instructors from landscape architecture, 1 from interactive art & design, 1 from mechanical engineering, and 1 from computer science. The number of participants was limited by scheduling constraints.

For the workshop, the IRT asked each participating instructor to present one of their design ideation deliverables, along with their methods of assessment. Participants engaged in extensive discussion, sharing perspectives based on practices in their field and classroom experiences. Together, participants identified common pain points and potential solutions. Participants responded with periodic sighing and murmurs of enthusiasm. Topics included the sizes of courses, teaching teams, and student teams; the level of details in assignment and rubric specifications; challenges in tracking students' progress and incorporation of feedback; and, difficulty assessing contributions in team assignments. The workshop thus fostered developing comparative understanding and co-creation of shared meanings, goals, and outcomes. Finding value in the research, two instructors chose to participate further by writing grant proposals and papers together. We further discuss how interests, understandings, stakes, and collaboration evolved in the section *Co-Design: Build a Design Education Stakeholders Community* (Section 3.5).

Two members of the IRT transcribed the individual and workshop discussion recordings and then performed qualitative codings of the transcribed data. Using Charmaz' approach [18], first, they performed initial coding of 3 individual discussions. Then, the whole IRT met to make the codes consistent and bring them into alignment. Next, we performed focused coding of remaining data. Based on relationships among codes, we organized them into categories. In a subsequent step, the larger we, including the two participating instructor co-authors, revised the codes, as needed, to suitably represent salient themes.

We next present findings, integrated with discussion, and implications for new forms of computation to address needs and provide support. We derived 3 categories through our grounded theory analysis of qualitative data, guided by our research questions and sensitizing concepts: 1) Design Process, 2) Assessment and Feedback Challenges, and 3) Computational Support. As the present research addresses how AI could support design education, which hinges on assessment and feedback across fields, here we only present themes from the second category and those pertaining to assessment and feedback in the third. We detail only those findings that complement prior work. We feature participant quotes illustrating phenomena.

3.3 Assessment and Feedback Challenges: Findings + Discussion

We present and interpret findings involving how instructors use rubrics of criteria in the assessment of student design work. Instructors specify grading rubrics, in accordance with learning objectives, through which students demonstrate performance [147]. Instructors' rubrics for design assessment assign weights to manifestations of student work processes, as well as resulting products. We contribute applying creative cognition's family resemblance principle to understand how rubrics work in practices, their functions and limitations. We problematize issues that arise in assessing contributions to team projects.

Feedback based on the assessment of various design characteristics plays a vital role in helping students make progress. We discuss various forms of feedback, e.g., verbal and redlining. We discuss challenges that arise both for instructors, who need to frequently provide feedback, and for students, who need to incorporate it.

3.3.1 Rubrics of Criteria and their Limits

Assessment criteria, in the form of grading rubrics, operationalize a variety of design characteristics that instructors find important for students to learn and demonstrate in an assignment. As D8 puts, criteria are *"kinds of rules that you can turn into a rubric."* In comparison with fact-oriented [148] assessment, such as through examinations that involve memorizing and reproducing, design students are required to demonstrate ability to reflect, apply, and understand [149].

D2: *I look at not just how well-functioning something is...conceptually, it meets what you planned...and also technically, it's been implemented correctly and functioning right. Also, aesthetics.*

When assessing, instructors in our study are interested not only in design product outcomes, but further, in the processes students perform to achieve them. Like de la Harpe et al. [74], we found that the weights, which product and process are assigned, vary across contexts. In line with Greene et al.'s work [132], we observe the need for focus on the process.

D8: *We tend to grade them on the process and if the process is coherent, but we also try to look at the outcome...In a professional context, you can look at products...and say this is a good design or not. But in a pedagogical context, maybe we shouldn't do this.*

We found that instructors in different fields, in this study, focus on different design characteristics, e.g., landscape architecture emphasizes visual aspects, mechanical engineering is more functionally oriented, and art & design emphasizes interaction aspects. D11's computer science and engineering assignments simultaneously focus on visual, interaction, and functional design characteristics. These differences in emphasis, on diverse design characteristics, is consistent both with prior work [79] and with instructors statements: design lacks universal criteria. Assessing it can be subjective, because in the end, some level of personal preference gets involved.

Consistently assessing design projects is thus, in itself a wicked problem. There are many variables to consider. The projects differ in objectives, personalities within teams, tastes of instructors, and situated aspects of sponsor capability and communication. Even when teams work on the same problem, their approaches may significantly differ. In D8's words, *"There's no right an-*

swer." Instructors create rubrics to provide structure. At the same time, recognizing and assessing the uniqueness of projects is vital. Project deliverables, criteria, and rubrics become amenable to change [29].

We contribute a new understanding that creative cognition's *family resemblance principle* [33] models the way that rubrics function in design courses. Instructors specify rubrics, but don't generally believe that particular elements of their rubrics are exactly necessary or sufficient. Rather, good solutions to the design problems tend to exhibit characteristics that the rubrics specify.

D5: *This is one of those hard things, because...some students could just do something really simple and it's just like, "Uh yes, that's it". And somebody else could have something really complex and you're like, "Yes, that's it".*

We use this new understanding, of how rubrics work according to the family resemblance principle, in conjunction with ideas from instructors, to inform design of computational support, such as AI-based design analytics and dashboards (Section 3.4).

3.3.2 Assessing Contributions to Team Projects

We find that assessing contributions in team projects is challenging and problematic. Sometimes some students get 'floated' in a course: other team members' efforts carry them. It becomes difficult for instructors to assess contributions.

The topic became a major discussion point during the workshop. One of the participating instructors brought this idea up, while envisioning computational means for providing insights into students' design processes. Soon, every instructor present in the workshop started identifying with the pain point and shared similar experiences.

D2: *All the roles are defined by the team. I kind of go and check who's doing what, but still, if the team is all agreed on certain each individual's role, they're happy with it, I can't really punish the individual person.*

D9: *I had 36 teams this semester and saw different kinds of team dynamic issues...I would be interested in things that help us in maybe more objectively gauge student contributions to the project.*

On the one hand, these problems point to social loafing where members make less efforts as responsibility is diffused and social matching where members conform to ideas [150, 151]. Creative cognition studies have shown that this reduces diversity of ideas [151]. Moreover, such students advancing—without learning necessary concepts and skills—can become problematic in future settings. In this vein, developing means for assessing individuals' contributions seems beneficial. However, how to use them may vary situationally. D2's suggests that a measure of each team members' contributions would be useful, but instructors should not make it the single factor in assigning grades to members.

On the other hand, the design classroom is a community of practice, i.e., "groups of people who share a concern or a passion for [design] and learn how to do it better as they interact regularly" [27]. In a community of practice, peripheral participation from some members is expected and legitimate. As these members gain understandings and skills, they take more central roles and contribute more. Learning science research has shown that in team settings, students who are less knowledgeable learn from those who know more [152].

The present research contributes AI-based design analytics (Section 3.4) and computing them team member-wise has the potential to address the challenge of assessing contributions to team projects.

3.3.3 Frequent Feedback

Students need specific and actionable feedback to understand where a design is lacking and make progress [153]. Instructors recognize that timely feedback is critical to student performance. At the same time, like prior investigations [122], we found that instructors walk a fine line between interfering, which could hinder students' ideation, and providing them with sufficiently frequent feedback to keep them on track and stimulate development. Like Oh et al. [25], we found that instructors provide feedback using a variety of modalities. We also found that a common pain point among instructors is that they invest time into giving feedback, then often find that students fail to keep track of it and address it as they revise their designs. More broadly, our study reveals students' lack of utilization of resources and incorporation of feedback provided to them.

3.3.3.1 Feedback Modalities

One common technique that design instructors use to provide feedback is to markup project documents. A salient form of markup is *redlining* (Figure 3.1) [89, 91, 154]. Practitioners either use a colored pen, typically red, to annotate printed materials, or do it electronically, with the help of computer tools. In conjunction with other forms of markup, such as sketches and annotations on tracing paper, the visual feedback helps designers iteratively formulate design problems and their solutions [154].

Figure 3.1: D6's feedback on a student's design through redlining. D6 marks up the problem areas and draws arrows connecting feedback written in respective boxes. The student receives feedback of using too much white space on the left and right sides, title taking more than needed space, text not styled per the instructor's guidelines, dominating background and small building size in the lower section, and an expected human figure.

D9: *We markup the documents ...and then give them a grade based on the rubrics that we have,* *along with the feedback. Feedback includes 'something is missing', 'talk about something different here'...it may even be feedback that corrects some of their grammar.*

Given overall workloads, instructors sometimes face challenges in providing quantitative assessment. In order for students to understand their performance, instructors may provide verbal feedback.

D4: *I give them verbal feedback when they're doing their presentation, but I don't necessarily give them the complete grade scores right after their project, which they really want to know, I know that.*

The present research builds on these understandings—in conjunction with the understandings regarding students' lack of incorporation of feedback (presented next)—to derive new computational support that can assist instructors in their feedback processes.

3.3.3.2 Students' Lack of Utilization of Resources and Incorporation of Feedback

In design courses, instructors often give students a variety of feedback opportunities. For example, D2 repeatedly says to her students that they can ask her to be a part of their ideation process and join their group meetings. Alas, she reports that students rarely contact her. D7 described that his course has 4 TAs, who are available to students 9-5 during the work week. This gives students the opportunity to consult with the TAs. He reports that they rarely do. Consequently, instructors are left with a sense that students could do better; if only.

Instructors sometimes ask TAs to regularly check student progress. This helps them become aware of issues, such as students who did not make enough progress, problems within a team, and lack of understanding of aspects of assignments. Balancing between interfering and guiding, instructors intervene when necessary. However, as D2 noted, *"It creates another layer. It's not just giving me all the information right away, so it takes some time."* A problem for students and instructors is to maintain awareness of distributed information.

Another problem in student responses during design education is that the feedback does not get incorporated, even though iterative design is explicitly taught. When discussing feedback processes, this problem was brought up by one of the participating instructors during the workshop.

Similar to the assessment of team contributions, this problem struck a chord with all participating instructors.

D11: *I say [to students that] you have to fix something...In the programming studio [course], we'll say, you have to use less saturation. You cannot have all these things be saturated. And sometimes I'll end up saying it over and over before they fix it.*

D6: *We have to kind of remind them, and then [our feedback] doesn't always all get incorporated. But that is how you get such polished products...that you continuously address some of [the feedback].*

Relatedly, we note that students often do not understand instructor feedback. Bridging this gap in understandings itself becomes an iterative process. Instructors point to the principles they taught, which students should demonstrate in their work as an evidence of meeting learning objectives.

D11: *If there are some students that are struggling, then we can give them extra feedback...[Some] violate the principles that I teach them, and then I say hey, the principles, we talked about that. And they're like oh, you mean [that]...they get it.*

In the next section, we develop ideas for assisting instructors in providing feedback plus addressing needs for how to track feedback.

3.4 Situating Analytics for AI

How could analytics help solve these design education problems involving assessment and feedback? Supporting design education with AI is challenging due to fundamental differences between the social and computational domains [13]. This interface border zone [73], where the fields intersect, is fundamentally sociotechnical. Building on the fundamental work of Suchman [12] and Dourish [13], we find that traversing the social / AI border requires understanding situated practice. The situated practices that we focus on take place in this border zone, involving instructors' design assessment and feedback. We build on these situated practices as a basis for discovering new approaches for deriving and presenting AI-based design creativity analytics.

In the user experience, analytics and how users interact with them, in the context of the task at hand, become inseparable. Thus, we define *situating analytics* as a methodology for conveying the meaning of measures that align with design rubrics, by contextually integrating the presentation of measures with associated design work. Here, space of tasks at hand involves assessing and giving feedback on design in educational processes of project-based learning. Situating analytics is based on identifying and building with contextual properties, derived through understanding practices. Interfaces traverse the border zone between the analytics and humans—here, students and instructors—making the analytics material [155] in the human experience, i.e., ascribing meanings, mediating emotions, and supporting interpretation.

We use the situating analytics paradigm to develop ideas for integration of AI as *algorithm-in-the-loop* [156, 157] support for design assessment and feedback, toward validating outcomes of instructor and student interactions with AI as accurate, reliable, and fair. As part of this, we develop ideas for a comprehensible and controllable AI integration, which can "thereby [increase] the users' self-efficacy, leading to reliable... & trustworthy systems" [158].

We invoke creative cognition's family resemblance principle as a basis for understanding the roles and limitations of AI-based analytics. We then propose: (1) approaches for deriving AI-based design creativity analytics that align with and support instructors' criteria; and (2) situating instructor and student interaction with the analytics—to foster transparency as well as validate assessment and feedback—through presentation, via dashboards integrated with design environments.

3.4.1 Situating Analytics as a Paradigm for Conveying the Meaning of Measures

We offer our approach to situating analytics, for assessing creativity in project-based design learning, as a paradigm for making the meaning of measures transparent to users. To derive analytics, one AI approach is to extract features such as colors, number of words, and number of images, and then use a machine learning model to predict scores with high accuracy [41]. However, such an approach fails to provide actionable insight to users. Instructors in our study highlight the need for approaches that can help students understand shortcomings in their work and give them insights about how to improve. Such assessments need to be learning objectives-based, in order to bridge the gap between instructors' and students' understandings. Further, instructors expressed that they should be able to indicate whether AI is performing as expected. These needs directly correspond to Shneiderman's principles: AI integration should be comprehensible and controllable [158].

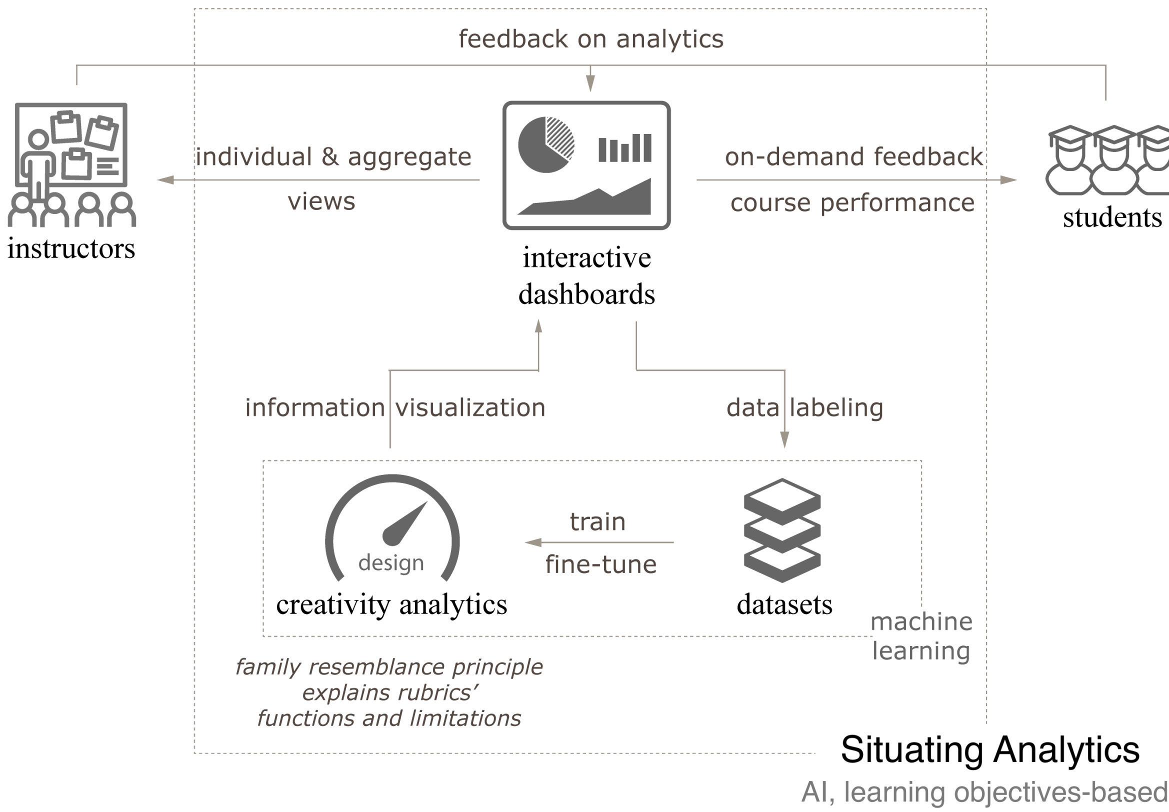

Figure 3.2: Diagram of a situating analytics approach to transparent design assessment and feedback. Transparent, learning objectives-based design creativity analytics—based on building with contextual properties derived through understanding practices—are presented via interactive dashboards to support teaching and learning, and simultaneously, to gather feedback for validating and refining analytics. (Use of icons under Creative Commons license. See attribution [1].)

D5: *There's a kind of disconnect between [students] turning in a [design] and they getting a number [back]...Why is it a 'B'?...[We] need to have a better tool communicating [the assessment] to the students.*

D6: *Yeah that would be cool I think if we could develop some metrics to build into a consistent rubric. It will spit out the rubric scores and then the professor can say, well that's right or wrong.*

We diagram a situating analytics approach to transparent design assessment and feedback (see

Figure 3.2). A co-design approach continuously involves instructor and student stakeholders in developing transparent, learning objectives-based analytics. Project-based learning analytics are derived using AI models that can constitute a basis for explaining the meaning of measures to the users. Analytics are presented via dashboards integrated with learning environments. Dashboard interaction can simultaneously support instructors and students in processes of teaching and learning, while at the same time gathering their feedback on the validity and utility of analytics. Dashboards need to present the approaches used to derive analytics, for intelligibility and accountability [159], in order to make how assessment works transparent. Through dashboard affordances, instructors and students will indicate whether and how an analytic and its derivation align with situated practices. They will articulate expected values and rationales when a computed analytic deviates. This data can be used to label analytics datasets that can be used to train supervised [40] and fine-tune unsupervised models [160], as well as to in/validate analytics.

In Section 3.4.2, situating analytics in practice, we prescribe AI approaches that can constitute a basis for conveying the meaning of design creativity analytics. In Section 3.4.3, we advocate dashboards integrated with design learning environments for situating instructor and student interaction with the analytics.

3.4.2 Design Creativity Analytics

Design courses involve project-based learning. Creativity is vital. As Blikstein puts it, in project-based environments, there is a need for analytics that can assess "much more complex [characteristics], such as creativity [and] the ability to find solutions to ill-structured problems" [56]. To measure creativity, its definition becomes essential. However, as discussed, established research avoids the epistemological trap of defining creativity in an absolute way (Section 1.1.2). Measuring creativity thus seems incompatible with most uses of learning analytics, which are to assess "specific and limited tasks" [56].

This is where the *family resemblance principle* [33] enables operationalizing design creativity analytics. Although creativity lacks an exact definition, according to family resemblance, specific characteristics of a design are likely to indicate creativity, even while no particular characteristic is necessary or sufficient (Section 1.1.2). In this way, we find that the family resemblance principle explains the rubrics that instructors specify for assignments. Solutions are expected to exhibit characteristics, but not compulsorily. Extending the scope of this principle, design analytics have the potential to play the same incomplete, yet useful, role as design rubrics.

Thus, we prescribe situating design creativity analytics that indicate the likelihood of quality design, without being absolutely definitive. We argue against any one computed analytic being seen as necessary or sufficient. Rather, we advocate *suites of design analytics*, each with the potential to measure characteristics that often occur in good design solutions. *Design analytics suites need not be complete, in order to add value.*

We note that situating analytics corresponding to instructors' every rubric element may not be possible. For example, it can be challenging to computationally assess *"a spatial sequence [in an architecture] that leads [a person] to an intense highlight moment (D1)"*. Our goal here is to augment, not replace instructors' roles.

The instructors in our study believe that some feedback about design, which they now provide "by hand", can be computed—and provided on demand—in order to help students to more quickly understand problems and correct them. We argue that AI techniques are suitable for various design characteristics, providing means to compute on demand, transparent, learning objectives-based analytics. Further, consistent with ongoing forays into process-oriented creative work [161, 162, 163, 164], we found that design process is important in how instructors assess student work. Thus, a derivation of analytics would beneficially incorporate process characteristics.

We develop a set of situated *design creativity analytics*, to computationally derive actionable measures for giving students feedback on project-based learning work. Each analytic is mapped, by family resemblance, to a criterion, which we discovered in design education rubrics, across situated contexts of our study (Table 3.1). Some design creativity analytics correspond to previous creative cognition ideation metrics [34, 35]: *Fluency*, or the number of ideas (D2, D8, D9); and *Flexibility*, or the diversity of ideas (D8, D9, D11). Other design creativity analytics are, as far as we are aware, new to creative cognition: *Visual Consistency* (D1, D6, D11), or the presentation of similar information using similar attributes; *Multiscale Organization* (D6, D11), or the presentation of ideas as a hierarchy; and *Legible Contrast* (D5, D6, D11), or the juxtaposition of hues for legible presentation. To further distinguish, a prior creative cognition ideation metric, Visual Presentation [35], assesses aspects of Visual Consistency, but it does not assess Multiscale Organization and Legible Contrast. Appendix B enumerates additional potential learning analytics, which we identified, and find relevant to design education.

D6:*...[the computer] could look and say you only got 2 drawings at the [larger] scale where you got 35 at the [smaller] scale, [hence] you need to do more analysis of the larger scale.*

D11: *...[the computer could tell] where high contrast is vs. where there is less contrast. [This matters] because contrast takes human visual attention.*

3.4.2.1 Fluency

Instructors in our study use the number of ideas—known as the Fluency ideation metric in prior creativity research [34, 35]—as a key analytic for assessing creative design. For example, D8 requires students to brainstorm and come up with as many ideas as possible in the beginning phases of design, stating that the activity's "goal is quantity". According to Darwinian theories of creativity, more the number of ideas, higher is the likelihood that one of them will survive and grow to be creative [165, 166]. In the brainstorming process, students develop ideas in textual as well as visual forms. Language and imagery represent complementary processes of human cognition [167] and their combination is known to aid the formation of mental models [168, 169].

Prior creativity research developed computational means of assessing Fluency, such as the number of text [35], image [35], and sketch [72] elements that comprise a free-form visual semantic composition of ideas. A recent investigation found that Fluency assessments provide instructors with first-order insights into student effort levels on design projects [57]. The paper reported that instructors desire an advanced analysis of design work, such as extraction of ideas contained within text and image elements.

To compute design creativity Fluency analytics, researchers can use state-of-the-art AI-based content recognition—e.g., natural language processing (NLP) [170] and computer vision models

[171]—to extract ideas contained within text, image, and sketch elements. An example is Mackeprang et al.'s investigation for supporting collaborative design, where they use NLP models to extract ideas contained within participants' text entries [43]. As computer vision models have been found to generate highly accurate image [171] and sketch [172] descriptions, researchers can utilize them to extend Mackeprang's approach to designs that compose diverse types of elements, such as text, image, sketch, video, and embedded docs and maps.

3.4.2.2 Flexibility

We invoke the ideation metric Flexibility/Variety from creative cognition [34, 35] to represent instructors' use of the number of categories of ideas as an analytic for assessing design work. For example, D9 requires students to "organize [generated] ideas into concepts", and then include "at least four 'different' concepts" and "a discussion of the advantages and disadvantages of each concept". In design creativity research, Flexibility/Variety represents the span of the explored solution space [33, 35]. Flexibility measures opportunities for remote associations, which are vital in developing creative solutions [173].

Prior creativity research developed computational means for assessing Flexibility, such as the number of different webpages, websites, and website types that the users collected ideas from [35]. Recent work on collaborative design ideation examined the use of external knowledge graphs such as Wikidata and DBpedia to compute semantic distances among ideas, toward assessing Flexibility/Variety [43]. Researchers have used computational methods to organize ideas into a tree structure. Examples include Linsey et al.'s use of WordNet [102], Fu et al.'s latent semantic analysis in conjunction with a posterior probability based method [174], and Vattam et al.'s use of functional hierarchies [175]. The tree structure supports analogical thinking and has been used to assess semantic distances among ideas. A promising approach is to investigate vector word representations to compute semantic distances [176].

To compute design creativity Flexibility analytics, first, contextualized vector representations of words [176] can be created using text sources, such as books, scholarly articles, and patents in a design field. As noted above, AI-based computational models can be used to extract ideas from text, image, and sketch elements of a design. Then, using the vector word representations, semantic distances among these ideas can be calculated [176]. Categorizing ideas, based on these semantic distances, has the potential to provide measures that correspond to instructors' understandings and assessments. Graph visualization of ideas, based on semantic distance, has the potential to support reflection. Walking across these graphs has been shown to support analogy formation [177, 102].

3.4.2.3 Visual Consistency

Visual Consistency refers to using graphical attributes—such as size, color, and font—in similar ways for presenting similar types of information. For example, all section headings could be in boldface, 14-pt and all subsections headings in italic, 12-pt Helvetica font. Instructors teach students to follow principles of visual design, through which even complex information can be presented with clarity [178, 179]. D11 provides guiding instructions to students for developing a *visual program*—which includes "a grid structure, consistent type sizes and styles, and a color plan"—based on Meggs' definition: "a system of parameters used consistently to unify a series or sequence of designs" [180].

Prior creativity research lacks computational methods for assessing consistency. Human raters assigned scores for the use of visual design principles such as whitespace, arrangement, and coherence, as part of assessing Visual Presentation creative cognition ideation metric [35].

To compute design creativity Visual Consistency analytics, in cases where a design schema stores information about the type of each element (e.g., heading, subheading, caption), the attribute values for elements of the same type can be compared. Inconsistencies can be highlighted. Alas, such schematized feedback is absent from typical design environments, e.g., Photoshop, Illustrator, and InDesign. In such cases, elements could first be clustered using attributes such as position and size, as well as ideas contained in them. Then, if two clusters have similar relative positioning of elements within them, then the attributes of the corresponding elements can be compared, to highlight inconsistencies. Further, the number of clusters can indicate the use of whitespace for organizing information.

Recently, clustering algorithms based on users' implicit actions of organizing content in a 2D design space have shown promising results [181]. Design environments can utilize such algorithms, when process data is saved. For example, they can incorporate user actions of selecting, moving, and manipulating elements together within the design space toward determining spatial proximity.

3.4.2.4 Multiscale Organization

Multiscale Organization refers to the visual and conceptual representation of ideas using hierarchy. It is a foundational element in design [69]. D6 requires students to visualize their landscape architecture project data "from national to regional to site scale". D11 engages students in multiscale organization, through a collaborative, zoomable design space and guides students to "use scale to nest sets of elements, where appropriate, to create readability, since [they] are sure to have more elements than can fit on screen, and in human working memory, at one time".

Multiscale organization supports designers in exploring, juxtaposing, and synthesizing ideas and their relationships across multiple scales [16]. It allows them to connect with and develop a design, by shifting their cognitive point of view to different scales or levels [69]. Design environments supporting multiscale organization—e.g., Photoshop, Illustrator, and IdeaMâché [70]—allow going beyond 2D and assembling content at multiple zoom levels or scales. Lupfer et al. found that, for students using IdeaMâché to perform free-form web curation, multiscale organization supported students' iterative and reflective ideation processes when working on design projects [70].

To compute Multiscale Organization analytics, as a starting point, the number of scales or zoom levels at which students have organized their ideas can be computed. Then, the assessment can focus on consistency aspects discussed in the last section, i.e., compare clusters scale-wise to find whether similar information is presented using similar attributes. For multiscale clustering of student design, researchers can investigate AMOEBA, which uses Delaunay triangulation to compare distances among different elements and recursively determine spatially nested groups [3]. They can likewise investigate identifying nested groups using Self-Organizing Map (SOM) techniques, which preserve topology and have proven effective for spatial clustering [182].

Further, techniques based on process data [181] can be extended to multiscale clustering algorithms. For example, user actions of zooming in or out, followed by the selection and manipulation of a set of elements, can be used as a factor in determining different scales of content organization.

3.4.2.5 Legible Contrast

Contrast refers to color properties that can be juxtaposed to produce a range of visual effects [183]. D11 specifies in assignment: "...appropriate use (not too much!) of contrast". This characteristic is based on visual design principles articulated by Tufte [178], which have been used by prior creativity research for assessing visual presentation [35]. In regard to contrast, Tufte invokes Imhof's first rule of color composition: large adjacent areas of pure, bright colors are loud and unbearable, but when used sparingly, can help achieve extraordinary effects [184, 178]. In Figure 3.1, such adjacent areas reduces the focus on the building in the foreground. As D6 redlined, "Tone down photoshopped sky background. It dominates".

To compute design creativity Legible Contrast analytics, image processing and computer vision algorithms that are capable of identifying regions of high contrast [185] can be utilized. Such algorithms operate on extremely small blocks, so one potential way is to find the percentage of such blocks in an overall design. Another possibility is to detect thick lines or boxes by finding long sequences of these high contrast blocks. For smaller occurrences, an interactive machine learning approach [160] would be beneficial for obtaining instructor feedback, to iteratively improve identification of whether the use of high contrast is excessive in the context of a design.

As a large collection of examples with such feedback becomes ready, deep learning and domain adaptation techniques [40] can be explored. As we discuss below, a learning analytics dashboard can play a new role in this collection of labeled examples.

3.4.3 Situating Interaction with Design Analytics

Enhancing sociocultural contexts with AI is challenging due to the multitude of cognitive, affective, and latent factors in play [186]. The statistical intelligence of AI/ML algorithms may provide an interpretation that drastically differs from common sense human interpretation of data

[140]. While algorithm-in-the-loop AI approaches have been shown to improve accuracy in multiple contexts, they are prone to bias and errors [156]. A successful human-AI system requires "careful design of the fine structure of interaction" that makes AI integration *controllable* and *comprehensible*, which thereby increases users' reliability and trust in the system [158].

We propose integrating analytics dashboards with design learning environments, as a means for jointly (1) situating human-AI interactions within the contexts of design courses; and (2) directly meeting instructors' and students' practices and needs. As part of situating instructor and student interactions with dashboards, we develop ideas for affordances through which users can in/validate analytics and provide feedback that can be used to improve AI algorithms. Through analytics dashboards integrated with design, instructor processes of providing feedback and students needs for receiving such feedback have the potential to be functionally rendered as isomorphic with AI algorithms' iterative needs for feedback on examples, in order to constitute labeled datasets for training recognizers. That is, we call for analytics dashboards integrated with design, in order to pair instructors giving feedback and students receiving it, with feedback on AI-computed design creativity analytics.

Turnbull explains indexicality as involving maps over space and time, which convey information that "can only be [completely] understood within the ... specifics of the circumstances and cannot be generalized beyond that context" [187]. We use this idea to motivate the need for connecting design analytics dashboards with learning environments. Design analytics dashboards are indexical, i.e., they derive and present analytics based on characteristics that exist and can best be understood within the context of actual design work. Indexically connecting AI-based design creativity analytics with design elements in learning environments can improve comprehensibility. Such indexical connections can support users in understanding the basis of analytics and in/validating them, and through this process, providing feedback that AI designers can utilize to improve recognition.

3.4.3.1 Design Analytics Dashboards

In our study, instructors brainstormed ways in which analytics [188] and visualizations [189] could help them gain insight into students' performance. They expressed needs for tracking design processes across multiple levels. For example, D2 would like to be able to compare points a student got last week, or on previous projects, to find out if the student has improved. D4 imagined visualizations that would reduce time and effort to make reports at the end of a semester, or at the end of the project, so that there isn't a need to *"write every single thing from scratch"*. D6 wants to see measures of how an individual student or team worked with an artifact (process scores / analytics), coupled with characteristics of that artifact product.

D6: *Is there a way...we can keep track of whether or not they got better presentation scores, based on how many times they practice?...We can see where they fixed [their design] before they actually come in and present.*

Dashboard, defined as "a visual display of the most important information needed to achieve one or more objectives" [190] can be useful in these regards. The tabular format—each row presenting an observation, with columns corresponding to attributes—supports "flexible querying and many perspectives for data exploration" [191]. Prior learning analytics dashboards [54] have facilitated quick understanding of the progress of student teams and individuals.

However, their efficacy has mostly been investigated in lecture-based learning contexts. Using dashboards for project-based learning requires personalization [192]. As design assessment lacks absolute criteria, design analytics dashboards specifically need to enable each instructor to select and combine characteristics, based on their pedagogic orientation and the project at hand.

In design education contexts, AI-based design creativity analytics will measure complex characteristics. As Bellamy et al. discuss, *bias* can result from systematic error in AI training data or models; bias can produce unfair outcomes for individuals or groups within a population [193]. Greene et al.'s algorithm-in-the-loop research advocates for users' rights to challenge assessments that affect them, as AI algorithms often fail to adapt to novel circumstances [156]. According to Woodruff, users' ability to challenge and change AI decisions—thus allowing contestability and recourse—are approaches for addressing algorithmic bias [194]. Further, to design AI technology that users can trust, Woodruff et al. discuss the need for strong participation of stakeholders and consideration of multiple perspectives [195].

Design analytics dashboards have the potential to situate instructor and student interaction in forms that mitigate complex problems posed by the use of AI. Dashboards can actively engage instructor and student stakeholders in continuous validation and refinement of AI models and derived design analytics. Dashboards can concurrently make analytics available in design learning contexts and gather data about how the analytics, as well as their presentation, affect design learning experiences. In situating stakeholders' interaction, comprehensibility and controllability will be vital.

Comprehensibility: In Section 3.3.3.2, we described how explaining the basis of assessment to students is important in helping them improve their work. As D11 expressed, *"[Some] violate the principles that I teach them, and then I say hey, the principles, we talked about that. And they're like oh, you mean [that]"*. Analogously, instructors expressed the need for AI-based analytics to be comprehensible. Recalling D5, (Section 3.4.1): *"There's a kind of disconnect between [students] turning in a [design] and they getting a number [back from AI]...why is it a 'B'?"*.

To aid comprehensibility, dashboards need to explain (make visible) the underlying basis for computing analytics. For *Fluency* analytics, dashboards can present ideas extracted from text, image, multimedia, and sketch elements. For *Flexibility* analytics, dashboards can relate extracted ideas by visualizing semantic distances among them. For *Visual Consistency* analytics, dashboards can identify elements that use dissimilar attributes despite being of the same type (Figure 3.3). For *Multiscale Organization* analytics, similarly, dashboards can identify inconsistent elements scale-wise. For *Legible Contrast* analytics, dashboards can present regions of high contrast. Connecting the presentation with learning environment can further aid in comprehensibility (Section 3.4.3.2).

We note that our approach of explaining analytics involves making the lower-level inferences that AI makes transparent. These include various conceptual and visual aspects present in the design (as detailed in Section 3.4.2). Examples of conceptual lower-level inferences include ideas extracted from text, image, and sketch elements. Examples of visual lower-level inferences include clusters and areas of high contrast present in the design. High-level inferences are the analytics—Fluency, Flexibility, Visual Consistency, Multiscale Organization, and Legible Contrast—derived using these lower-level inferences. To maintain focus, the present research only develops approaches for explaining how analytics are derived using lower-level inferences, not the lower-level inferences themselves, and so on. For example, for comprehensibility and controllability of Fluency analytics, we present to the user the ideas considered toward computing the Fluency, but not how AI determined whether a given image is of a cat or a dog.

Controllability: In case the AI assessment does not make sense, instructors want outcomes to be controllable. As D6 expressed (Section 3.4.1), *"the professor can say, well [the assessment is] right or wrong"*. The presentation of the underlying basis allows instructors and students to challenge not only an analytics score, but also the approach used to compute it. To aid controllability, dashboards can include affordances for instructors and students to express if and how a given design analytic makes sense.

Human-AI interaction through dashboards can also be structured as a form of labeling data, to enable instructors to simultaneously give feedback to students and developers, and about datasets. For example, if instructors disagree with any computed score and/or approach, they can be presented with affordances to enter an alternative score and the rationale behind it. With an increasing number of examples, machine learning algorithms—including deep learning with attention mechanisms [196]—can be investigated to more closely model instructors' rationale and assign scores for a new design in the given context. Further, using transfer learning [197], the data can potentially be utilized in other course contexts.

In the context of situating instructor and student interaction via dashboards, we note that the purpose is to make AI integration controllable and comprehensible, as an approach for addressing algorithmic efficacy and bias. We are not advocating poorly performing AI algorithms, which can increase the load on students and instructors, by requiring extensive input from them when algorithms produce incorrect results.

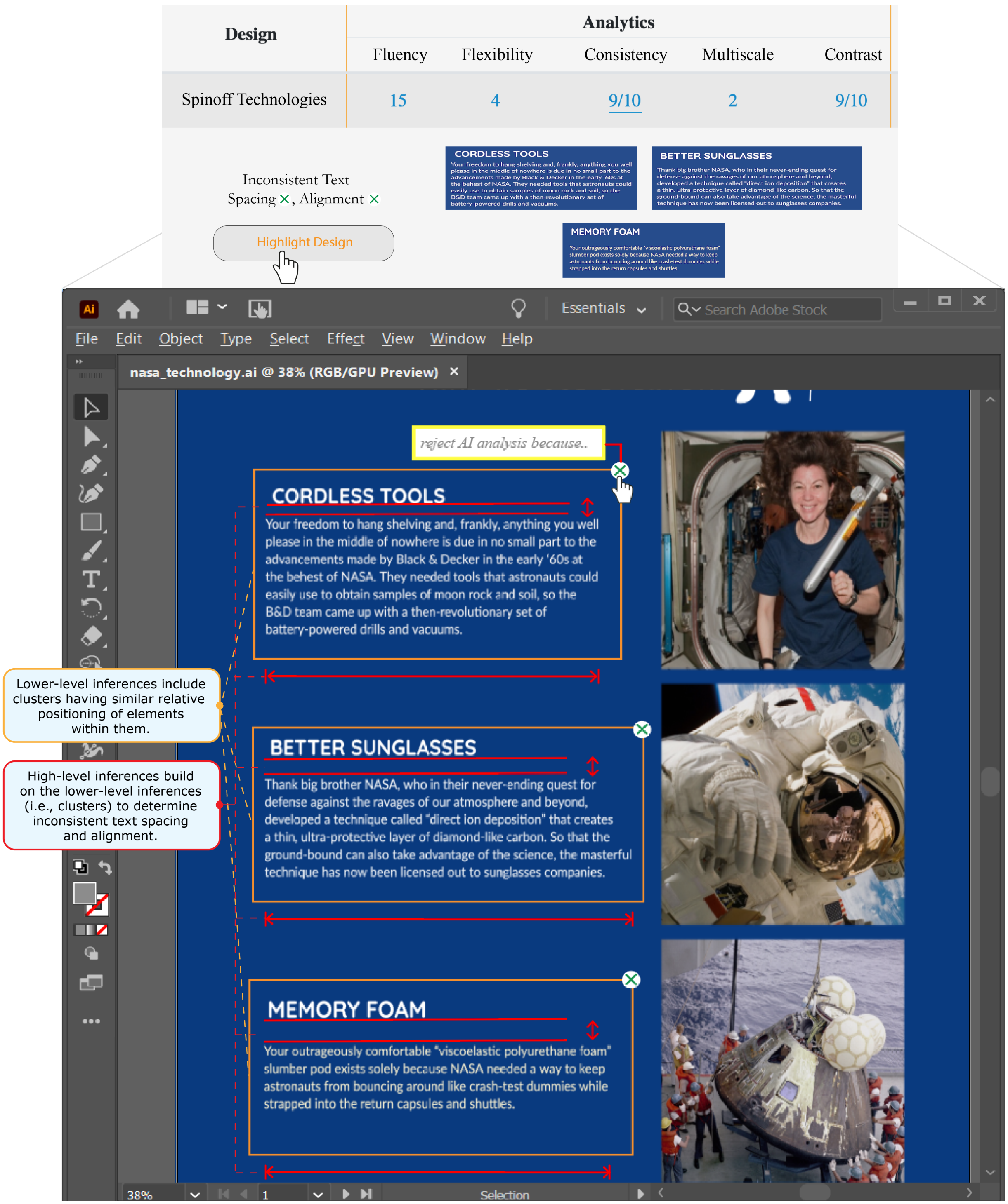

Figure 3.3: A mockup of dashboard integration with design learning environments to support comprehensibility and controllability. Design analytics dashboards are indexical: they present analytics that can be understood more effectively within the context of actual design work. This mock-up shows clusters (in orange) used as the basis for Visual Consistency analytics and inconsistencies identified (in red). Users can challenge assessment (activating cross symbols) and input rationale (in the presented text box). See attribution for 'NASA Technology' design example [2].

3.4.3.2 Integrating Dashboards with Design Environments

As described above, dashboards are indexical: they present design creativity analytics—based on characteristics valuable to the instructor and student users—which refer to and can be understood, more completely, within the context of the actual design work. The indexicality of the dashboard can function as a mediating mechanism between AI-based analytics and design work. This integration can improve the comprehensibility of analytics by connecting them to corresponding design elements in learning environments.

For *Fluency* and *Flexibility* design creativity analytics, if the user wants to understand the spatial and conceptual contexts in which an extracted idea occurs, a dashboard can support activating the idea and highlighting its occurrences within a design. The user could then indicate whether the AI-algorithm correctly extracted an idea from respective spatial and conceptual contexts, and input if a different idea would be more suitable. For *Visual Consistency* and *Multiscale Organization*, likewise, if the user wants to see the basis of comparison of attributes of two elements, a dashboard can support redlining inconsistencies and presenting—within the design—the scales and clusters based on which different sets of elements are being compared to each other (Figure 3.3). The user can then indicate whether the AI-algorithm correctly identified scales and clusters. For *Legible Contrast*, a dashboard can likewise support redlining the regions determined as high contrast and viewing them within the design, in the context of surrounding elements, to allow the user in validating whether the use of contrast is indeed excessive.

Tracking Feedback: Integrating dashboards with design can help students and instructors track feedback. Further above, we discussed students' lack of incorporation of feedback, despite reminders, is a common pain point for the instructors in our study. Instructors find it frustrating, as they put significant efforts into providing frequent feedback. Verbal critique and physical redlining of physical copies are often lost and feedback provided via learning management systems [198] also does not get incorporated. The problems that instructors point out continue to exhibit across iterations.

Firstly, to enable connecting verbal and physical feedback with design elements, in learning environments, computational support needs to facilitate digitization. Secondly, the needs for connecting and tracking immediately suggest a combination of node-link [199, 200], compound [201], and spatial hypertext [202] approaches. Using these approaches as the basis, dashboards can visualize where the feedback originated and how the design changed over iterations, including feedback incorporation, or failure to incorporate, to varying extents. Design versioning features such as visual “diffs” [203] will facilitate making design feedback and iteration visible.

Direct support for redlining [89, 204] students’ assignments, within digital environments of design tools—and creating connected dashboard representations—also needs to be investigated. Redlining can be added, as a special CSCW layer, in tools such as Photoshop and InDesign. In conjunction with community feedback on creative design [205], this design process integration has the potential to transform the processes of teaching and learning. As students iterate on design work, they can respond using this layer, which will trigger notifications to instructors that students have addressed particular feedback. In addition, the layer can support instructors’ needs to flag the presence of feedback and if it has been addressed.

D6: *...if students could just show their work and we could just draw straight on it. That would help us a lot. It’s instant feedback, you know.*

D2: *Yeah, and also sometimes if I can kind of put a flag, and give me a reminder if this person [incorporated the particular feedback].*

By identifying commonalities, involving feedback across deliverables, recurrent problems from this layer can be highlighted to instructors and students on dashboards. This information can stimulate instructors’ pedagogical intervention and students’ seeking helpful resources. The problems can be categorized at student, team, and course levels for effective understanding.

3.5 Co-Design: Build Stakeholder Community to Support with AI

We advocate using co-design methods for investigating AI support for human and social computing contexts. These methods actively involve users across stages of design processes [75, 58, 206]. The methods situate users as the ‘experts of their experience’ [75]. Their expertise involves forms of tacit and shared knowledge and communication [76]. Making the tacit *visible* plays a crucial role in designing support for participants' work practices [77]. Co-design provides designers with methods that support users in expressing and sharing their expertise [75].

We actively involved our "users" through co-design approaches. Their stakes in the process and levels of engagement increased. They valued the work more. They become more forthcoming and more available. Participating instructors shared ideas on addressing common pain points. Findings and implications are derived *with* them, not *for* them. Design with stakeholders, rather than for stakeholders, deepens involvement [207].

As described in the methodology (Section 3.2), finding value in the research, two instructors chose to participate further. Participating instructors expressed interest, and began adopting a design ideation environment [208] that the initial research team is developing. They have begun using assessment and feedback support capabilities designed in conjunction with ideas presented in the dissertation. The PI and one of the participating instructors, who became a co-author, together published a work-in-progress based on user experiences with an initial version of design creativity analytics and dashboards in course contexts. Further, the PI and two participating instructors who became co-authors have started writing grant proposals together, addressing design education challenges, identified across course contexts, through the co-design process.

Learning analytics and dashboard environments that represent them to users, function as boundary objects [209]: "physical or conceptual entities that each [stakeholder] interprets in its own way, but that provide common referents or points of articulation to ground conversations" [210]. These boundary objects, in conjunction with co-design methods, can facilitate addressing "the 'middle space' where learning and analytics meet" [210], and answering questions such as whether analytic computations provide actionable insights to stakeholders. In the course contexts of our study, design creativity analytics, presented via dashboards integrated with learning environments, have started providing common referents for instructors, students, and developers of AI-based computational support. In our initial observations, a triumvirate of situated artifacts—design rubrics, AI-based analytics, and dashboards that enable controlling and comprehending outcomes—support grounding conversations among stakeholders. Further research is needed.

Stakes and engagement evolve through co-design, to manifest growing interest, usefulness, and involvement. We envision the community growing further, while providing new opportunities for multidisciplinary design discourse and collaboration.

3.6 Conclusion

We developed a case study for human-centered AI, engaging aspects of co-design with instructors across diverse fields and developing understandings of situated practices, in order to investigate whether and, if so, how new forms of computation, e.g., AI, could support teaching and assessing design. Our findings are affirmative. Using a grounded theory approach to analyze data from workshops and discussions, as well as design artifacts from courses, we formulated categories and presented relevant themes, focusing on: 1) assessment and feedback challenges and 2) implications for AI-based analytics.

Our case study contributes new theory: (1) understanding of uses and limitations of rubrics through creative cognition's family resemblance principle; (2) *situating analytics*, as a paradigm for for conveying the meaning of measures that align with design rubrics, by contextually integrating the presentation of measures with associated design work; (3) operationalizing design creativity analytics, based on family resemblances in design rubrics, which has the potential to provide students with actionable, on demand feedback, and thus support them in learning to do design through iterating on their design work; and (4) dashboards, integrated with design environments, for situating instructor and student algorithm-in-the-loop interaction with AI-based analytics.

We invoke the family resemblance principle to contribute new understanding of how rubrics work, and likewise for how AI-based design creativity analytics, derived from process and product data, can productively be incorporated into education. Family resemblance tells us that no particular characteristic is essential, but together—in a rubric or analytics dashboard—they tend to indicate good design work. As D5 expressed, one student could do something really simple, while another something really complex; both may produce good designs. Hence, while rubrics and analytics provide vantage points, they do not provide a God's eye view [211]. Analytics' purpose is to augment, not replace, instructors' ongoing interpretation and engagement.

We lay out *situating analytics* as a paradigm for conveying the meaning of measures to the users. In the present research, this means for assessing design in educational processes of project-based learning. Goals are to bridge comprehensibility, controllability, and actionability gaps between users and AI. Gaps between users and AI have been recognized in the healthcare domain, where despite success in lab settings, AI-based clinical decision support tools sometimes fail in practice, due to, "a lack of consideration for clinicians' workflow" [212]. Through a situated, co-design approach, we discovered characteristics that instructors seek when assessing design. Next, based on these contextual properties, we identified analytics corresponding to salient conceptual and visual characteristics, such as Fluency, Flexibility, Visual Consistency, Legible Contrast, and Multiscale Organization of ideas. Then, for deriving these analytics, we focused on AI-based approaches that can constitute a basis for conveying the meaning of measures to the users.

We argue that explaining AI outcomes is vital to human-centered AI; it will make AI more valuable to people and society. A 'father' of modern AI tweeted, *"Suppose you have cancer and you have to choose between a black box AI surgeon that cannot explain how it works but has a 90% cure rate and a human surgeon with an 80% cure rate. Do you want the AI surgeon to be illegal?"* [213]. In response, there was a huge debate. Many opposed the black box AI, arguing that the 90% rate system might not have been trained on a representative sample and the system cannot explain or make that transparent to the patient. In the present research, without explaining analytics, students would fail to understand why they are assigned certain scores, and especially, how they can improve their designs. Instructors would not relate analytics to design processes.

One common pain point for instructors in our study is assessing contribution in team projects. Further research can investigate computing design creativity analytics, team member-wise, to understand efficacies of providing instructors and students with insights that are complementary to teamwork assessment—e.g., peer ratings on contribution to work, interaction with others, and skills—using tools such as CATME [214].

Kerne identified the interface as an integrated conceptual and sensory border zone, which supports interplay among humans and technologies [215]. To address challenges of traversing social

/ AI border zones for supporting design course contexts, we developed ideas for how to give design analytics dashboards participatory, multi-function roles in algorithm-in-the-loop assessment. Dashboards that situate presentation of and interaction with analytics have the potential to stimulate instructors' intervention and students' continuous improvement of work. The dashboards are expected to help students, by giving them new forms of on demand, transparent feedback on their design work. The dashboards will help instructors, by giving them new views, both individual and aggregate, which provide insights about what is going on in their classes, in terms of how students are accomplishing and failing. In turn, instructors can use these dashboard views to formulate new plans for what students need.

In addition to meeting instructor and student practices and needs, dashboards can simultaneously situate their algorithm-in-the-loop interactions to iteratively refine and validate derived analytics. By providing these stakeholders with affordances for indicating whether or not an analytic and its derivation make sense, and input of expected value and rationale when AI does not match, dashboards have the potential to help generate and refine labeled analytics datasets, and so make models work better.

To increase the intelligibility of analytics, we propose integrating dashboard presentations with design learning environments. Connecting the instructor, as well as AI assessment and feedback, with design elements of concern can help address problems of how students lose track of feedback, another common pain point for instructors in our study. We further developed ideas involving digital redlining over student work, as a special feedback layer, in various design tools, to assist with tracking.

Situating instructor and student interaction with analytics—by integrating dashboards with actual design work, in design environments—has the potential to transform design education. To keep pace, tools such as Photoshop, Illustrator, and Sketch need to incorporate support for learning. We expect such integrated environments to become vitally important in project-based education. The situating analytics approach can be expected to, further, add value to writing tools, such as Word and Docs.

A bonus is that the feedback features have the potential to be repurposed in professional design and writing work. Inasmuch as students learn and create in integrated environments, they are likely to want to keep using them, as they graduate and become professionals. Project-oriented work, involving creative tasks, is the least susceptible to being eliminated by automation [216]. Ironically, the incorporation of AI—as scaffolding in design education—into project-based work, thus has the potential to become a mainstay of human involvement and performance in the future of work.

Understanding beneficial practices and challenges is critical to developing AI support for educating skilled designers, who can solve complex, sociocultural problems [217]. Building a community of stakeholders, through a co-design approach, creates a supportive environment and provides a foundation for sharing complementary expertise. This is vital for conceptualizing transformational tools, techniques, and environments [218]. Future research can beneficially involve more stakeholders—such as students, teaching assistants, graders, administrators, industry representatives—and span a range of institutions. It can employ additional qualitative methods—such as participant observation and contextual inquiry during various forms of assessments within a course—to further understand situated practices and identify contextual properties. Further, future research can further develop and validate situating analytics for design education. It can investigate how situating analytics can contribute both to other design contexts and to other educational contexts.

This research has the potential to enable transforming the role of AI in project-based education and work. New forms of computation can take support roles, based in situating analytics that provide on demand assessment and feedback. Situating analytics will integrate interaction with them into user experiences, supporting learning and work. That is, derivation of and interaction with analytics need to be interwoven, in order for the analytics to be rendered meaningful and actionable components of education and work experiences. The role of algorithms-in-the-loop is to give assistance. We do not advocate replacing or reducing instructors. Rather, we theorize situating analytics as an alternative, complementary channel or modality of assessment and feedback, which can add value to design education. In this vision, instructors and students sustain as inceptors, facilitators, and arbiters of creativity.

4. INTEGRATING MEASUREMENT AND PRESENTATION OF A VISUAL DESIGN CHARACTERISTIC FOR EXPLAINING AI-BASED ANALYTICS

We use the situating analytics paradigm, which we developed in the last chapter, to guide our derivation and presentation of AI-based analytics. We focus on providing instructors insights into students' visual design work, regarding how they use space and scale to organize elements.

Design education demands continue to grow, resulting in larger class sizes; instructors face challenges in providing timely assessment and feedback [5]. Actionable and justified assessment and feedback, which students can utilize for improvement, are critical to their learning and progress [44]. Alternative forms of human assistance that have proven useful in supporting instructor efforts include peer [6, 7] and crowd [8, 9, 10]. The present research investigates AI support, which can complement human assistance by providing the ability to process big data at speed [11].

Multiscale design, according to Lupfer et al., is "the use of space and scale to explore and articulate relationships, [which] involves the juxtaposition and synthesis of diverse design elements" [16]. *Scale* refers to the zoom level: elements at the same scale are "equally legible at the same viewport zoom" [70]. The juxtaposition and synthesis of elements, across scales, often results in nested spatial groups or *clusters* (See example in Figure 4.1). Prior computational approaches provide effective approaches for the assessment of visual organization—e.g., in website design—at a single scale [42, 41, 219]. The research gap is a dearth of prior computational approaches to assessment of multiscale design.

Why does multiscale design matter? Ray and Charles Eames demonstrated how we humans conceptualize our knowledge of the universe across powers of ten [220]. Tufte articulates micro / macro readings, a strategy for constructing data narratives in which designers employ "fine details" that accumulate to form "larger coherent structures" [178]. Perlin and Fox conceptualized and actualized the importance of organizing information across scales, in the form of a zoomable user interface (ZUI) [221]. Bederson discusses how a ZUI helps people in developing a mental map of the information by taking advantage of human memory and spatial perception [222]. According

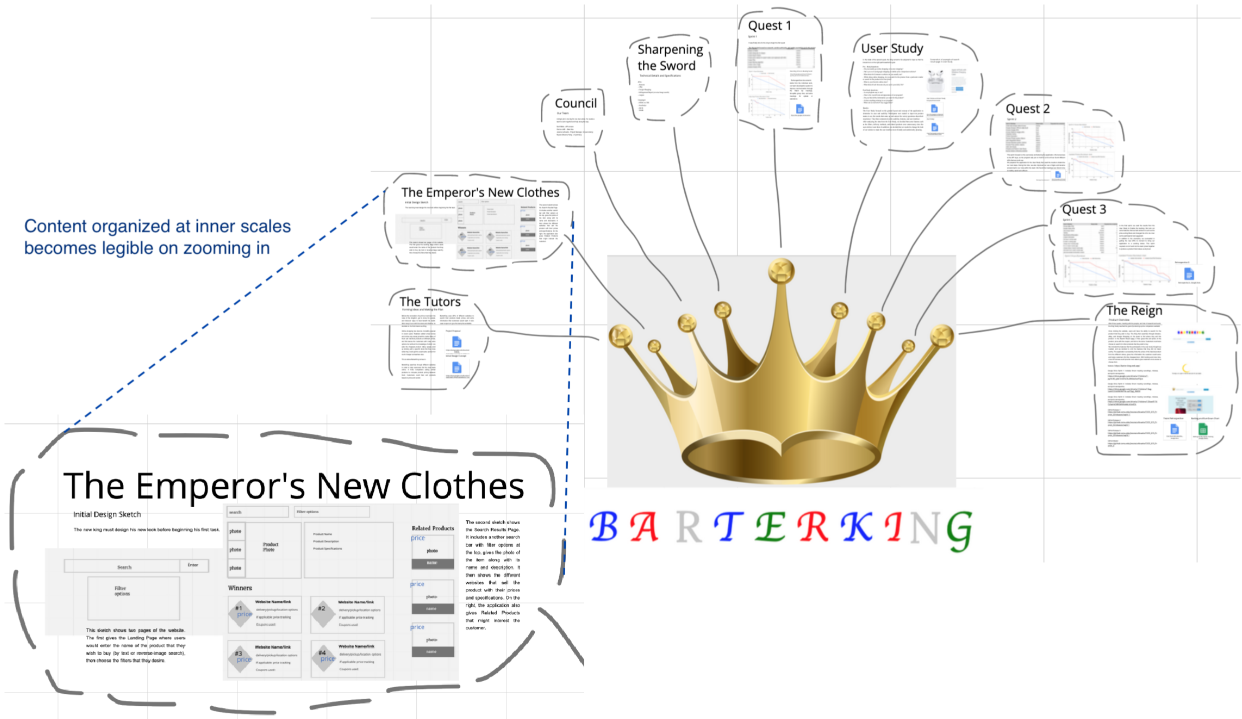

Figure 4.1: Multiscale design produced by a student in instructor I4's human-computer interaction project. The student organized different phases of the project as clusters of design elements, nested across scales of magnification. At the outer scale, the crown, its caption, and spikes connected with each phase together form an overall cluster. In the bottom left, we see the view as the user zooms in. Elements deeply nested in one scale become legible at the next. We observe nine nested clusters at the inner scale.

to Barba, multiscale design supports a holistic analysis and development of ideas, across multiple scales, by allowing people to shift cognitive point of view up and down hierarchies [69]. Bar-Yam describes multiscale design as an approach to manage increasing complexity [223]. Design environments supporting multiscale design—e.g., Photoshop, Illustrator, InDesign, and IdeaMâché [70]—enable going beyond 2D by organizing content across scales, i.e., across a range of zoom levels. Lupfer et al.'s investigation of a landscape architecture classroom found that creating and connecting representations across scales pervades student design work on projects that analyze conditions, and then conceptually and schematically form design proposals to meet the situated needs of sites involving waterways and land use [16]. Design students' use of a multiscale design environment has been found to support them in iterative and reflective ideation [70].

Summers and Shah discuss that deriving objective measures of complex characteristics within a design can benefit educators in assessing student capabilities [224]. To our knowledge, there is no empirical evidence yet that objective measures of students' multiscale design benefit instructors in their educational processes of design. Hence, our first research question is:

RQ1: Is multiscale design a significant characteristic of visual design that instructors value in situated course contexts?

To probe this, the present research begins to explore AI techniques that allow derivation and presentation of analytics—the number of scales and clusters—measuring multiscale design organization. Like recent *AI-based analytics* research [14, 15], we adopt a broad view of analytics, which includes not only facts but also inferences. In deriving and presenting AI-based analytics, we focus on conveying the meaning of measures, by supporting instructors in relating the analytics to students' design project work. If instructors find the analytics useful and think that the analytics can be given to students on-demand, then the ability to relate analytics to design work can also later support students in understanding shortcomings and improving their work. Hence, our corresponding research questions are:

RQ2: How, if at all, can multiscale design be computationally modeled so as to constitute a basis for deriving and conveying the meaning of AI-based analytics?

RQ3: How, if at all, do AI-based multiscale design analytics and how they are presented support instructors' assessment and feedback experiences in situated course contexts?

In order to develop a computational model of multiscale design that constitutes a basis for conveying the meaning of AI-based analytics, we explored *AMOEBA* clustering algorithm, which recursively determines spatially proximal groups of elements [3]. We chose AMOEBA for three main reasons: 1) its ability to computationally model nested spatial relationships; 2) the possibility of making the basis of clustering visible to the user, which thus aids conveying the meaning of derived analytics; and 3) AMOEBA's computational efficiency and the ability to operate without the number of clusters as an input, which one does not expect to be fixed in advance for design works. We adapted AMOEBA for design contexts where elements have dimensions, i.e., width and height, as the algorithm had previously been used for point elements. Using the adapted AMOEBA as the basis, we derived multiscale design analytics. We derived quantitative measures of scale and cluster recognition performance, including precision, recall, and F-score.

To investigate whether and how multiscale design analytics—as well as how they are presented—support instructors' assessment and feedback, we introduced the analytics into course contexts. We situated instructors' interaction with the analytics, via a dashboard technology probe, integrated with a free-form, multiscale design environment. Through this integration, we made relationships visible between the analytics and particular design element assemblages that they measure. We used the technology probe [17] methodology to simultaneously: (a) understand instructors' needs and desires in a real-world setting, (b) collect data through field-testing of analytics and their presentation, and (c) stimulate instructors' and our own thinking about technological possibilities. We derived initial findings from a qualitative analysis of instructors' interviews regarding their experiences with multiscale design analytics.

In the next section, we present prior work relevant to various aspects of our investigation. We then describe our probe, comprised by the adapted AMOEBA algorithm and the dashboard integrated with multiscale design environment. We follow with a mixed methods evaluation methodology and findings, including quantitative measures of precision, recall, and F-score, and qualitative analyses of both the AI recognition and instructor experiences. We discuss how ambiguity in the interpretation of design work serves as a resource for explaining AI-based analytics. We derive implications for users' independent assessment and interactive understanding of AI outcomes. We reflect on the role of perceptual grouping principles in visual organization and potential for improving the AI performance. We conclude by reiterating the contributions of the present research and potential broader impacts.

4.1 Prior Work

We situate amidst and differentiate from prior research that investigated multiscale design organization and its assessment, computational approaches for design analysis, the use of spatial clustering algorithms for modeling design organization, and the use of learning analytics and dashboards.

4.1.1 Multiscale Design Organization: Strategies and Assessment

We draw on multiscale design strategies introduced in Section 1.1.6, in particular, the use of scale transitions for defining spatial relationships among design elements: zooming out shows "encompassing contexts" and zooming in shows "nested details". Prior work assessed multiscale design organization by counting the number of scales, based on the number of times one needs to zoom in, in order to make inner elements, i.e., the *nested details* legible [70]. Prior work used the organization of ideas in spatial groups as one of the criteria for a holistic evaluation of ideas in a free-form environment [35]. However, human raters performed these assessments. The present research develops computational methods for assessing the number of scales and clusters present in a design.

4.1.2 Computational Design Assessment

Gero and Maher show that computational models for creative design, such as based on AI, have improved our understanding of design [225]. Researchers' investigation contexts include idea generation [36, 37, 38], re-design [39, 40], and assessment [41, 42]. We focus on the use of computational approaches for design assessment, in particular, the visual organization of ideas. Simultaneously, we consider whether these prior approaches make the visual and conceptual basis underlying assessment visible to the user.

Reinecke et al. assessed website aesthetics by developing a regression model based on attributes such as color, symmetry, and the number of images and text groups [41]. The aim of the work, however, was to predict users' first impression of a website, not the presentation of the underlying basis or actionable insights to any end-users per se. Oulasvirta et al.'s Aalto Interface Metrics web service is aimed at providing assessments of a graphical user interface design, to help designers in identifying and addressing the shortcomings [42]. In addition to providing the computed metrics—for characteristics, such as visual clutter, colorfulness, and whitespace—they make visible the approach used for computing the metric, as well as, in some cases, provide a visualization of the corresponding design characteristic.

Our work most closely resembles Koch and Oulasvirta's computational layout perception of website design [226], in regard to computationally identifying the hierarchical groupings in the given design. The differences, however, are that the website design is typically legible at a single zoom level and it is not performed in a free-form space. To our knowledge, the present research is the first attempt to assess free-form, multiscale design organization.

4.1.3 Spatial Clustering

Spatial clustering refers to the partitioning of "spatial data into a series of meaningful subclasses called spatial clusters, such that [elements] in the same cluster are similar to each other, and are dissimilar to those in different clusters" [227]. Taking inspiration from recent success on the use of clustering for modeling design in 2D space [181], we investigate the potential of algorithms that identify clusters within clusters, for the purpose of modeling multiscale design organization.

Liu et al. and Deng et al. classify spatial clustering algorithms into seven groups: partitioning (e.g., K-Means and CLARANS), hierarchical (e.g., CURE, CHAMELEON, and BIRCH), density-based (e.g., DBSCAN, OPTICS, and ADACLUS), graph-based (e.g., AMOEBA and AUTOCLUST), grid-based (e.g., WaveCluster and STING), model-based (e.g., EM and Geo-SOM), and combinatorial (e.g., CSM and ICC) [227, 228]. Further, as they discuss, a majority (19 out of 23, except AMOEBA, AUTOCLUST, ADACLUS, and ICC) of these algorithms require prior knowledge, such as the number of clusters and merge/split conditions. In design contexts—where differences in the representation of ideas are a rule rather than an exception [225]—one does not expect the number of clusters to be fixed in advance.

Out of the four algorithms that are not heavily reliant on prior knowledge, only AMOEBA [3] and AUTOCLUST [229] have been employed for multiscale clustering, i.e., clusters within clusters. AUTOCLUST is an extension of AMOEBA, with additional criteria addressing adjacency of sparse and high-density clusters. Both these algorithms take a Delaunay triangulation based approach to determine the nested clusters. The algorithms, however, were previously used for

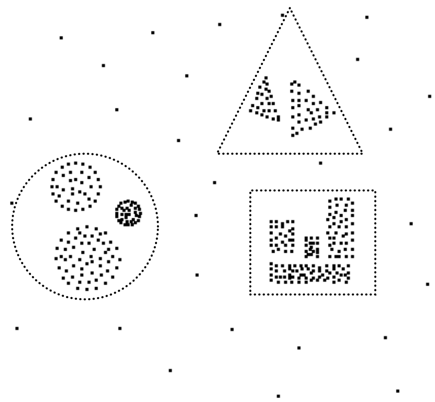

Figure 4.2: The AMOEBA algorithm performs multiscale clustering of point elements [3]. We adapted the algorithm to perform multiscale clustering of design elements that have dimensions, i.e., width and height. Reprinted from [3].

multiscale clustering of point elements (Figure 4.2). The present research adapts the Delaunay triangulation based approach for design contexts, in which the elements typically have dimensions, i.e., width and height (Figure 4.1). We evaluate the efficacy of the adapted algorithm through precision, recall, and F-score measures.

4.1.4 Learning Analytics and Dashboards

Learning analytics and dashboards technologies have been found to support instructors in identifying student problems and intervening, which improved student retention and success [230]. For design education contexts, which involve project-based learning, we draw on the need for measuring complex characteristics that can provide instructors insights into students' creative processes (Section 1.1.4). The present research begins to address the need by deriving and presenting AI-based analytics—the number of scales and clusters—corresponding to students' multiscale design organization. We focus on conveying the meaning of analytics. We report findings from qualitative analysis of instructors' interviews, regarding the usefulness and limitations of AI-based analytics and their presentation.

4.2 Technology Probe

Our probe is comprised by: (1) a free-form, multiscale design environment, in which students work; (2) the adapted AMOEBA algorithm for multiscale clustering of student design work; and (3) a dashboard integrated with the multiscale design environment, which makes relationships visible between analytics and student design work.

4.2.1 Multiscale Design Environment

We draw on a free-form, multiscale environment that supports real-time collaborative design ideation, by helping users discover and interpret relationships through visual thinking [208]. Users flexibly organize design elements, across space and scale, by invoking creative strategies of collect, assemble, sketch, write, shift perspective (pan and zoom), and exhibit [72]. Design element types include text, image, video, shape, sketch, Google docs, and Google maps. Users collect elements by dragging and dropping a variety of content types from the web. They assemble these elements through transformation operations, such as scale (resize), rotate, layer, crop, and blend. They pan and zoom to compose content across space and scale. They annotate ideas through sketching and writing. They exhibit their work using permanent URLs. Figure 4.1 shows an example work produced using the environment.

The environment stores design representations, as JSON, including semantic information, such as position (x and y), width, height, and any transforms applied to the constituent elements. We used this semantic information as input to our adapted multiscale clustering algorithm.

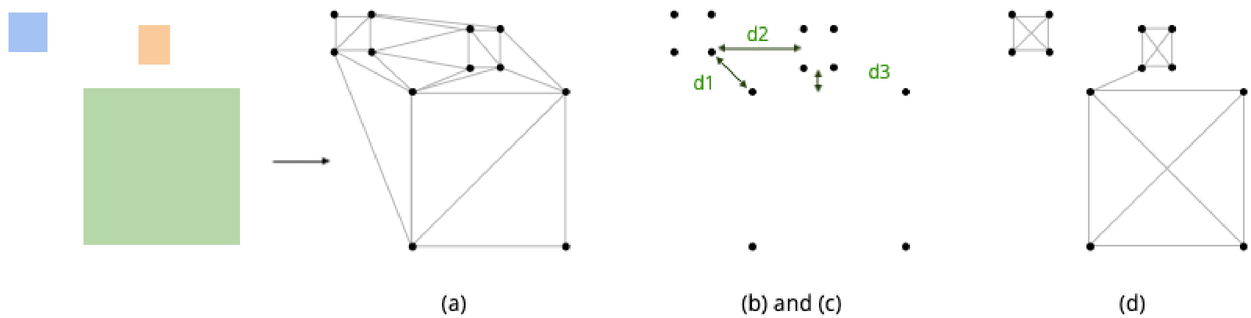

Figure 4.3: Adapting AMOEBA for elements having width and height. The adapted algorithm adds four aspects to the original algorithm: (a) compute Delaunay graph using all vertices of an element, (b) instead of using Delaunay edges directly, find the shortest distance between elements represented by the edge, (c) relatedly, ignore any edges between vertices of the same element, and (d) when calculating subgraphs, add edges among all vertices of an element. See detailed reasoning in Section 4.2.2.

4.2.2 Adapted AMOEBA Multiscale Clustering Algorithm

AMOEBA has been used for multiscale clustering of point elements, using Delaunay triangulation to recursively determine nested clusters based on distances among elements [3]. It compares the lengths of edges that are incident on each element with the average global edge length among all elements in the Delaunay graph. A single hyperparameter α (alpha) controls the homogeneity/heterogeneity of clusters: the bigger the value of this hyperparameter, the more heterogeneous distances exist amongst elements in a cluster. The algorithm does not need the number of clusters as input, and it is computationally efficient, requiring only *O(nlogn)* time.

We adapted the algorithm to cluster design elements that are dimensional, i.e., have width and height. We added four aspects to the original algorithm (see Figure 4.3; the combined algorithm with these changes can be seen in Appendix C):

(a) When computing the Delaunay graph, input each element as its four vertices. If we only consider the center, then elements having large width/height will have an incorrect large separation with nearby elements, in the graph.

(b) When iterating through the edges of the Delaunay graph for computing distances, instead of directly using the edge length, find the shortest distance between the two elements connected by each edge. Delaunay triangulation forms edges with nearby elements, but the edges in the Delaunay graph are not always the shortest distance.

(c) Relatedly, ignore any edges between the vertices of the same element. The distance of an element to itself is zero.

(d) When calculating subgraphs, collapse vertices of each element by adding edges among all its vertices. Without such intra-element edges, two vertices of the same element having inter-element edges with two nearby elements will incorrectly fall into two different subgraphs.

4.2.3 Dashboard Integrated with the Multiscale Design Environment

We developed a dashboard interface that allows instructors to manage a course, its assignments, and submission to the assignments. We put the situating analytics paradigm we developed in Chapter 3 into practice, connecting the dashboard presentation of the analytics with the design environment in which students performed their work. When an instructor views an assignment page (see Figure 4.4 top) on the dashboard, it presents to them the students submissions and scale and cluster analytics derived through the AI (i.e., adapted AMOEBA) analysis of submissions. Instructors can click the 'Submitted Design' link to view the design submitted design in its original form (e.g., Figure 4.1), or they can click on an analytic to view its basis within the submitted design.

When an instructor clicks an analytic on the dashboard assignment view, it opens the respective submission within the design environment and presents to the instructor an animation of the scales and clusters—determined by AI—within the design (Figure 4.4 bottom). This makes the relationships visible between design element assemblages and the analytics that measure them, thus allowing instructors to view the basis of derived analytics. All clusters that are at the same scale are presented using the same background color, so as to help discern nested relationships.

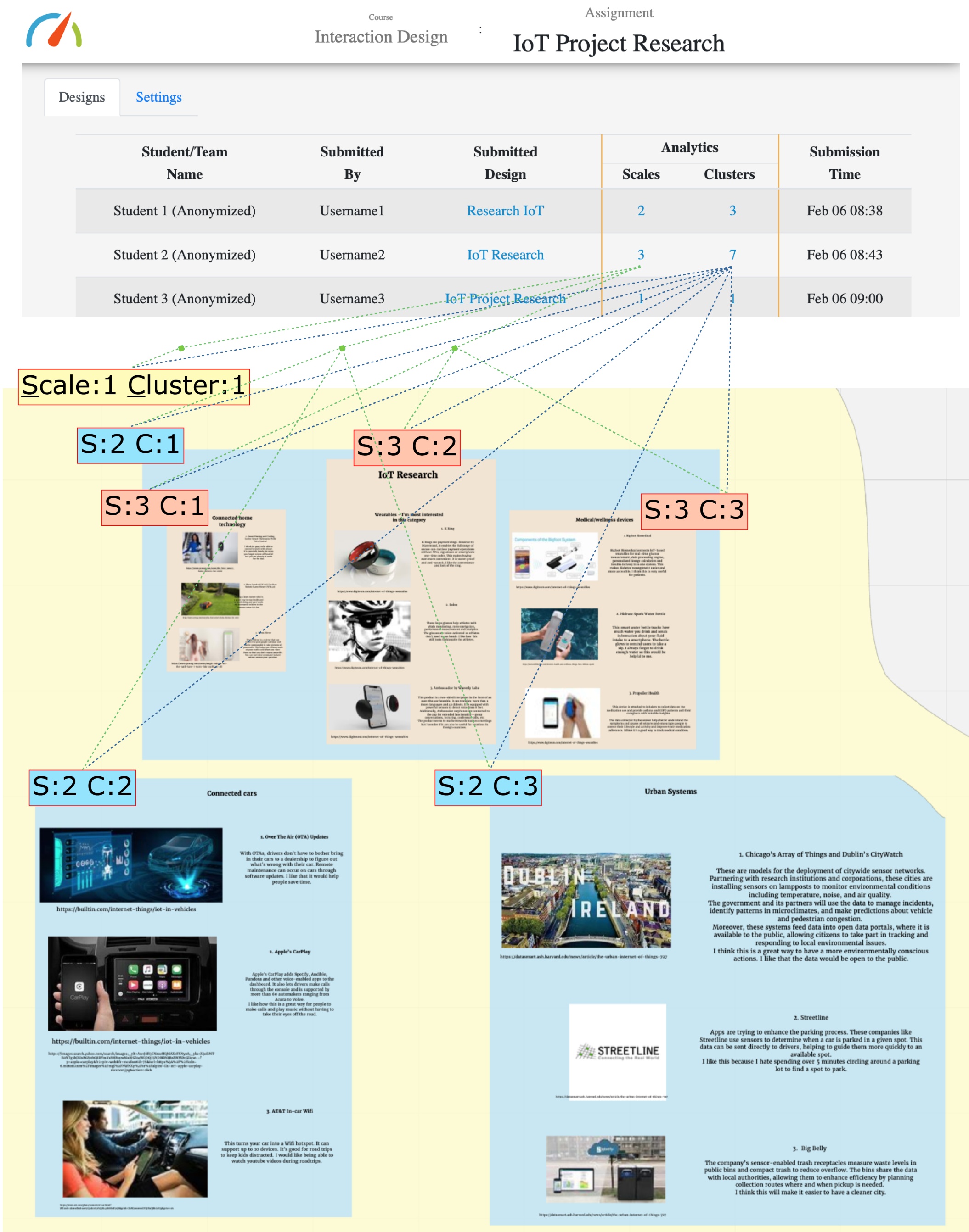

Figure 4.4: We integrated a dashboard with the actual design environment with the goal of explaining the analytics to users, in context. Clicking an analytic presented on the dashboard interface opens the actual design environment and shows the corresponding AI identified nested scales, with all clusters at a particular scale rendered in the same background color. In the above figure, the outermost scale has one cluster—including all design elements—which is rendered in yellow color. The next inner scale has three clusters (one at top and two at bottom), which are rendered in blue. The innermost scale has three clusters—within the top blue cluster—which are rendered in brown. The visualization makes relationships visible, between particular design element assemblages and analytics that describe and measure them. It enables instructors to interpret and comprehend what the analytics mean.

4.3 Methodology

We describe our methodology of: (1) creating a labeled dataset of scales and clusters; (2) deriving precision, recall, and F-score measures for multiscale spatial clustering by adapting prior information retrieval measures; and (3) qualitative data gathering through instructor interviews and its analysis drawing on Charmaz's approach to grounded theory.

4.3.1 Labeling Scales and Clusters

Prior work evaluates clustering performance against a set of 30-60 labeled clusters. This includes both quantitative [231, 232] and qualitative [233, 227, 3] evaluations. As we evaluated both scales and clusters, we aimed to have at least as many labels as prior work for both. We selected designs from a variety of courses, in proportion to the number of works produced in each course. At an operational level, the selection criteria were: design work with a non-trivial number of elements ($\sim$30 or more), which were organized at multiple scales (2 or more), i.e., zoom levels. We also selected a few designs having a single scale or a small number of elements, to validate that the AI does not fail in such cases.

We developed guidelines for human raters to label clusters across scales. Following the guidelines, two raters labeled scales and clusters within a total of 39 student design works. Through the labeling process, we obtained an average total of 107 scales and 653 clusters. The inter-rater reliability score (Cohen's kappa) for scales was 0.88, indicating a near-perfect [234] agreement. The inter-rater reliability score for clusters was 0.71, indicating a substantial agreement.

In the absence of labels, it is challenging to ensure apriori that a sample set is representative of the population. Hence, we did a posterior analysis of the scale and cluster distribution within the labeled set. The distribution plot (Figure 4.5) indicates a variety of organizations within the labeled set. A majority of the designs use 2 or 3 scales (average = 2.79), which aligns with prior work [70].

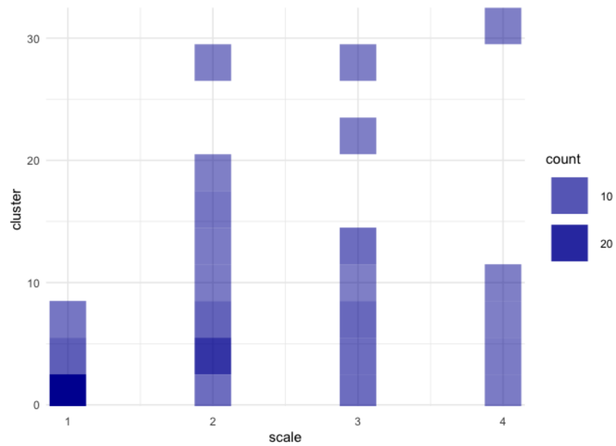

Figure 4.5: The distribution of the numbers of clusters across different scales, in our labeled set, indicates a variety of organizations used as the basis for evaluating the multiscale analysis approach. We used cluster bins of size 3 to smoothen the y-axis representation.

4.3.2 Precision, Recall, F-Score Measures for Multiscale Clustering

Precision, recall, and F-score are widely used measures for evaluating clustering performance in both spatial and non-spatial domains [231, 235]. However, in spatial domains, these measures have only been used for evaluating clusters at a single scale. For multiscale spatial clustering, prior work only performed qualitative evaluation. Hence, for quantitative evaluation, we adapted measures developed for multilevel clustering in data mining and information retrieval [236].

Let human labeled clusters be represented by classes (L_1, L_2, L_3, ..., L_c). Then, given a class L_r of size n_r and an algorithm-identified cluster C_i of size n_i , if n_{ri} elements in the cluster C_i belong to L_r, then precision (P), recall (R), and F-Score (F) are computed as:

$$P(L_r, C_i) = \frac{n_{ri}}{n_i} \quad , \quad R(L_r, C_i) = \frac{n_{ri}}{n_r}$$

$$F(L_r, C_i) = \frac{2 * P(L_r, C_i) * R(L_r, C_i)}{P(L_r, C_i) + R(L_r, C_i)}$$

Then, for each L_r, its F-score is the maximum F-score value attained at any node in the tree T of nested clusters, i.e.,

$$F(L_r) = \max_{C_i \in T} F(L_r, C_i)$$

The net F-score of the clustering algorithm is then the aggregate individual class F-Score weighted by the class size:

$$F\text{-}Score_{cluster} = \sum_{r=1}^{c} \frac{n_r}{n} F(L_r)$$

Extending this approach for evaluating scale recognition performance, given s_h human-labeled scales and s_a algorithm-identified scales, for each human-labeled scale M_r, form its pair with the maximum matching scale S_i, based on the sum of $F(L_{ri})$ values, i.e., the F-scores of matching clusters present on the scale i:

$$F(M_r, S_i) = \sum_{ri} \frac{n_r}{n} F(L_{ri}) \qquad F(L_{ri}) : F(L_r)\ that\ matches\ on\ scale\ i$$

$$F(M_r) = \max_{1 \leq i \leq s_h} F(M_r, S_i)$$

If more than one human-labeled scales match the same algorithm-identified scale, form the pair with the human-labeled scale for which the sum of $F(M_r)$ values is higher. Then, for s_{ha} matches between s_h human-labeled scales and s_a algorithm-identified scales:

$$P_{scale} = \frac{s_{ha}}{s_a} \quad , \quad R_{scale} = \frac{s_{ha}}{s_h} \quad , \quad F\text{-}Score_{scale} = \frac{2 * P_{scale} * R_{scale}}{P_{scale} + R_{scale}}$$

ID	Gender	Course	Field	Role	Semester
I1	F	Digital Media Design, UI/UX for Games	Interactive Art & Design	Course Instructor	Spring 20
I2	F	Interaction Design	Interactive Art & Design	Course Instructor	Spring 20
I3	F	Engineering Design	Mechanical Engineering	Course Instructor	Summer 20
I4	M	Programming Studio	Computer Science and Engineering	Course Instructor	Spring 20
I5	M	Programming Studio	Computer Science and Engineering	Course Instructor	Spring 20
I6	F	Programming Studio	Computer Science and Engineering	Lab Instructor / TA	Spring 20
I7	M	Programming Studio	Computer Science and Engineering	Lab Instructor / TA	Spring 20
I8	M	Programming Studio	Computer Science and Engineering	Lab Instructor / TA	Spring 20
I9	M	Programming Studio	Computer Science and Engineering	Lab Instructor / TA	Spring 20

Table 4.1: Five design course professors and four teaching assistants in the role of lab instructor interacted with AI-based analytics, presented via the dashboard integrated with multiscale, free-form design environment.

4.3.3 Design Course Deployments: Instructor Interviews

Five design course instructors and four teaching assistants (TAs) in the role of lab instructor—across three departments—interacted with AI-based analytics, presented via the dashboard integrated with multiscale, free-form design environment, during Spring and Summer 2020 (Table 4.1). We conducted semi-structured interviews with the instructors and TAs regarding their experiences at the end of the respective courses. At times, instructors explored analytics during the interview. Due to Covid-19's impact, instructors spent significant time moving courses online, which adversely affected planned use.

Drawing on Charmaz' approach to grounded theory qualitative data analysis [18], two authors first performed initial coding of three interview transcripts. They met to bring their initial codes into alignment, and formed tentative categories. Then, they performed focused coding of the remaining interview transcripts, revising codes and categories, as needed, to suitably represent the salient phenomena. We present the categories, including participant quotes illustrating the phenomena, in Section 4.4.3. Interview questions can be seen in Appendix D.

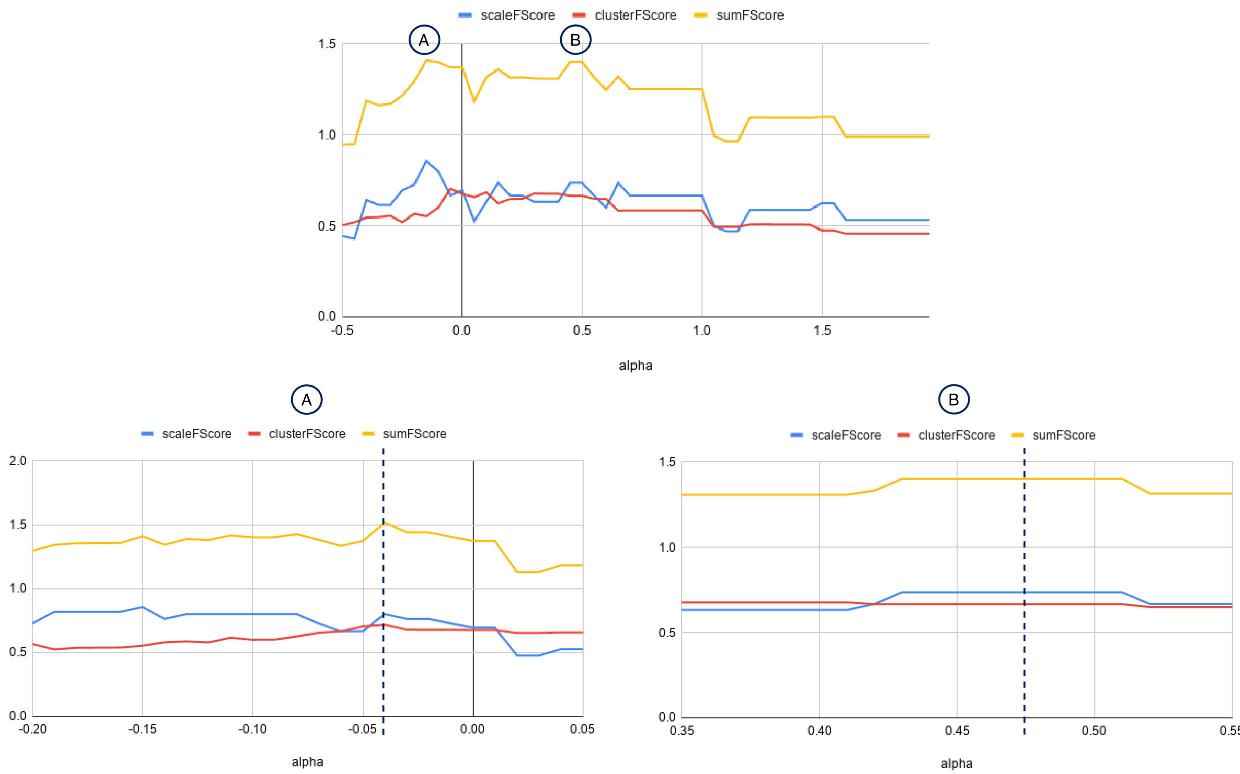

Figure 4.6: Hyperparameter tuning: AMOEBA uses a single hyperparamater α to control the heterogenity of distances among elements that are clustered together. We used 10% of the labeled data to tune the hyperparameter, searching for optimal values on both sides of default value suggested by prior work (α=1). Observing two peaks (A and B), we inspected the values of the hyperparameter at a finer level of granularity near these peaks (bottom charts). We selected two hyperparameter values (-0.04 and 0.475) from the analysis.

4.4 Mixed Methods Findings

We begin with quantitative measures—precision, recall, and F-Score—of AI performance. We next present patterns of AI mismatch with human labels, found through a qualitative analysis of AI scale and cluster recognition. We follow this with findings from our qualitative analysis of instructor interviews, regarding their experiences with the AI-based analytics, presented via the dashboard integrated with the multiscale design environment.

	α	**Precision**	**Recall**	**F-Score**
Scale	-0.040	0.531	0.686	0.599
	0.475	0.720	0.628	0.671
Cluster	-0.040	0.658	0.911	0.694
	0.475	0.536	0.979	0.61

Table 4.2: Precision, Recall, and F-Score measures, for the two values of hyperparameter α at which we observed peak aggregate performance for scale and cluster recognition in the tuning phase.

4.4.1 AI Performance: Precision, Recall, F-Score Measures

Using the approach specified in Section 4.3.2, we computed the precision, recall, and F-Score measures. In Section 4.2.2, we described that AMOEBA uses a hyperparamater α to control the heterogenity of distances among elements that are clustered together. As we adapted the AMOEBA algorithm in a new domain (i.e., design), we used 10% of the labeled data to tune the hyperparameter. As shown in Figure 4.6, searching on both sides of the default hyperparameter value (α=1) suggested by prior work [3], we found two peaks (A and B) for the aggregate F-Score of scale and cluster recognition. We further checked the values at a finer level of granularity near both the peaks (bottom charts in Figure 4.6). Based on the performance plots, we selected the value of -0.04 for peak A and 0.475 for peak B. We report precision, recall, F-Score measure for scale and cluster recognition for both the hyperparameter values (Table 4.2). We note that the range near peak B (α=0.45-0.5) is more stable, i.e., there is no sharp increase or decrease in performance near that value. We discuss the results in Section 4.5.4. The adapted AMOEBA algorithm's use of α can be seen in Appendix C.

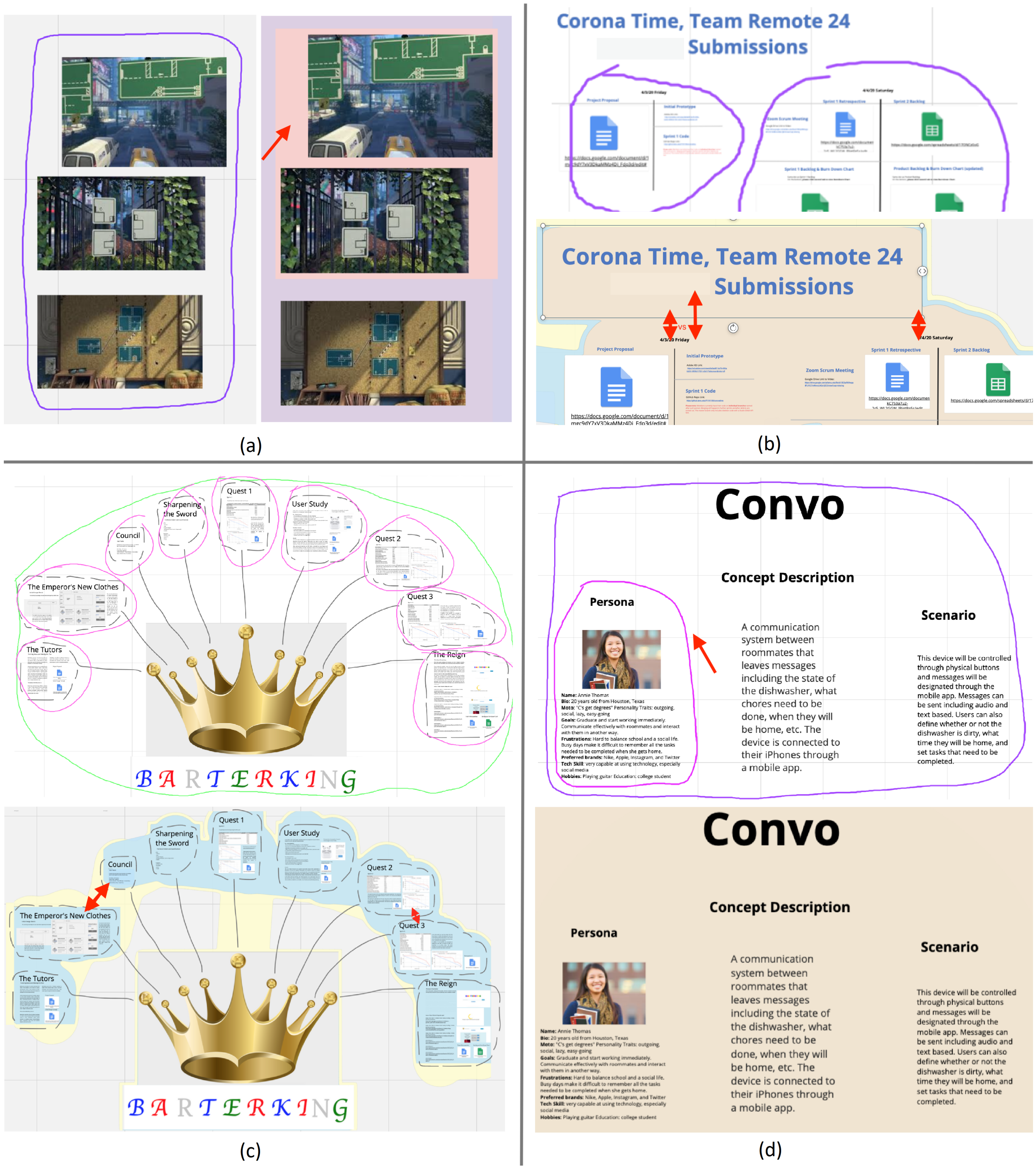

Figure 4.7: Patterns where current recognition falls short due to only using spatial proximity as the basis: (a) Extra scales and clusters: A human rater labeled (left) a cluster of all three images based on legibility vs AI (right) further clustered the top images at a new scale based on spatial distances; (b) Fewer clusters: A human rater labeled (top) two clusters vs AI (bottom) identified one cluster combining heading with the elements below due to their small distance from the bounding box of the heading; (c) Fewer clusters: A human rater labeled (top) nine clusters based on similarity of representation vs AI (bottom) identified two clusters due to relatively smaller distances at places; and (d) Fewer scales: A human rater labeled (top) an inner scale as the content becomes legible on further zooming in vs AI (bottom) identified only one scale based on spatial distances.

4.4.2 AI Performance: Qualitative Analysis

Similar to prior work, we performed visual inspection of output clusters, to qualitatively evaluate AI performance. We found four patterns where the current recognition falls short, due to only using spatial proximity as the basis:

(a) *Extra scales and clusters*: As Figure 4.7 (a) shows, while the AI is identifying extra scales and clusters on that scale, based on spatial distances, a human would not do so, as all content is equally legible. Such occurrences, therefore, result in lower scale precision as well as lower cluster precision scores.

(b) *Fewer clusters (headings)*: As Figure 4.7 (b) shows, due to small spatial distances with the bounding box of a heading, AI collapses multiple clusters into one and includes the heading with these otherwise distant clusters. Hence, the AI not only does not find many clusters a human rater labeled, but rather, it finds an additional, different cluster. Such occurrences, therefore, result in both lower cluster precision and recall scores.

(c) *Fewer clusters (similarity)*: As Figure 4.7 (c) shows, a human rater labeled more clusters due to similarities of representation whereas AI is based on proximity. Similar to (b), not identifying clusters labeled by a human rater, and identifying additional ones, such occurrences result in both lower cluster precision and recall scores.

(d) *Fewer scales*: As Figure 4.7 (d) shows, a human rater identified an inner scale because the content becomes legible on zooming in, whereas due to similar spatial distances, AI identified only one scale. Such occurrences, therefore, result in a lower scale recall score.

4.4.3 Qualitative Analysis of Instructor Interviews

We developed four categories through the analysis of interview data, which illustrate (1) how analytics provide useful information, (2) use of analytics for assessment and feedback, (3) potential benefits of analytics for students, and (4) support for exploration and understanding of analytics.

4.4.3.1 Analytics Provide Useful Information

Instructors in our study reported that analytics help them develop understandings of students' work. They gain insights such as how students developed and presented structure in their design, and whether students are able to effectively use the multiscale design environment.

> Q1 I1: *I think using the dashboard and using the analytics is really helpful for me to kind of get an understanding of what [students are] doing.*

> Q2 I2: *I've been thinking like, you know, [scales and clusters] could be a very useful information for me, you know in terms of how students developed structure and present that structure at different levels.*

> Q3 I9: *If there are multiple scales and clusters...they are at least using the environment efficiently. So if this number is extremely low for everybody...then maybe you need to take a tutorial on [the environment].*

4.4.3.2 Using Analytics for Assessment and Feedback

Instructors had a range of responses regarding the use of analytics for assessment and feedback. I1 and I4 think that the analytics can become a part of their rubrics and feedback they give to students. And that can motivate and provide students guidance on what instructors are looking for in their design.

> Q4 I1: *I think [these analytics and my rubrics] complement each other. I think it will be very helpful...if there's a way that I can just sort of make a rubric on [dashboard] and attach to when they get their feedback.*

> Q5 I4: *You know, give them something to shoot for...I think that I would say...here are the things that I'd like to see in your design...I think that I would definitely like to assign scales as a part of the rubric to say, I would like to see the big picture from out here, and then when you zoom in, see more.*

I9 shared that instead of directly using analytics for assessment and feedback, they find analytics a quick-to-use indicator of underlying problems.

> Q6 I9: *So, I won't use the values in the column to directly give them points...But it's better than having to go to every [design] and look for every single issue or having a much larger rubric that I ran by...So think of the analytics as the symptoms and [then] you actually identify diseases.*

A couple of instructors expressed that they did not find much use of utilizing multiscale analytics for assessment.

> Q7 I6: *If I was going to use it in a project [as a student] then I think it would be [more] helpful to me as compared to like if I'm just grading someone else's work.*

> Q8 I7: *I didn't find much use as such like of scales and clusters...how can I use these as an instructor?*

4.4.3.3 Potential Benefits of Analytics for Students

Instructors expressed that students can benefit from seeing these analytics, as these can become a source of reflection, as well as help students in understanding their progress.

> Q9 I5: *I'm all for giving students as much information as they can use and you know...they can use [analytics] to look at their progress.*

> Q10 I1: *Yeah, I would love students to explore more zoom levels...because usually, I think it is more like...I see it as an overall picture...but they don't really utilize being able to kind of go in to certain areas or zooming in to certain parts and elaborating.*

> Q11 I1: *[Show students] maybe spatial clusters just so that they could be more aware about how they separate.*

Instructors shared that students would benefit by seeing the analytics, but they do not want to enforce a specific organization. Their goal is to help students effectively use and reflect on multiscale organization.

> Q12 I4: *I wouldn't want them all to look the same like you don't want to go somewhere and see every painting looks the same but it was almost as if some people were painting with boards and nails and hammers versus paintbrushes and paint. They just didn't really get what they're supposed to be putting on the [design]. So then it was just like not as effective.*

> Q13 I1: *[While] they have to become a little bit more mindful of [space]...okay, then my sketches have to be in this area and my flower research is going to this area...Just seeing how they lay out everything themselves...I would rather not control whether intentionally or unintentionally at all how they see spatial clusters.*

4.4.3.4 Means for Exploration and Understanding of Analytics

Instructors found features that support exploration and understanding of analytics useful. These include analytics links that take the user to the actual design and the animation of scales and clusters within the design.

> Q14 I1: *I think I'm really enjoying these like little links that I can kind of click on it...with the scales or clusters like they can take me to those.*

> Q15 I9: *So, regarding scales, what I was able to infer [from animation] was that... there is one zoom level that has a particular region... and then they have a different zoom level that focuses on different regions and so on.*

As instructors were able to inspect the basis underlying the analytics, they expressed where AI has a mismatch with their interpretation.

Q16 I1: *I think I now have a better understanding of spatial clusters...the animation of colors changing...the outer cluster...one...two...I think I just get confused by the one added [extra] on scale 3.*

Q17 I3: *There are only four [clusters] here...and I'm not sure why [it shows here] two different ones...you've got a couple [extra] clusters.*

Further, instructors provided recommendations to enhance understandings of analytics, such as by including examples and interactive means of exploring design examples that use scale in different ways.

Q18 I4: *An example of how scales would work.. maybe one two and three static images [on dashboard] showing a zoom level of 0, at 1, and at 2.. and what additional information you can see suddenly at that [level]. And sort of an example of best practices instead of students just try to figure it out.*

Q19 I5: *So, how do I maneuver to other scales?*

Q20 I1: *Okay. I was wondering if.. whether it would be possible to kind of.. maybe like pinpoint or just kind of go to the precise scale.*

4.5 Discussion + Implications

We put the situating analytics paradigm into practice, to develop a new approach of explaining AI-based analytics, by integrating their measurement and presentation. We compare and contrast our alternative emphasis on 'explaining' AI-based analytics to prior research on 'explainable AI'. We follow by considering how ambiguity in the interpretation of design work can serve as a resource for explaining AI-based analytics. We next interpret our initial research questions in terms of our findings. We discuss the significance of multiscale design in situated course contexts. We discuss and derive implications for: (1) modeling multiscale design, including incorporation of additional perceptual grouping principles; and (2) developing novel support for assessment and feedback in design courses by explaining analytics.

4.5.1 Integrating Measurement and Presentation: An Approach to Explaining AI-based Analytics

Suchman seminally articulated that designing a useful AI system requires understanding users' *situated practice*, i.e., how their actions develop purpose and intelligibility within particular circumstances. Suchman emphasizes a system's transparency—how it conveys an AI's intended purpose to users and establishes accountability—as requisite for effectively supporting situated practice [12]. Dourish breaks this principle down to focus on how translating ideas between intellectually different domains of situated practice (social) and technology (computational) "can be both exceptionally valuable and unexpectedly difficult" [13].

In the present study, the AI is the adapted AMOEBA clustering algorithm. The context involves design courses in interactive art & design, mechanical engineering, and computer science. The characteristics of multiscale design that we measure are based on design practice [16]. The analytics measuring these characteristics—the number of scales and clusters—are based on AI's identification of nested clusters in a design using spatial positions and dimensions of the constituent elements. We built on the situating analytics paradigm we developed in Chapter 3, to make the basis of analytics visible to the instructors. We integrated a dashboard with an actual design environment, connecting analytics with the particular sets of spatially related elements they measure (Figure 4.4). Through this approach, instructors were able to get a quick sense of students' design organization, and then drill down to the work, as well as to pinpoint where the analytics mismatch their interpretation (Q2, Q16, Q17). Our argument for this approach is two-pronged:

1. We chose a particular characteristic, multiscale design, to measure. We did this instead of just getting grades for a large set of assignments and building a bottom up recognizer from that data. Such a recognizer is typically based on an arbitrary aggregation of high dimensional features ("black box") that would likely map to characteristic not explicitly discernible to design instructors or students. On the other hand, our analytics make sense to the design instructors. They can potentially inform students' design process for improving the characteristic; and

2. The interface directly presents the linkages between the analytics and the design data that they are derived from.

We found that the current approach "explains" the analytics by showing the instructors what is being measured in the design work. We argue that interface's ability to work this way is jointly dependent on the choice of the visual design analytic, because this characteristic is based on design practice, and so makes intuitive sense to designers. While there is a great scope for improving the AI performance (Section 4.5.4), in the present form, in terms of constituting a basis for explaining analytics, it is already able to serve instructors at a level where they are able to relate the measures to students' design organization.

Explainable AI (XAI) is a related approach, which heretofore has referred to work supporting human interpretability and comprehensibility [237], focused on how users "understand, appropriately trust, and effectively manage" the function of AI algorithms, in practice [238]. For example, Samek et al. explain AI in terms of descriptions of their algorithms, e.g., decomposition and partial derivatives [239]. In contrast, explaining analytics refers to giving users understanding of what analytics are describing and how they work. Using XAI can further support explaining analytics, by drawing on relationships between how AI functions and the characteristics that make sense to users, in context.

An XAI approach can superimpose a simple visualization of the present AI's Delaunay graph in a view of the design elements. This enables users to see the lengths of edges that the AI used to determine whether a cluster is nested. Another use of XAI in the present case would be a slider that allows users to see how changing the hyperparameter (Section 4.2.2, Appendix C) affects AI output of scales and clusters. Our findings suggest that AI designers support users' understanding by enabling making the Delaunay graph visible and providing a slider for showing affects of varying the hyperparameter. Likewise, at present, the only features our AI uses are positions and dimensions of design elements. In Section 4.5.4, we advocate other features—e.g., similarity based on size, color, and font—which AI can utilize. In that scenario, an XAI approach that shows the contribution of different features in determining the overall distance among elements can be useful.

4.5.2 Ambiguity as a Resource for Explaining AI-based Analytics

We need labeled data or ground truth to train and validate AI models, to make machines "artificially intelligent", to make them do what we humans do. But, which human(s) amongst us? For example, two humans looking at a picture can interpret it differently [240]. Hence, when humans label data, their interpretations, assumptions, and biases creep into AI, which is trained by / tuned to particular instance-label mappings. Addressing such AI biases, which can result in unfair outcomes, such as for individuals or groups within a population, is a key challenge that researchers are working to recognize and address [193, 194, 157].

In our study, while the two raters had a near-perfect agreement on scales (kappa=0.88)—i.e., how many times they needed to zoom in to see content legibly—and substantial agreement on clusters (kappa=0.71), sometimes one of the raters would label one more scale and several clusters on that scale. Likewise, the two raters, sometimes clustered certain elements differently, based on how they interpreted relative proximity.

Gestalt theory explains how humans perceptually group elements based on principles, such as proximity, symmetry, and continuity [241]. While certain principles have more effect than others, it is possible for two raters to group a set of elements differently, based on the relative degree to which different principles play a role in their perception [242]. Based on the same reasoning, the actual users of our system, at times, may perceptually group elements differently than one or both of the raters, i.e., different from what the AI has become tuned to recognize. Moreover, the analytics are not meant to constrain creative processes (Q12, Q13). Rather, novelty is preferred (Q12). Further, there can be additional reasons for a mismatch, e.g., difficulty in comprehending AI outcomes or AI performance limitations [243].

Implications. AI solutions that form the basis of design analytics should account for diverse perspectives and interpretations. As Gero and Maher express, in design, "inconsistencies are a rule rather than an exception" [225]. According to Gaver et al.'s maxim, in designing technology, the need for interpretation or the presence of ambiguity "is a resource that [we] should neither ignore nor repress” [244]. Gaver et al.’s prescription is to *cast doubt on sources to provoke independent assessment* [244]. We observe that this corresponds to recent guidelines for human-AI interaction, which prescribe, *make clear why the system did what it did* [245]. Our approach to explaining analytics aligns with these prescriptions. The ambiguity does not prevent the analytics from being useful.

Incorporating additional Gaver et al. and human-AI interaction guidelines can further benefit AI-based analytics research. For example, Gaver et al.’s *use imprecise representations to emphasize uncertainty* can be incorporated, e.g., by presenting scales and clusters as ranges of numbers, factoring in inter-rater reliability and AI performance measures.

Going further, AI support can be made *contestable*, providing users affordances that allow them to argue for where the AI fails to match their interpretations (Q16 , Q17). According to Woodruff et al., the ability to challenge and change AI decisions is critical to addressing bias in AI models [194]. Such input from users can become a source of diverse labeled data for the designers of AI, enabling them to bring the *artificial* closer to human intelligence, over time.

4.5.3 Significance of Multiscale Design in Situated Course Contexts

Our study shows that instructors value multiscale design as a characteristic of visual design, in situated course contexts. I4 would like students to present the big picture and then details when one zooms in (Q5). I1 would “love students to explore more zoom levels” as she does not see them “zooming in to certain parts and elaborating” (Q10). I9 would like see that students are able to effectively use the multiscale design environment (Q3). I2 values students’ presentation of structure at different levels (Q2). At the same time, multiscale design analytics were not expected to serve as a catch all measure for design. I4 talks about not wanting all designs to look the same (Q12). I1 says, “I would rather not control...how they see spatial clusters” (Q13).

We consider creative cognition’s family resemblance principle, according to which, no particular characteristics are required in the designation of a work’s creativity [33]. Rather, a family of traits tends to serve as indicative. Multiscale design, as measured here, becomes one such trait. Improving the current measures, therefore, can further aid instructors in their design education processes. Simultaneously, deriving analytics for a family of traits—e.g., aesthetics, feasibility, and originality [123, 122]—is a fruitful research avenue. As no particular trait is sufficient, there is literally no way for the analytics to be perfect. But if they work well enough, they can provide instructors and students with sufficient insights so as to (1) provide first order feedback; and (2) stimulate ongoing work.

4.5.4 Multidimensional Modeling of Multiscale Design: Improving AI-based Analytics

Design education, across fields—such as art, design, and architecture—values visual organization principles. [74]. In certain contexts, visual design has become a dominant factor in how a work is perceived [246]. The industry at the same time is producing a plethora of tools—e.g., Photoshop, Illustrator, Canva, Sketch, and Figma—to help people perform visual design. Thus, developing means for assessing peoples' visual designs and providing them instantaneous feedback, through transparent analytics, has the potential to prove transformational in supporting their creative enterprises.

Our study focused on developing multiscale design analytics. We adapted the AMOEBA algorithm for multiscale clustering of design work and evaluated the performance of the adapted algorithm using precision, recall, and F-score measures (Table 4.2). If we compare the cluster F-Score measure (0.694) with that for single-scale spatial clustering (0.79-1.0) [231], we find a drop of 0.1 to 0.3 points in performance. But determining nested clusters is more complex. Drawing analogy from natural language processing, the state-of-the-art model for named entity recognition achieves 0.95 F-score, entity linking 0.82, and relationship extraction 0.76 [247], i.e., as the complexity of the cognitive task increases, performance decreases.

Lower precision scores (Table 4.2; scale: 0.531/0.72, cluster: 0.658/0.536, for two values of α) and qualitative analysis of AI performance (Figure 4.7 (a)) correlate with instructor experiences (Q16, Q17), that AI is identifying extra scales and clusters. Lower scale recall (0.686/0.628) corresponds to analysis that shows AI is identifying fewer scales (Figure 4.7 (d)). Recall measure for clusters, albeit higher, is not perfect (0.911/0.979). In Figure 4.7 (c), correspondingly, analysis shows AI identifying fewer clusters. In these mismatches, we observe that grouping principles other than proximity—e.g., similarities of size, shape, and font—come into play. On the one hand, students seeing a mismatch between AI recognition and their intent can stimulate iterating on design to consistently separate elements. On the other hand, while proximity is a salient factor [248], humans don't always reduce spatial organization to distances.

Implications. Incorporating additional *intelligence* will be beneficial for improving AI performance. Similar to how we adapted AMOEBA for dimensional design data, it seems promising to adapt AMOEBA's variations, such as those incorporating attribute similarity [227]. In addition to proximity and similarity, incorporating the gestalt principle of 'common region'—that Koch and Oulasvirta found useful in the hierarchical grouping of website design [226]—can potentially benefit. Similar to how modeling based on user actions has proven effective in a 2D design environment [181], incorporating process data such as the user's zooming in/out can help model multiscale design. Determining visible area of design elements can further help address AI's identification of fewer clusters (Figure 4.7 (b)).

4.5.5 Supporting Assessment and Feedback in Design Courses by Explaining Analytics

Learning analytics have been found to serve a variety of purposes, including providing means for scaling feedback [45, 46], informing pedagogical action [47, 48], and supporting students' self-reflection [49, 50] and skill development [51, 52]. However, for learning analytics to be effective in open-ended, project-based contexts, there is a need to assess complex characteristics that can give insights into students' creative strategies and abilities [56]. Gero's function-behavior-structure framework supports AI analysis of complex creative design processes, such as mutation and analogy [249]. Summers and Shah discuss that quantifiable measures of complex characteristics—e.g., information entropy, coupling between components, and solvability—can help assess students' abilities to address real-world problems [224].

Our study investigates complex multiscale design organization characteristics and provides evidence for the usefulness of learning analytics in open-ended and creative project-based learning contexts of design courses. Instructors in our study reported that multiscale design analytics can support them directly or indirectly in assessment and feedback processes (Q4 , Q5 , Q6). Instructors found that multiscale analytics have the potential to inform pedagogical intervention, based on whether or not students are able to effectively utilize the design environment (Q3). They shared that providing these analytics to students can help them reflect (Q9, Q11) and improve their multiscale design skills (Q5).

Implications. The community should investigate new analytics based on assessments of complex characteristics in project-based learning contexts. While challenging to develop, these analytics have the potential to give insights into higher-order learning processes. In the present study, organization principles that design instructors expect their students to demonstrate (Q5, Q13) map to the *synthesis* category—including higher-order processes such as create, compose, and organize—in Bloom's taxonomy [250, 251]. Likewise, other project-based learning contexts—e.g., arts and humanities [252]—will likely benefit from analytics based on assessments of complex characteristics.

Simultaneously, we note an opportunity of investigating the potential of existing computational analysis work, toward supporting assessment and feedback in course contexts. For example, courses that teach website design can utilize computational support that quantifies website aesthetics—e.g., Oulasvirta et al.'s interface metrics service [42] or Reinecke et al.'s machine learning model [41]—for assessment and feedback.

In line with prior work advocating interactive explanations [253], we argue that users will benefit when interfaces to complex analytics draw on an ensemble of pedagogical and visualization techniques. Based on instructors' experiences and suggestions, we find initial evidence for these techniques, for facilitating understanding and interaction:

- *Animation*: As Tversky explains, animations can aid perception and comprehension of the fine structure of spatial and temporal relationships among different pieces of content [254]. Mayer and Moreno found that adding animations enhances learner understanding, when compared with only using verbal forms of study material [255]. In our study, we found that animation aids the understanding of complex characteristics (Q15, Q16). Further, according to Tversky, interactivity such as close-ups, zooming, and control of speed is likely to lead to a better understanding [254].

These can prove useful in addressing participants' desires, such as maneuvering to different scales (Q20 , Q19).

- *Indexical Representation*: Turnbull explains indexicality as maps over space and time which convey information that can only be completely understood within the specifics of the given context [187]. In this present study, we found the indexicality of dashboard—i.e., connecting analytics with the actual design element assemblages that they measure—supported instructors in understanding the analytics. (Q14 , Q15). In addition, an intermediate representation on the dashboard, which indexically connects with the actual design work can further support the understanding of complex analytics. For example, a tree-like pendogram visualization of scales and clusters [3] on the dashboard can support understanding the nested structure and exploring a specific scale or cluster (Q20).

- *Exemplification*: The use of examples in teaching provides a device for explanation and communication [256, 257]. It helps students in concept formation and development of skills. Exemplification has the potential to help students understand analytics' complexity and so inform their design processes, to meet instructors' expectations (Q18).

Investigating whether and how these techniques contribute to interfaces for presenting a range of complex analytics would be a fruitful avenue for future research. Such research can pinpoint whether particular techniques are more useful in certain educational disciplines, in comparison to others.

The current research provides evidence for *descriptive analytics*, i.e., analytics provide insights into student work [258]. Going ahead, with data from the past iterations of a course, these analytics have the potential to function as *prescriptive analytics*, i.e., provide instructors and students alerts and suggestions based on computational modeling of the relationship between analytics and students' course performance [230, 258].

Also, so far, to probe the potential usefulness, we presented the analytics to all the instructors. However, in our study, not all instructors found all analytics to be useful (Q7 , Q8). Hence, for returning users, the analytics can be made selectable, i.e., instructors can configure a subset of analytics when setting up a course and its assignments.

4.6 Conclusion

Our study demonstrates the potential of design analytics for providing instructors transparent insights into student design work and so support their assessment and feedback efforts. The present research contributes: (1) how to explain multiscale design analytics derived using AI—thus validating the usefulness of situating analytics paradigm—by linking dashboard presentation of design analytics with the actual design work that they measure and characterize; (2) the adapted AMOEBA algorithm, for addressing multiscale clustering of dimensional elements; (3) measures of spatial scale and cluster recognition derived based on prior work in information retrieval; (4) establishing baseline performance, through a quantitative evaluation of AI's recognition of multiscale design characteristics; and (5) implications for explaining AI-based analytics through a qualitative analysis of instructors' experiences with the analytics in situated course contexts.

While we are yet to achieve the best possible recognition, we have identified multiscale design as a significant characteristic and begun to address how to use AI for measuring it. We have created a technology probe that enables us to begin to understand the implications of multiscale design analytics, in practice.

We invoked creative cognition's family resemblance principle to explain that there is no single way to define or interpret creativity. Although this poses a challenge toward measuring creativity, it simultaneously creates opportunities for the interface to "explain" the specific way AI is measuring particular characteristics, which allows users to agree or disagree with the AI. Our integration of the dashboard presentation with the actual design environment allowed instructors to independently assess the basis of analytics, i.e., the particular sets of design element assemblages that the AI determined as nested clusters. This makes the interface to the AI-based analytics visible, or as Bellotti and Edwards said, intelligible and accountable [159].

Explaining AI-based analytics is vital—similar to the needs observed for healthcare [259, 260] and criminal justice [261] domains—as the measures can directly impact outcomes for an individual. Analytics that do not connect with students' design work would have little meaning for instructors, if at all. Students, if provided with analytics, would fail to understand and address the shortcomings that they indicate. Further, explaining analytics creates opportunities for developing interface affordances that allow users to contest the AI outcomes, which is particularly important when AI recognition may not match human interpretation due to multiple reasons, such as the training/tuning data used, limited AI performance, and insufficient clarity in the presentation of AI outcomes.

Significant implications for future research are stimulated by the current level of investigation of the particular multiscale design analytics in particular situated course context classrooms. We need further investigation of how these as well as new multiscale design analytics affect other design education contexts and design in industry. Such research can investigate the extent to which different pedagogical and visualization techniques—e.g., animation, indexical representation, and exemplification—are beneficial in specific contexts. Transparent insights on design work can prove vital in improving learners' creative strategies and abilities, which in turn can stimulate economic growth and innovation [262]. Continued efforts toward simultaneously satisfying the dual goals of explainability and performance, across a range of contexts, has the potential to create broader impacts by providing inroads to addressing complex sociotechnical challenges, such as ensuring reliability and trust [158] in the use of AI systems.

5. IMPLICATIONS AND CONCLUSION

In this chapter, based on our investigations, we derive implications for supporting creative human learning experiences through the use of AI. We begin by discussing the transformational impact that making AI useful for complex human activity can bring to society, focusing on measures for assessment and feedback on creative learning. We follow with the need and value of studying practices and how engaging stakeholders in co-design facilitates the process. We then focus on the need for making AI transparent, including the foundational role that the situating analytics paradigm can play in future research. Lastly, we advocate aligning the role of analytics with instructors' approach of scaffolding but not constraining creativity.

5.1 Investigate How AI Can Be Useful for Supporting Complex Human Activity

As Rikakis points out, the usefulness of AI is being realized in contexts ranging from image search to purchase recommendation to manufacturing. However, there is a need for addressing AI's slow progress in education contexts, which involve "rich experiences aimed at advancing human intelligence" [186]. The value of addressing the need is visible from the U.S. National Science Foundation's (NSF) designation of 'AI-Augmented Learning' as one of eight key areas for fundamental research in AI, for advancing national competitiveness and transforming society [263].

The present research focused on the use of AI for addressing the growing demands for design education. Aligned with Rikakis' observation, our survey in Chapter 2 provides evidence that there is a scarcity of computational approaches, including AI, for assessing and providing feedback on students' design work. Because design education is so based in contextualized human experiences, ideas for supporting assessment and feedback using AI may appear like pie-in-the-sky. Realizing such ideas would seem akin to a magician pulling a rabbit out of a hat.

In Chapter 3, building on the fundamental work of Suchman [12] and Dourish [13], we showed that understanding situated practices is vital to developing AI support for sociotechnical design education contexts. We used co-design methods to understand situated practices. Based on the understandings, we developed approaches for deriving AI-based analytics for measuring a range of design characteristics, i.e., why and how a rabbit can be pulled out of a hat. In Chapter 4, we pulled one rabbit out of a hat, developing AI-based analytics that measure multiscale design organization and studying how they affect instructors' assessment and feedback, in situated course contexts. We discussed how these analytics correspond to higher-order learning processes. We discussed how insights into these processes can prove vital in improving learners' creative strategies and abilities.

Our investigation holds implications for developing AI support for design education and industry contexts, as well as other contexts that involve creative learning. While we pulled out one rabbit, there is a need for pulling many. Based on the understandings developed using creative cognition's family resemblance principle [33], we advocate suites of AI-based analytics: together they tend to indicate good design work. Future research can develop AI support for forms of measuring and presenting Fluency, Flexibility, Visual Consistency, Multiscale Organization, and Legible Contrast, for which we presented approaches in Chapter 3. Likewise, develop AI support for measuring complex characteristics in a range of creative contexts, including arts, writing, and computer programming.

Improving people's creative abilities is vital, as it can stimulate innovation and thus economic growth. Further, in Chapter 3, we note that creative tasks are the least susceptible to being eliminated by automation. Hence, the incorporation of AI for supporting these tasks has the potential to become a mainstay of human involvement and performance in the future of work.

5.2 Understand Situated Practices Through Co-Design

Design education inducts students into a 'community of practice' [27], where students learn and develop skills, often through socially situated understandings of work that practitioners perform [28]. Developing technology support for socially situated practices is fundamentally challenging, as it requires bridging intellectually different domains [13]. To address the challenges, it becomes vital to identify contextual properties by studying situated practices [13].

Future research investigating AI support for rich human experiences will benefit from taking co-design [75, 58] approaches for studying practices. In the present research, co-design approaches facilitated studying practices through:

- *Involving Stakeholders*: We engaged co-design approaches—particularly, ongoing discourses through individual discussions and workshops—to identify contextual properties pertaining to how instructors assess and provide feedback on student work, in practice (Chapter 3). Using co-design approaches allowed us to not only elicit tacit forms of knowledge but involve stakeholders at a deep level. Instructors co-created shared meanings, discussing problems in assessment and feedback and potential solutions to address them. Finding value in research, two instructors in our study chose to participate further, by writing papers and grant proposals together. Through co-design approaches, "stakes" continued to increase, which facilitated iteratively developing better understandings of practices and needs.

- *Field-testing Technology*: We reaped benefits by using the co-design method of technology probes [17] to understand how analytics affect instructors' situated assessment and feedback practices. We deployed a technology probe—comprised by AI-based analytics and a dashboard integrated with a multiscale design environment—in situated course contexts (Chapter 4). It allowed instructors to experience the technology in a real-world setting. It allowed us to collect data from field-testing. Based on their experiences, instructors provided us with rich feedback regarding the significance of analytics, their potential as rubric elements, and suggestions for explaining the analytics measures. This way, the probe method allows continuous refinement of technology to make it suited for users' situated practice.

5.3 Explaining Analytics: A Key to Intelligent Support for Design Education

Much research has addressed *explainable AI*, which focuses on conveying, to users, the algorithms that recognizers invoke. An emergent finding of this research is that what matters to various users, alternatively, is explaining AI analytics, that is, conveying what measures mean, so that they derive value from them, in situated contexts of practice.

The need here is to explain the what the capabilities of an AI mean, in practice, as part of user experiences. We argue that this matters more than, or at least complements, explaining the algorithms that make up the underlying mechanism. As Doran et al. describe, explaining the mechanism involves showing users how inputs are mathematically mapped to outputs [237]. This necessitates a level of technical understanding by users. On the other hand, explaining what the capabilities of AI means involves providing users with means through which they can understand how an AI makes high-level inferences, based on certain lower-level inferences. According to Doran et al., it corresponds to aspects of comprehensibility, e.g., AI making a high-level inference that a given image is of a factory, because lower-level inferences include halogen lights, floor, and machines. In the present case, of design education, explaining a capability involves conveying what analytics mean in the context of the design work whose characteristics they measure. High-level inferences involve how the present design analytics are based on lower-level conceptual (e.g., ideas) and visual (e.g., clusters) inferences from design work.

Transparency of AI-based systems is vital, to convey an AI's role in supporting situated practice and establish its accountability to users [12]. According to Bellotti and Edwards' landmark findings, an intelligible context-aware system should "provide users with an account of its actions, interpretations, and decisions" [159]. Subsequently, Shneiderman shows why user interface design that supports human comprehensibility is critical to developing reliable AI systems [158]. Without transparency, questions and concerns regarding the accountability and reliability of AI would only grow. For example, what if an AI-based clinical decision support system makes an incorrect diagnosis? Or what if a self-driving car misreads the speed limit? Or what if an AI system always filters out women or people of color job candidates? Unless the AI explains its decisions, placing trust in "black box" operations can result in unfair, dangerous, or even catastrophic outcomes.

We argue that the need for explaining AI outcomes is likewise important to address in education. What if the AI makes an incorrect determination of a student's performance? It could affect their career and well-being. Moreover, even if AI makes a correct determination, it would be of limited use without explaining outcomes. For example, AI showing to students that their 'Visual

Consistency' score is 80 out of 100—without explaining where they earned 80 points and where they lost 20—would leave it to the students to figure out what could be right or wrong. This may cause them stress and anxiety. It is not clearly actionable.

Instructors in our study highlighted the importance of explaining outcomes on assignments to students. Explaining the basis of assessment is already a substantial challenge that human graders deal with in situated practice. D11's experience exemplifies: *"[Some of the students] violate the principles that I teach them, and then I say hey, the principles, we talked about that. And they're like oh, you mean [that]".* Instructors recommend an analogous role for AI-based assessment. According to them, explaining analytics based on learning objectives is vital toward bridging the gap between instructors' and students' understandings. Recalling D5's statement from Chapter 3 (Section 3.4.1), *"There's a kind of disconnect between [students] turning in a [design] and they getting a number [back from AI]...Why is it a 'B'?".* Without addressing this need, AI would not be able to help students improve their work. Instructors would not be able to relate AI outcomes to student work.

In Chapter 3 (Section 3.4.1), we introduced *situating analytics*, a paradigm for conveying the meaning of measures that align with design rubrics—by contextually integrating their presentation with what and how they are measuring—here, associated design work. Building on the ideas, in Chapter 4 (Section 4.2.3), we integrated a dashboard with its corresponding design environment, thereby making connections visible between the analytics and particular sets of design element assemblages that they measure. We found that our approach explains the AI-based multiscale design analytics to instructors. They were able to use our multiscale analytics to garner a sense of students' design organization, and so pinpoint where the AI-based analytics did not match their interpretation. We thus expect the situating analytics paradigm to play a foundational role in future research, for explaining a range of analytics. For Fluency, Flexibility, Visual Consistency, Multiscale Organization, and Legible Contrast, future research can build on the approaches we presented in Chapter 3.

In Chapter 4 (Section 4.5.1), we discussed how explainable AI (XAI) [237]—which focuses on

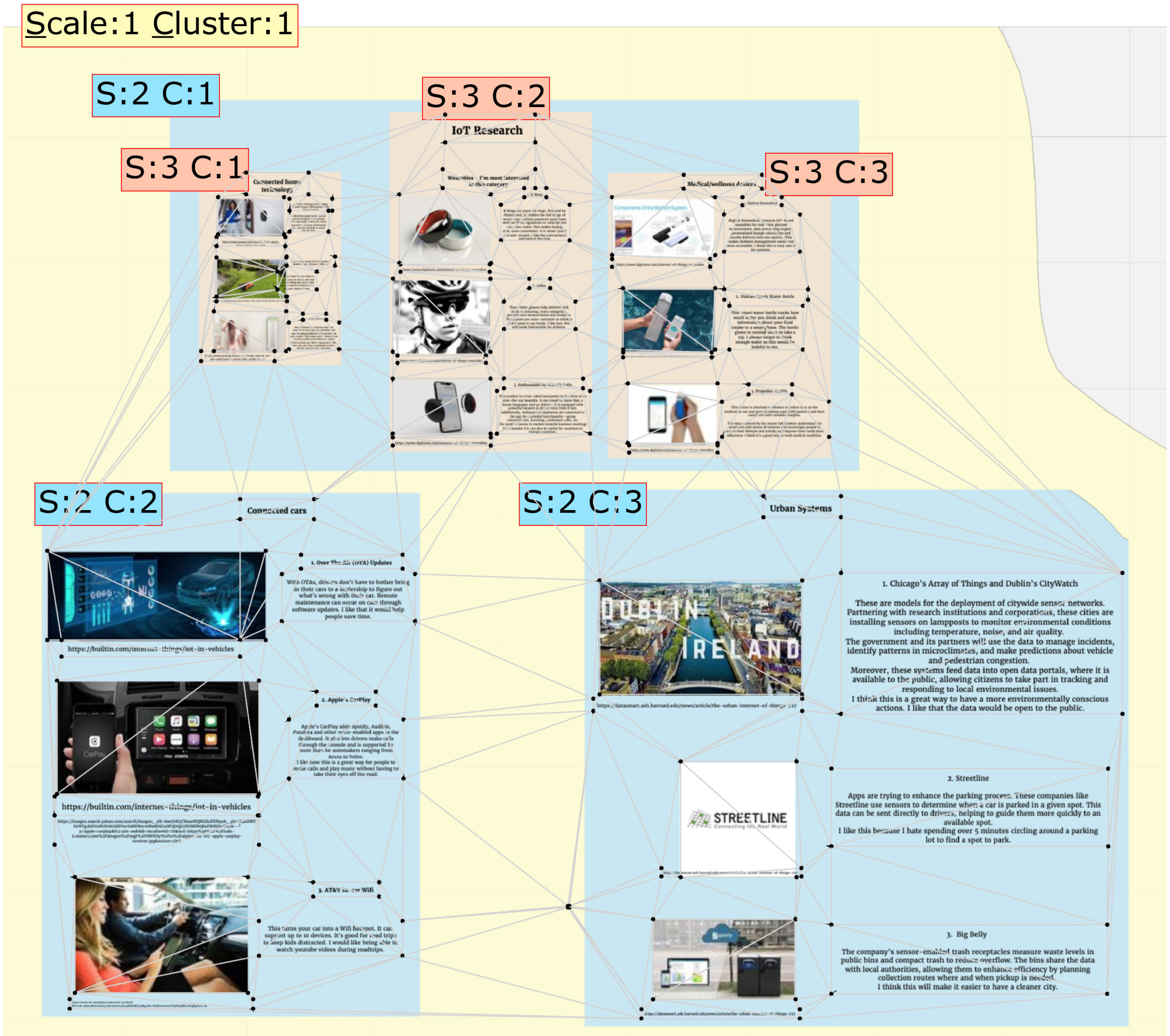

Figure 5.1: A mockup showing how explainable AI (XAI) approaches can further support conveying the basis of analytics. We overlay the Delaunay graph—used to determine nested clusters based on spatial distances—on the top of Figure 4.4's scale and cluster renderings. Making the Delaunay graph visible supports users in seeing the lengths of edges that the AI uses to determine whether a cluster is nested.

the human interpretability and comprehensibility of how an AI algorithm functions—can further support explaining AI-based analytics. For example, for multiscale design analytics, presenting the Delaunay graph can enable users to see the lengths of edges that the AI used to determine whether a cluster is nested. Figure 5.1 shows a mockup we created by overlaying the Delaunay graph on the Figure 4.4's presentation of nested clusters. Likewise, beneficial future research can use XAI techniques to augment the approaches we developed for explaining other analytics.

In Chapter 4 (Section 4.5.5), we also presented initial evidence for techniques—animation, indexical representation, and exemplification—that can improve explaining AI-based analytics. Future research can investigate these techniques in diverse contexts, so as to determine the extent to which they support explaining various analytics.

5.4 Make AI Interactive and Controllable

According to Bellotti and Edwards, providing users with control of a computational system is essential, as the system cannot deterministically predict the desired course of action, due to too many contextual variables involved [159]. To give users this control of AI, Shneiderman emphasizes the need for carefully designing human-AI interaction [158]. Controllability can help AI adapt to users' practices, as well as prevent harmful or fatal situations. For example, in language translation, AI is able to adapt based on the user edits and retranslate similar occurrences [264]. An example of the fatal behavior is the use of AI for excessive automation in the Boeing MAX 737 MCAS system. Analyzing the MCAS failure, among the lessons learned, Mongan and Kohli include: "AI systems...should have a simple, fast, and lasting mechanism for override" [265]. Woodruff explains the need for controllability in terms of 'contestability' and 'recourse', to allow users to challenge and change AI outcomes, through actionable inputs [194].

Our investigation highlights the need for controllability in designing AI support for education contexts. As we discussed in Chapter 3 (Section 3.4.3), AI approaches are prone to systematic errors in training data or models. Thus, derived analytics may lack accuracy. Instructors in our study expressed the need for indicating whether AI is performing as expected. Recalling D6's statement, *"spit out the rubric scores and then the professor can say, well that's right or wrong"*. Relatedly, in Chapter 4 (Section 4.4.3), instructors reported multiscale design analytics not matching their interpretation at times.

The situating analytics paradigm focuses on designing user interaction to gather feedback on the validity and utility of AI outcomes. We developed ideas for providing affordances to instructors and students, for controllability, through the dashboard integrated with the design environment (Chapter 3, Section 3.4.3 and Figure 3.3). Through these affordances, users can indicate when their interpretation does not match with the AI assessment. The affordances allow the users to indicate mismatches for both high-level (e.g., visual consistency analytic) and low-level (e.g., clusters) inferences. As we discussed in Chapters 3 and 4, providing instructors means to override AI will not only correct the outcomes, but further can be expected to provide a continuous source of new labeled data to the designers of AI, which, in turn, can help them iteratively bring the artificial closer to human intelligence. In Section 3.4.3, we argued that the processes of frequent feedback in design classrooms have the potential to be functionally rendered as isomorphic with AI algorithms' iterative needs for labeled examples for training recognizers.

Part of controllability thus involves giving users the ability to "label" examples in which the feedback from the AI's model runs counter to their own mental models. In this regard, we note that users' mental models do not remain fixed. Instructors in our study, as well as prior work [29], described that design rubrics evolve based on the direction that a project takes. The evolution of design rubrics has the potential to be functionally rendered as isomorphic with the resulting evolution in the needs for labeled data. Beneficial future research can develop controllable AI interfaces and investigate how they affect instructor and student experiences, as well as how AI can evolve based on new labeled examples, in order to support situated course contexts.

5.5 Scaffold Creativity Without Constraining

Pedagogical processes that lead students to creative outcomes are central in design education [20]. Deliverables, rubrics, and criteria are amenable to change, so as to support students in emergent creative directions [29]. At the same time, instructors need to scaffold creative processes—e.g., prescribing activities while encouraging design reasoning, positioning themselves as a source of expertise, and facilitating further development of interesting aspects of the project—so as to help students make progress in their creative work [142]. Hence, instructors need to walk a fine line between encouraging creativity and providing structure.

We find that AI-based analytics have the potential to provide analogous support to instructors' walking a fine line, where they scaffold students' creative processes but not constrain them. In Chapter 3 (Section 3.4.2), we interpreted our findings using the creative cognition's family resemblance principle [33], to identify the same incomplete, yet useful role for analytics, as design rubrics: no particular analytics are necessary or sufficient; together they tend to indicate creative work. As D7 said, *"There is no right answer"*. As D5 expressed, *"some students could just do something really simple and it's just like,'Uh yes, that's it'. And somebody else could have something really complex and you're like, 'Yes, that's it"'*.

Chapter 4 echoes this, in the study of instructors' experience with multiscale design analytics (Section 4.4.3). According to instructors, analytics can help students reflect and become conscious of their design processes. For example, I1 suggested showing scale and cluster analytics to students so that they can explore zoom levels and become aware of how they separate. I4 suggested providing students with examples of multiscale organization, so as to scaffold their processes. At the same time, the goal is not to constrain creativity. As I4 expressed, *"I wouldn't want...all [designs] to look the same like you don't want to go somewhere and see every painting looks the same"*. Likewise, I1 does not want to control how students organize elements in clusters.

Hence, we do not advocate setting specific analytics values as goals, unless instructors require students to meet certain quantitative performance, as measured by the analytics. Rather, analytics are expected to serve as grist for reflection and stimulate ongoing work. Like design rubrics, analytics cannot be perfect. They do not need to be perfect in order to provide value. Their purpose is to augment, not replace, instructors' ongoing interpretation and engagement.

5.6 Conclusion: Situating Analytics for Design Education

We investigated the potential of AI for supporting situated design education assessment and feedback practices. We took an interdisciplinary approach, drawing and connecting ideas from human-computer interaction, design education, co-design, creative cognition, and multiple areas of computer science. We sustained this interdisciplinary approach through and across phases of research: conceptualizing, defining methods, designing, developing, interpreting, and theory-building.

To understand practices and needs, we engaged instructors—from Landscape Architecture & Urban Planning, Architecture, Interactive Art & Design, Mechanical Engineering, and Computer Science and Engineering—in co-design. We closely involved them in different phases of research through individual discussions, workshops, and technology probes methods. We developed understandings of how instructors perform assessment, including approaches they use for diverse design characteristics, as well as challenges they face in providing and tracking students' incorporation of feedback. By interpreting our findings using creative cognition's family resemblance principle, we identified an opportunity for AI-based analytics to play the same incomplete, yet useful, role as design rubrics. Drawing on computational techniques—e.g., natural language processing, computer vision, and machine learning—we developed approaches for measuring diverse design characteristics. We employed human-computer interaction principles for conveying the meaning of measures and effectively integrating the computational support into instructors' and students' experiences.

Based on understandings of practices, and drawing on prior research by Suchman and Dourish, we show that making AI-based systems transparent is vital for supporting design education experiences. We contribute situating analytics as a paradigm for conveying the meaning of measures, which align with design rubrics, by contextually integrating the presentation of the measures with associated design work. Following this paradigm, we developed approaches for deriving and presenting AI-based design creativity analytics—Fluency, Flexibility, Visual Consistency, Multiscale Organization, and Legible Contrast—which measure complex characteristics in students' design work. Building on the ideas, we realized AI-based multiscale design organization analytics. We introduced them in situated course contexts, via a dashboard integrated with an actual design environment. We found that linking the analytics with design element assemblages they measure helped instructors in understanding what analytics mean. Based on instructors' experiences, we derived implications for conveying the meaning of complex measures, including techniques such as animation, indexical representation, and exemplification, as well as representations that take into account the ambiguity in the interpretation of design work and uncertainty associated with

AI-based approaches.

The present research has implications for developing AI-based approaches to support rich human experiences, as situated in creative contexts of education and industry. Developing a range of AI-based analytics that measure complex characteristics in creative human enterprises can transform learning by doing, through instantaneous, on demand feedback. Understanding situated practice will remain vital to developing transparent AI support, so as to establish its intelligibility and accountability to users. While analytics have the potential to stimulate ongoing creative work, they could result in people's increased attention to those aspects of their work for which automated assessment and feedback become available. How does this affect their design creativity as a whole? Likewise, it will be important to continue identifying further significant characteristics of design, which may be general or context specific, and how AI can measure them. On the whole, for design education, continuing investigation can beneficially address how AI-based situated analytics affect the education experiences and outcomes of instructors and students, how to keep improving interaction that explains AI capabilities, and how combinations of design analytics contribute. More broadly, we need to understand how these principles can benefit other fields, or turn out to be domain specific.

REFERENCES

[1] Noun Project: Free Icons for Everything, "Dashboard by Rafael Garcia Motta from the Noun Project; dashboard by Aybige from the Noun Project; students by Philipp Petzka from the Noun Project; and workshop by DailyPM from the Noun Project.." https://thenounproject.com/, 2019. Last accessed: 2019-11-26.

[2] Adobe, "Free Poster Template with images from https://www.nasa.gov/ and text based on thrillist.com/tech/11-nasa-technologies-we-use-every-day | Adobe Spark." https://spark.adobe.com/post/1TK56DIfH24Xz/, 2020. Last accessed: 2020-10-19.

[3] V. Estivill-Castro and I. Lee, "Multi-level clustering and its visualization for exploratory spatial analysis," *GeoInformatica*, vol. 6, no. 2, pp. 123–152, 2002.

[4] C. Wrigley and K. Straker, "Design thinking pedagogy: The educational design ladder," *Innovations in Education and Teaching International*, vol. 54, no. 4, pp. 374–385, 2017.

[5] M. Krause, T. Garncarz, J. Song, E. M. Gerber, B. P. Bailey, and S. P. Dow, "Critique style guide: Improving crowdsourced design feedback with a natural language model," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 4627–4639, 2017.

[6] J. K. McDonald, P. J. Rich, and N. B. Gubler, "The perceived value of informal, peer critique in the instructional design studio," *TechTrends*, vol. 63, no. 2, pp. 149–159, 2019.

[7] K. Papadopoulos, L. Sritanyaratana, and S. R. Klemmer, "Community TAs scale high-touch learning, provide student-staff brokering, and build esprit de corps," in *Proceedings of the first ACM conference on Learning@ scale conference*, pp. 163–164, ACM, 2014.

[8] S. Dow, E. Gerber, and A. Wong, "A pilot study of using crowds in the classroom," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 227–236, ACM, 2013.

[9] J. S. Hui, E. M. Gerber, and S. P. Dow, "Crowd-based design activities: helping students connect with users online," in *Proceedings of the 2014 conference on Designing interactive systems*, pp. 875–884, ACM, 2014.

[10] A. Yuan, K. Luther, M. Krause, S. I. Vennix, S. P. Dow, and B. Hartmann, "Almost an expert: The effects of rubrics and expertise on perceived value of crowdsourced design critiques," in *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, pp. 1005–1017, ACM, 2016.

[11] A. Kittur, L. Yu, T. Hope, J. Chan, H. Lifshitz-Assaf, K. Gilon, F. Ng, R. E. Kraut, and D. Shahaf, "Scaling up analogical innovation with crowds and ai," *Proceedings of the National Academy of Sciences*, vol. 116, no. 6, pp. 1870–1877, 2019.

[12] L. A. Suchman, *Plans and situated actions: The problem of human-machine communication*. Cambridge university press, 1987.

[13] P. Dourish, "What we talk about when we talk about context," *Personal and ubiquitous computing*, vol. 8, no. 1, pp. 19–30, 2004.

[14] M.-L. How, S.-M. Cheah, Y.-J. Chan, A. C. Khor, and E. M. P. Say, "Artificial intelligence-enhanced decision support for informing global sustainable development: A human-centric ai-thinking approach," *Information*, vol. 11, no. 1, p. 39, 2020.

[15] H. Lim, "Design for computer-mediated multilingual communication with ai support," in *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pp. 93–96, 2018.

[16] N. Lupfer, H. Fowler, A. Valdez, A. Webb, J. Merrill, G. Newman, and A. Kerne, "Multi-scale Design Strategies in a Landscape Architecture Classroom," in *Proceedings of the 2018 on Designing Interactive Systems Conference 2018*, pp. 1081–1093, ACM, 2018.

[17] H. Hutchinson, W. Mackay, B. Westerlund, B. B. Bederson, A. Druin, C. Plaisant, M. Beaudouin-Lafon, S. Conversy, H. Evans, H. Hansen, and Others, "Technology probes:

inspiring design for and with families," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 17–24, ACM, 2003.

[18] K. Charmaz, *Constructing grounded theory*. Sage, 2014.

[19] G. A. Bowen, "Grounded Theory and Sensitizing Concepts," *International Journal of Qualitative Methods*, vol. 5, no. 3, pp. 12–23, 2006.

[20] R. K. Sawyer, "Teaching and learning how to create in schools of art and design," *Journal of the Learning Sciences*, vol. 27, no. 1, pp. 137–181, 2018.

[21] C. L. Dym, A. M. Agogino, O. Eris, D. D. Frey, and L. J. Leifer, "Engineering Design Thinking, Teaching, and Learning," *Journal of Engineering Education*, vol. 94, pp. 103–120, Jan 2005.

[22] C. B. Brandt, K. Cennamo, S. Douglas, M. Vernon, M. McGrath, and Y. Reimer, "A theoretical framework for the studio as a learning environment," *International Journal of Technology and Design Education*, vol. 23, no. 2, pp. 329–348, 2013.

[23] D. P. Crismond and R. S. Adams, "The informed design teaching and learning matrix," *Journal of Engineering Education*, vol. 101, no. 4, pp. 738–797, 2012.

[24] D. Dannels, A. H. Gaffney, and K. N. Martin, "Beyond content, deeper than delivery: What critique feedback reveals about communication expectations in design education," *International Journal for the Scholarship of teaching and Learning*, vol. 2, no. 2, p. 12, 2008.

[25] Y. Oh, S. Ishizaki, M. D. Gross, and E. Y.-L. Do, "A theoretical framework of design critiquing in architecture studios," *Design Studies*, vol. 34, no. 3, pp. 302–325, 2013.

[26] D. Vyas, G. van der Veer, and A. Nijholt, "Creative practices in the design studio culture: collaboration and communication," *Cognition, Technology & Work*, vol. 15, no. 4, pp. 415–443, 2013.

[27] J. Lave and E. Wenger, *Situated learning: Legitimate peripheral participation*. Cambridge university press, 1991.

[28] A. Shreeve, E. Sims, and P. Trowler, "'A kind of exchange': learning from art and design teaching," *Higher Education Research & Development*, vol. 29, no. 2, pp. 125–138, 2010.

[29] R. K. Sawyer, "Teaching creativity in art and design studio classes: A systematic literature review," *Educational research review*, vol. 22, pp. 99–113, 2017.

[30] J. Osmond and M. Tovey, "The Threshold of Uncertainty in Teaching Design.," *Design and Technology Education*, vol. 20, no. 2, pp. 50–57, 2015.

[31] J. Frich, M. Mose Biskjaer, and P. Dalsgaard, "Twenty Years of Creativity Research in Human-Computer Interaction: Current State and Future Directions," in *Proceedings of the 2018 Designing Interactive Systems Conference*, pp. 1235–1257, ACM, 2018.

[32] J. C. Kaufman and R. A. Beghetto, "Beyond big and little: The four c model of creativity.," *Review of general psychology*, vol. 13, no. 1, p. 1, 2009.

[33] S. M. Smith, T. B. Ward, and R. A. Finke, *The creative cognition approach*. MIT press, 1995.

[34] J. J. Shah, S. M. Smith, and N. Vargas-Hernandez, "Metrics for measuring ideation effectiveness," *Design studies*, vol. 24, no. 2, pp. 111–134, 2003.

[35] A. Kerne, A. M. Webb, S. M. Smith, R. Linder, N. Lupfer, Y. Qu, J. Moeller, and S. Damaraju, "Using metrics of curation to evaluate information-based ideation," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 21, no. 3, pp. 1–48, 2014.

[36] J. S. Gero and M. L. Maher, "Mutation and analogy to support creativity in computer-aided design," *CAAD Futures'91*, pp. 241–249, 1991.

[37] T. Kurtoglu, M. I. Campbell, and J. S. Linsey, "An experimental study on the effects of a computational design tool on concept generation," *Design Studies*, vol. 30, no. 6, pp. 676–703, 2009.

[38] J. Koch, A. Lucero, L. Hegemann, and A. Oulasvirta, “May ai? design ideation with cooperative contextual bandits,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2019.

[39] R. Kumar, J. O. Talton, S. Ahmad, and S. R. Klemmer, “Bricolage: example-based retargeting for web design,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2197–2206, 2011.

[40] Z. Wang, J. Yang, H. Jin, E. Shechtman, A. Agarwala, J. Brandt, and T. S. Huang, “Deepfont: Identify your font from an image,” in *Proceedings of the 23rd ACM international conference on Multimedia*, pp. 451–459, ACM, 2015.

[41] K. Reinecke, T. Yeh, L. Miratrix, R. Mardiko, Y. Zhao, J. Liu, and K. Z. Gajos, “Predicting users’ first impressions of website aesthetics with a quantification of perceived visual complexity and colorfulness,” in *Proc. CHI*, pp. 2049–2058, ACM, 2013.

[42] A. Oulasvirta, S. De Pascale, J. Koch, T. Langerak, J. Jokinen, K. Todi, M. Laine, M. Kristhombuge, Y. Zhu, A. Miniukovich, and Others, “Aalto Interface Metrics (AIM) A Service and Codebase for Computational GUI Evaluation,” in *The 31st Annual ACM Symposium on User Interface Software and Technology Adjunct Proceedings*, pp. 16–19, 2018.

[43] M. Mackeprang, A. Khiat, and C. Müller-Birn, “Concept Validation during Collaborative Ideation and Its Effect on Ideation Outcome,” in *CHI Extended Abstracts*, pp. 1–6, ACM, 2018.

[44] T. J. Ngoon, C. A. Fraser, A. S. Weingarten, M. Dontcheva, and S. Klemmer, “Interactive guidance techniques for improving creative feedback,” in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–11, 2018.

[45] A. Pardo, J. Jovanovic, S. Dawson, D. Gašević, and N. Mirriahi, “Using learning analytics to scale the provision of personalised feedback,” *British Journal of Educational Technology*, vol. 50, no. 1, pp. 128–138, 2019.

[46] L.-A. Lim, S. Gentili, A. Pardo, V. Kovanović, A. Whitelock-Wainwright, D. Gašević, and S. Dawson, “What changes, and for whom? a study of the impact of learning analytics-based process feedback in a large course,” *Learning and Instruction*, p. 101202, 2019.

[47] L. Lockyer, E. Heathcote, and S. Dawson, “Informing pedagogical action: Aligning learning analytics with learning design,” *American Behavioral Scientist*, vol. 57, no. 10, pp. 1439–1459, 2013.

[48] S. B. Shum and R. D. Crick, “Learning dispositions and transferable competencies: pedagogy, modelling and learning analytics,” in *Proceedings of the 2nd international conference on learning analytics and knowledge*, pp. 92–101, 2012.

[49] A. F. Wise, “Designing pedagogical interventions to support student use of learning analytics,” in *Proceedings of the fourth international conference on learning analytics and knowledge*, pp. 203–211, 2014.

[50] K. Verbert, E. Duval, J. Klerkx, S. Govaerts, and J. L. Santos, “Learning analytics dashboard applications,” *American Behavioral Scientist*, vol. 57, no. 10, pp. 1500–1509, 2013.

[51] A. Shibani, S. Knight, and S. B. Shum, “Educator perspectives on learning analytics in classroom practice,” *The Internet and Higher Education*, vol. 46, p. 100730, 2020.

[52] H. Tarmazdi, R. Vivian, C. Szabo, K. Falkner, and N. Falkner, “Using learning analytics to visualise computer science teamwork,” in *Proceedings of the 2015 ACM Conference on Innovation and technology in computer science education*, pp. 165–170, 2015.

[53] S. Dawson, L. Macfadyen, F. R. Evan, T. Foulsham, and A. Kingstone, “Using technology to encourage self-directed learning: The Collaborative Lecture Annotation System (CLAS),” in *Australasian Society for Computers in Learning in Tertiatry Education*, pp. 246–255, 2012.

[54] E. Duval, “Attention please!: learning analytics for visualization and recommendation,” in *Proceedings of the 1st international conference on learning analytics and knowledge*, pp. 9–17, ACM, 2011.

[55] K. Verbert, S. Govaerts, E. Duval, J. L. Santos, F. Assche, G. Parra, and J. Klerkx, "Learning dashboards: an overview and future research opportunities," *Personal and Ubiquitous Computing*, vol. 18, no. 6, pp. 1499–1514, 2014.

[56] P. Blikstein, "Using learning analytics to assess students' behavior in open-ended programming tasks," in *Proceedings of the 1st international conference on learning analytics and knowledge*, pp. 110–116, ACM, 2011.

[57] G. Britain, A. Jain, N. Lupfer, A. Kerne, A. Perrine, J. Seo, and A. Sungkajun, "Design is (A)live: An Environment Integrating Ideation and Assessment," in *CHI Late-Breaking Work*, pp. 1–8, ACM, 2020.

[58] E. B.-N. Sanders and P. J. Stappers, "Co-creation and the new landscapes of design," *Co-design*, vol. 4, no. 1, pp. 5–18, 2008.

[59] J. S. Dumas and J. Redish, *A practical guide to usability testing*. Intellect books, 1999.

[60] G. Salvendy, *Handbook of human factors and ergonomics*. John Wiley & Sons, 2012.

[61] M. Buchenau and J. F. Suri, "Experience prototyping," in *Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques*, pp. 424–433, 2000.

[62] F. S. Visser, P. J. Stappers, R. Van der Lugt, and E. B. Sanders, "Contextmapping: experiences from practice," *CoDesign*, vol. 1, no. 2, pp. 119–149, 2005.

[63] J. Gregory, "Scandinavian approaches to participatory design," *International Journal of Engineering Education*, vol. 19, no. 1, pp. 62–74, 2003.

[64] B. Gaver, T. Dunne, and E. Pacenti, "Design: cultural probes," *interactions*, vol. 6, no. 1, pp. 21–29, 1999.

[65] T. Mattelmäki *et al.*, *Design probes*. Aalto University, 2006.

[66] S. Bødker, K. Grønbæk, and M. Kyng, “Cooperative design: techniques and experiences from the Scandinavian scene,” in *Readings in Human–Computer Interaction*, pp. 215–224, Elsevier, 1995.

[67] M. J. Muller, “Participatory design: the third space in HCI,” *Human-computer interaction: Development process*, vol. 4235, pp. 165–185, 2003.

[68] C. Alexander, *Notes on the Synthesis of Form*, vol. 5. Harvard University Press, 1964.

[69] E. Barba, “Cognitive Point of View in Recursive Design,” *She Ji: The Journal of Design, Economics, and Innovation*, vol. 5, no. 2, pp. 147–162, 2019.

[70] N. Lupfer, A. Kerne, R. Linder, H. Fowler, V. Rajanna, M. Carrasco, and A. Valdez, “Multiscale Design Curation: Supporting Computer Science Students’ Iterative and Reflective Creative Processes,” in *Proceedings of the 2019 on Creativity and Cognition*, pp. 233–245, ACM, 2019.

[71] H. A. Simon, “How big is a chunk?: By combining data from several experiments, a basic human memory unit can be identified and measured,” *Science*, vol. 183, no. 4124, pp. 482–488, 1974.

[72] A. Kerne, N. Lupfer, R. Linder, Y. Qu, A. Valdez, A. Jain, K. Keith, M. Carrasco, J. Vanegas, and A. Billingsley, “Strategies of free-form web curation: Processes of creative engagement with prior work,” in *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition*, pp. 380–392, 2017.

[73] A. Kerne, “Interface ecosystem, the fundamental unit of information age ecology,” in *Proceedings of the 29th International Conference on Computer Graphics and Interactive Techniques. Electronic Art and Animation Catalog.*, pp. 142–145, 2002.

[74] B. De La Harpe, J. F. Peterson, N. Frankham, R. Zehner, D. Neale, E. Musgrave, and R. McDermott, “Assessment focus in studio: What is most prominent in architecture, art and design?,” *International Journal of Art & Design Education*, vol. 28, no. 1, pp. 37–51, 2009.

[75] E. B.-N. Sanders and P. J. Stappers, *Convivial toolbox: Generative research for the front end of design*. BIS Amsterdam, 2012.

[76] D. Schuler and A. Namioka, *Participatory design: Principles and practices*. CRC Press, 1993.

[77] L. Suchman, "Making work visible," *Communications of the ACM*, vol. 38, no. 9, pp. 56–64, 1995.

[78] P. Biernacki and D. Waldorf, "Snowball sampling: Problems and techniques of chain referral sampling," *Sociological methods & research*, vol. 10, no. 2, pp. 141–163, 1981.

[79] S. Yilmaz and S. R. Daly, "Feedback in concept development: Comparing design disciplines," *Design Studies*, vol. 45, pp. 137–158, 2016.

[80] H. De Wit, "Internationalisation of higher education in europe and its assessment, trends and issues," *NVAO The Hague, The Netherlands*, 2010.

[81] P. A. Board, "Accreditation standards," *Planning Accreditation Board, Chicago, IL*, 2017.

[82] J. Kolko, "Transformative learning in the design studio.," *interactions*, vol. 19, no. 6, pp. 82–83, 2012.

[83] J.-F. Gehanno, L. Rollin, and S. Darmoni, "Is the coverage of google scholar enough to be used alone for systematic reviews," *BMC medical informatics and decision making*, vol. 13, no. 1, p. 7, 2013.

[84] K. Charmaz, *Constructing grounded theory: A practical guide through qualitative analysis*. sage, 2006.

[85] S. M. Watkins, "Using the design process to teach functional apparel design," *Clothing and Textiles Research Journal*, vol. 7, no. 1, pp. 10–14, 1988.

[86] M. Barrett, "Guidelines for evaluation and assessment in art and design education 5-18 years.," *Journal of Art and Design Education*, vol. 9, no. 3, pp. 299–313, 1990.

[87] S. Sheppard, R. Jenison, A. Agogino, M. Brereton, L. Bocciarelli, J. Dally, J. Demel, C. Dym, D. Evans, and R. Faste, "Examples of freshman design education," *International Journal of Engineering Education*, vol. 13, no. 4, pp. 248–261, 1997.

[88] R. S. Adams and C. J. Atman, "Cognitive processes in iterative design behavior," in *FIE'99 Frontiers in Education. 29th Annual Frontiers in Education Conference. Designing the Future of Science and Engineering Education. Conference Proceedings (IEEE Cat. No. 99CH37011*, vol. 1, pp. 11A6–13, IEEE, 1999.

[89] T. Jung, E. Y.-L. Do, and M. D. Gross, "Immersive redlining and annotation of 3D design models on the web," in *Computers in Building*, pp. 81–98, Springer, 1999.

[90] Z. Ulusoy, "To design versus to understand design: the role of graphic representations and verbal expressions," *Design Studies*, vol. 20, no. 2, pp. 123–130, 1999.

[91] P. L. Brusasco, L. Caneparo, G. Carrara, A. Fioravanti, G. Novembri, and A. M. Zorgno, "Computer supported design studio," *Automation in Construction*, vol. 9, no. 4, pp. 393–408, 2000.

[92] J. K. Ochsner, "Behind the mask: a psychoanalytic perspective on interaction in the design studio," *Journal of Architectural Education*, vol. 53, no. 4, pp. 194–206, 2000.

[93] M. E. Cardella, C. J. Atman, R. S. Adams, and J. Turns, "Engineering student design processes: Looking at evaluation practices across problems," in *Proceedings of the 2002 American Society for Engineering Education Annual Conference and Exposition, Montreal, Canada*, 2002.

[94] W. P. Lewis and E. Bonollo, "An analysis of professional skills in design: implications for education and research," *Design studies*, vol. 23, no. 4, pp. 385–406, 2002.

[95] Y. J. Reimer and S. A. Douglas, "Teaching HCI design with the studio approach," *Computer science education*, vol. 13, no. 3, pp. 191–205, 2003.

[96] J. Zimmerman, "Position paper on design in hci education," in *International Conference on Human-Computer Interaction (Interact): Workshop of HCI Education*, Carnegie Mellon University, 2003.

[97] E. Blevis, Y. Rogers, M. Siegel, W. Hazlewood, and A. Stephano, "Integrating hci and design: Hci/d at iub, a design education case story," in *Zimmerman, J., Evenson, S., Baumann, K., & Purgathofer, P. Workshop on the relationship between design and HCI. ACM CHI 2004 conference on Human factors and computing systems. Vienna, Austria. http://k2. iguw. tuwien. ac. at*, vol. 8080, 2004.

[98] M. E. Cardella, C. J. Atman, and R. S. Adams, "Mapping between design activities and external representations for engineering student designers," *Design Studies*, vol. 27, no. 1, pp. 5–24, 2006.

[99] P. D. Adamczyk and M. B. Twidale, "Supporting multidisciplinary collaboration: requirements from novel hci education," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 1073–1076, 2007.

[100] K. K. Jeffries, "Diagnosing the creativity of designers: individual feedback within mass higher education," *Design Studies*, vol. 28, no. 5, pp. 485–497, 2007.

[101] D. P. Dannels and K. N. Martin, "Critiquing critiques: A genre analysis of feedback across novice to expert design studios," *Journal of Business and Technical Communication*, vol. 22, no. 2, pp. 135–159, 2008.

[102] J. S. Linsey, K. L. Wood, and A. B. Markman, "Increasing innovation: presentation and evaluation of the wordtree design-by-analogy method," in *ASME 2008 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, pp. 21–32, American Society of Mechanical Engineers Digital Collection, 2009.

[103] Y. Doppelt, "Assessing creative thinking in design-based learning," *International Journal of Technology and Design Education*, vol. 19, no. 1, pp. 55–65, 2009.

[104] H. Cai, E. Y.-L. Do, and C. M. Zimring, "Extended linkography and distance graph in design evaluation: an empirical study of the dual effects of inspiration sources in creative design," *Design studies*, vol. 31, no. 2, pp. 146–168, 2010.

[105] S. Chance, "Strategic by design: Iterative approaches to educational planning," *Planning for Higher Education*, vol. 38, no. 2, p. 40, 2010.

[106] S. P. Dow, A. Glassco, J. Kass, M. Schwarz, D. L. Schwartz, and S. R. Klemmer, "Parallel prototyping leads to better design results, more divergence, and increased self-efficacy," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 17, no. 4, pp. 1–24, 2010.

[107] B. M. Hanington, "Relevant and rigorous: Human-centered research and design education," *Design Issues*, vol. 26, no. 3, pp. 18–26, 2010.

[108] M. Lande and L. Leifer, "Difficulties student engineers face designing the future," *International Journal of Engineering Education*, vol. 26, no. 2, p. 271, 2010.

[109] D. P. Dannels, A. L. Housley Gaffney, and K. N. Martin, "Students' talk about the climate of feedback interventions in the critique," *Communication Education*, vol. 60, no. 1, pp. 95–114, 2011.

[110] J. J. Ham and M. A. Schnabel, "Web 2.0 virtual design studio: social networking as facilitator of design education," *Architectural science review*, vol. 54, no. 2, pp. 108–116, 2011.

[111] T. Kershaw, K. Holtta-Otto, and Y. S. Lee, "The effect of prototyping and critical feedback on fixation in engineering design," in *Proceedings of the Annual Meeting of the Cognitive Science Society*, pp. 807–812, 2011.

[112] S. Palmer and W. Hall, "An evaluation of a project-based learning initiative in engineering education," *European Journal of Engineering Education*, vol. 36, pp. 357–365, aug 2011.

[113] J. Y. Park, "Design education online: Learning delivery and evaluation," *International Journal of Art and Design Education*, vol. 30, pp. 176–187, jun 2011.

[114] J. McCarthy, "International design collaboration and mentoring for tertiary students through facebook," *Australasian Journal of Educational Technology*, vol. 28, no. 5, 2012.

[115] S. V. McLaren, "Assessment is for learning: supporting feedback," *International Journal of Technology and Design Education*, vol. 22, no. 2, pp. 227–245, 2012.

[116] T. Schrand and J. Eliason, "Feedback practices and signature pedagogies: what can the liberal arts learn from the design critique?," *Teaching in Higher Education*, vol. 17, no. 1, pp. 51–62, 2012.

[117] Y. L. Wong and K. W. M. Siu, "A model of creative design process for fostering creativity of students in design education," *International Journal of Technology and Design Education*, vol. 22, no. 4, pp. 437–450, 2012.

[118] D. Tinapple, L. Olson, and J. Sadauskas, "Critviz: Web-based software supporting peer critique in large creative classrooms," *Bulletin of the IEEE Technical Committee on Learning Technology*, vol. 15, no. 1, p. 29, 2013.

[119] X. Yuan and J.-h. Lee, "Toward a computational approach of creativity assessment in product design," in *International Conference on Computer-Aided Architectural Design Futures*, pp. 50–62, Springer, 2013.

[120] X. Wei, D. Weng, Y. Liu, and Y. Wang, "Teaching based on augmented reality for a technical creative design course," *Computers & Education*, vol. 81, pp. 221–234, 2015.

[121] A. Xu, H. Rao, S. P. Dow, and B. P. Bailey, "A classroom study of using crowd feedback in the iterative design process," in *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*, pp. 1637–1648, 2015.

[122] R. S. Adams, T. Forin, M. Chua, and D. Radcliffe, "Characterizing the work of coaching during design reviews," *Design Studies*, vol. 45, pp. 30–67, 2016.

[123] B. T. Christensen and L. J. Ball, "Dimensions of creative evaluation: Distinct design and reasoning strategies for aesthetic, functional and originality judgments," *Design Studies*, vol. 45, pp. 116–136, 2016.

[124] H. D. Arslan, G. Özer, and A. Özkiş, "Evaluation of final product integrated with intelligent systems in architectural education studios," *Online Journal of Art and Design*, vol. 5, no. 3, p. 119, 2017.

[125] C. M. Gray, M. Exter, and T. S. Krause, "Moving Towards Individual Competence From Group Work in Transdisciplinary Education," in *2017 ASEE Annual Conference & Exposition*, (Columbus, Ohio), ASEE Conferences, jun 2017. https://peer.asee.org/28691.

[126] C. M. Gray, "Democratizing assessment practices through multimodal critique in the design classroom," *International Journal of Technology and Design Education*, pp. 1–18, 2018.

[127] S. Soni, P. Tandon, and P. Khanna, "Knowledge system for early phase aesthetic concept generation in industrial design," *International Journal of Intelligent Systems Design and Computing*, vol. 2, no. 1, pp. 45–75, 2018.

[128] R. Buchanan, "Wicked problems in design thinking," *Design issues*, vol. 8, no. 2, pp. 5–21, 1992.

[129] M. Pandir and J. Knight, "Homepage aesthetics: The search for preference factors and the challenges of subjectivity," *Interacting with Computers*, vol. 18, no. 6, pp. 1351–1370, 2006.

[130] X. S. Zheng, I. Chakraborty, J. J.-W. Lin, and R. Rauschenberger, "Correlating low-level image statistics with users-rapid aesthetic and affective judgments of web pages," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1–10, 2009.

[131] K. Dorst and N. Cross, "Creativity in the design process: co-evolution of problem–solution," *Design studies*, vol. 22, no. 5, pp. 425–437, 2001.

[132] J. A. Greene, R. Freed, and R. K. Sawyer, "Fostering creative performance in art and design education via self-regulated learning," *Instructional Science*, vol. 47, no. 2, pp. 127–149, 2019.

[133] T. A. Dutton, "Design and studio pedagogy," *Journal of architectural education*, vol. 41, no. 1, pp. 16–25, 1987.

[134] L. Lindström, "Creativity: What is it? can you assess it? can it be taught?," *International Journal of Art & Design Education*, vol. 25, no. 1, pp. 53–66, 2006.

[135] H. W. J. Rittel and M. M. Webber, "Dilemmas in a general theory of planning," *Policy sciences*, vol. 4, no. 2, pp. 155–169, 1973.

[136] N. Cross, "Designerly ways of knowing," *Design studies*, vol. 3, no. 4, pp. 221–227, 1982.

[137] J. Zimmerman, J. Forlizzi, and S. Evenson, "Research through design as a method for interaction design research in HCI," in *Proceedings of the ACM CHI*, pp. 493–502, ACM, 2007.

[138] E. Stolterman, "The nature of design practice and implications for interaction design research," *International Journal of Design*, vol. 2, no. 1, 2008.

[139] Q. Yang, A. Steinfeld, C. Rosé, and J. Zimmerman, "Re-examining whether, why, and how human-ai interaction is uniquely difficult to design," in *Proceedings of the 2020 chi conference on human factors in computing systems*, pp. 1–13, 2020.

[140] G. Dove, K. Halskov, J. Forlizzi, and J. Zimmerman, "Ux design innovation: Challenges for working with machine learning as a design material," in *Proceedings of the 2017 chi conference on human factors in computing systems*, pp. 278–288, 2017.

[141] W. Chen, "Exploring the learning problems and resource usage of undergraduate industrial design students in design studio courses," *International Journal of Technology and Design Education*, vol. 26, no. 3, pp. 461–487, 2016.

[142] J. McDonnell, "Scaffolding practices: A study of design practitioner engagement in design education," *Design studies*, vol. 45, pp. 9–29, 2016.

[143] K. Luther, J.-L. Tolentino, W. Wu, A. Pavel, B. P. Bailey, M. Agrawala, B. Hartmann, and S. P. Dow, "Structuring, aggregating, and evaluating crowdsourced design critique," in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pp. 473–485, ACM, 2015.

[144] C. Sjöberg and T. Timpka, “Participatory design of information systems in health care,” *Journal of the American Medical Informatics Association*, vol. 5, no. 2, pp. 177–183, 1998.

[145] R. Luck, “Dialogue in participatory design,” *Design studies*, vol. 24, no. 6, pp. 523–535, 2003.

[146] C. Frauenberger, J. Good, and W. Keay-Bright, “Designing technology for children with special needs: bridging perspectives through participatory design,” *CoDesign*, vol. 7, no. 1, pp. 1–28, 2011.

[147] M. Davis, *Teaching Design: a guide to curriculum and pedagogy for college design faculty and teachers who use design in their classrooms*. Simon and Schuster, 2017.

[148] H. L. Erickson, *Concept-based curriculum and instruction: Teaching beyond the facts*. Corwin Press, 2002.

[149] C. Tang, P. Lai, D. Arthur, and S. F. Leung, “How do students prepare for traditional and portfolio assessment in a problem-based learning curriculum,” in *Themes and Variations in PBL: refereed proceedings of the 1999 Bi-ennial PBL Conference*, vol. 1, pp. 206–217, Australia ProblemBased LearningNetwork (PROBLARC), 1999.

[150] K. Kreijns, P. A. Kirschner, and W. Jochems, “Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: a review of the research,” *Computers in human behavior*, vol. 19, no. 3, pp. 335–353, 2003.

[151] N. W. Kohn and S. M. Smith, “Collaborative fixation: Effects of others’ ideas on brainstorming,” *Applied Cognitive Psychology*, vol. 25, no. 3, pp. 359–371, 2011.

[152] N. M. Webb, “Student interaction and learning in small groups,” *Review of Educational research*, vol. 52, no. 3, pp. 421–445, 1982.

[153] M.-J. Taylor and C. McCormack, “Juggling cats: investigating effective verbal feedback in graphic design critiques,” in *A paper to The Australian Council of University Art and Design Schools Annual Conference*, pp. 22–25, 2004.

[154] J. K. Whyte, B. Ewenstein, M. Hales, and J. Tidd, "Visual practices and the objects used in design," *Building Research & Information*, vol. 35, no. 1, pp. 18–27, 2007.

[155] E. Giaccardi and E. Karana, "Foundations of materials experience: An approach for hci," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, (New York, NY, USA), pp. 2447–2456, Association for Computing Machinery, 2015.

[156] B. Green and Y. Chen, "The principles and limits of algorithm-in-the-loop decision making," *Proceedings of the ACM on Human-Computer Interaction*, vol. 3, no. CSCW, p. 50, 2019.

[157] B. Green and Y. Chen, "Disparate interactions: An algorithm-in-the-loop analysis of fairness in risk assessments," in *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pp. 90–99, ACM, 2019.

[158] B. Shneiderman, "Human-centered artificial intelligence: Reliable, safe & trustworthy," *International Journal of Human–Computer Interaction*, pp. 1–10, 2020.

[159] V. Bellotti and K. Edwards, "Intelligibility and accountability: human considerations in context-aware systems," *Human–Computer Interaction*, vol. 16, no. 2-4, pp. 193–212, 2001.

[160] B. Jiang and J. Canny, "Interactive machine learning via a gpu-accelerated toolkit," in *Proceedings of the 22nd International Conference on Intelligent User Interfaces*, pp. 535–546, ACM, 2017.

[161] L. R. Lippard, *Six years: the dematerialization of the art object from 1966 to 1972*, vol. 364. Univ of California Press, 1997.

[162] A. Kaprow, "Happenings in the New York scene," in *The Improvisation Studies Reader*, pp. 254–260, Routledge, 2014.

[163] H. Higgins, *Fluxus experience*. Univ of California Press, 2002.

[164] W. C. Seitz, N. of modern art (New York, D. M. of Contemporary Arts (Tex.)., and S. F. M. of Modern Art (Calif.)., *The art of assemblage*. Museum of Modern Art New York, 1961.

[165] J. Guilford, "Creativity," *American Psychologist*, vol. 5, pp. 444–454, 1950.

[166] D. K. Simonton, "Creativity as blind variation and selective retention: Is the creative process darwinian?," *Psychological Inquiry*, vol. 10, no. 4, 1999.

[167] A. Baddeley, "Working memory," *Science*, vol. 255, no. 5044, pp. 556–559, 1992.

[168] A. M. Glenberg and W. E. Langston, "Comprehension of illustrated text: Pictures help to build mental models," *Journal of memory and language*, vol. 31, no. 2, pp. 129–151, 1992.

[169] R. E. Mayer and R. Moreno, "Nine ways to reduce cognitive load in multimedia learning," *Educational psychologist*, vol. 38, no. 1, pp. 43–52, 2003.

[170] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent trends in deep learning based natural language processing," *ieee Computational intelligenCe magazine*, vol. 13, no. 3, pp. 55–75, 2018.

[171] V. Ordonez, G. Kulkarni, and T. L. Berg, "Im2text: Describing images using 1 million captioned photographs," in *Advances in neural information processing systems*, pp. 1143–1151, 2011.

[172] Q. Yu, Y. Yang, F. Liu, Y.-Z. Song, T. Xiang, and T. M. Hospedales, "Sketch-a-net: A deep neural network that beats humans," *International journal of computer vision*, vol. 122, no. 3, pp. 411–425, 2017.

[173] S. Mednick, "The associative basis of the creative process.," *Psychological review*, vol. 69, no. 3, p. 220, 1962.

[174] K. Fu, J. Chan, J. Cagan, K. Kotovsky, C. Schunn, and K. Wood, "The meaning of "near" and "far": the impact of structuring design databases and the effect of distance of analogy on design output," *Journal of Mechanical Design*, vol. 135, no. 2, 2013.

[175] S. Vattam, B. Wiltgen, M. Helms, A. K. Goel, and J. Yen, "DANE: fostering creativity in and through biologically inspired design," in *Design Creativity 2010*, pp. 115–122, Springer, 2011.

[176] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1532–1543, 2014.

[177] D. Gentner, "Structure-mapping: A theoretical framework for analogy," *Cognitive science*, vol. 7, no. 2, pp. 155–170, 1983.

[178] E. R. Tufte, N. H. Goeler, and R. Benson, *Envisioning information*, vol. 126. Graphics press Cheshire, CT, 1990.

[179] J. Bertin, *Semiology of Graphics*. University of Wisconsin Press, 1983.

[180] P. B. Meggs, *Type and image: The language of graphic design*. John Wiley & Sons, 1992.

[181] P. Siangliulue, J. Chan, S. P. Dow, and K. Z. Gajos, "IdeaHound: improving large-scale collaborative ideation with crowd-powered real-time semantic modeling," in *Proc. UIST*, pp. 609–624, ACM, 2016.

[182] F. Bação, V. Lobo, and M. Painho, "The self-organizing map, the geo-som, and relevant variants for geosciences," *Computers & Geosciences*, vol. 31, no. 2, pp. 155–163, 2005.

[183] J. Itten, *The elements of color*. John Wiley & Sons, 1970.

[184] H. Windisch, *Schule der Farbenphotographie*. Heering-Verlag, 1941.

[185] W. Osberger and A. J. Maeder, "Automatic identification of perceptually important regions in an image," in *Proceedings. Fourteenth International Conference on Pattern Recognition (Cat. No. 98EX170)*, vol. 1, pp. 701–704, IEEE, 1998.

[186] T. Rikakis, A. Kelliher, J.-B. Huang, and H. Sundaram, "Progressive cyber-human intelligence for social good," *Interactions*, vol. 25, no. 4, pp. 52–56, 2018.

[187] D. Turnbull and H. Watson, *Maps Are Territories Science is an Atlas: A Portfolio of Exhibits*. University of Chicago Press, 1993.

[188] G. Siemens and P. Long, "Penetrating the fog: Analytics in learning and education.," *EDUCAUSE review*, vol. 46, no. 5, p. 30, 2011.

[189] S. Card, J. Mackinlay, and B. Shneiderman, *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999.

[190] S. Few, *Information Dashboard Design: Displaying data for at-a-glance monitoring*, vol. 5. Analytics Press Burlingame, CA, 2013.

[191] M. A. Yalçın, N. Elmqvist, and B. B. Bederson, "Keshif: Rapid and expressive tabular data exploration for novices," *IEEE transactions on visualization and computer graphics*, vol. 24, no. 8, pp. 2339–2352, 2017.

[192] C. Michel, E. Lavoué, S. George, and M. Ji, "Supporting awareness and self-regulation in project-based learning through personalized dashboards," *International Journal of Technology Enhanced Learning*, vol. 9, no. 2–3, 2017.

[193] R. K. Bellamy, K. Dey, M. Hind, S. C. Hoffman, S. Houde, K. Kannan, P. Lohia, J. Martino, S. Mehta, A. Mojsilovic, *et al.*, "Ai fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias," *arXiv preprint arXiv:1810.01943*, 2018.

[194] A. Woodruff, "10 things you should know about algorithmic fairness," *interactions*, vol. 26, no. 4, pp. 47–51, 2019.

[195] A. Woodruff, S. E. Fox, S. Rousso-Schindler, and J. Warshaw, "A qualitative exploration of perceptions of algorithmic fairness," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2018.

[196] D. A. Hudson and C. D. Manning, "Compositional attention networks for machine reasoning," *arXiv preprint arXiv:1803.03067*, 2018.

[197] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.

[198] B. Beatty and C. Ulasewicz, "Faculty perspectives on moving from Blackboard to the Moodle learning management system," *TechTrends*, vol. 50, no. 4, pp. 36–45, 2006.

[199] T. Berners-Lee, R. Cailliau, A. Luotonen, H. F. Nielsen, and A. Secret, "The world-wide web," *Communications of the ACM*, vol. 37, no. 8, pp. 76–82, 1994.

[200] T. Nelson, "On the xanadu project," *BYTE Magazine*, vol. 15, no. 9, pp. 298–299, 1990.

[201] T. Nelson, *Literary machines 93.1*. Mindful Pr, 1993.

[202] C. C. Marshall and F. M. Shipman III, "Spatial hypertext: designing for change," *Communications of the ACM*, vol. 38, no. 8, pp. 88–97, 1995.

[203] Sympli, "Versions: Git for Designers." https://versions.sympli.io/, 2020. Last accessed: 2020-03-30.

[204] T. Jung and E. Y.-L. Do, "Immersive redliner: collaborative design in cyberspace," in *ACADIA Eternity, infinity and virtuality in architecture*, pp. 185–194, CUMINCAD, 2000.

[205] J. Kim, M. Agrawala, and M. S. Bernstein, "Mosaic: designing online creative communities for sharing works-in-progress," in *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pp. 246–258, ACM, 2017.

[206] M. Kyng and J. Greenbaum, *Design at work: Cooperative design of computer systems*. Erlbaum, Lawrence, Associates, 1991.

[207] B. DiSalvo and C. DiSalvo, "Designing for democracy in education: Participatory design and the learning sciences," in *Proceedings of the Eleventh International Conference of the Learning Sciences*, pp. 793–799, Boulder, CO: International Society of the Learning Sciences, 2014.

[208] W. A. Hamilton, N. Lupfer, N. Botello, T. Tesch, A. Stacy, J. Merrill, B. Williford, F. R. Bentley, and A. Kerne, "Collaborative Live Media Curation: Shared Context for Participation in Online Learning," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, ACM, 2018.

[209] S. L. Star and J. R. Griesemer, “Institutional ecology,translations’ and boundary objects: Amateurs and professionals in Berkeley’s Museum of Vertebrate Zoology, 1907-39,” *Social studies of science*, vol. 19, no. 3, pp. 387–420, 1989.

[210] D. Suthers and K. Verbert, “Learning analytics as a middle space,” in *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, pp. 1–4, ACM, 2013.

[211] H. Sankey, “Scientific realism and the God’s eye point of view,” *Epistemologia*, vol. 27, no. 2, pp. 211–26, 2003.

[212] Q. Yang, A. Steinfeld, and J. Zimmerman, “Unremarkable ai: Fitting intelligent decision support into critical, clinical decision-making processes,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–11, 2019.

[213] G. Hinton, “Geoffrey Hinton on Twitter: "Suppose you have cancer and you have to choose between a black box AI surgeon that cannot explain how it works but has a 90% cure rate and a human surgeon with an 80% cure rate. Do you want the AI surgeon to be illegal?".” https://twitter.com/geoffreyhinton/status/1230592238490615816, 2020. Last accessed: 2020-03-19.

[214] M. L. Loughry, M. W. Ohland, and D. J. Woehr, “Assessing teamwork skills for assurance of learning using catme team tools,” *Journal of Marketing Education*, vol. 36, no. 1, pp. 5–19, 2014.

[215] A. Kerne, *CollageMachine: A model of interface ecology*. PhD thesis, New York University, Graduate School of Arts and Science, 2001.

[216] National Academies of Sciences Engineering and Medicine, *Workforce development and intelligence analysis for national security purposes: Proceedings of a workshop*. National Academies Press, 2018.

[217] J. A. Broadbent and N. Cross, “Design education in the information age,” *Journal of Engineering Design*, vol. 14, no. 4, pp. 439–446, 2003.

[218] G. Fischer, "Designing socio-technical environments in support of meta-design and social creativity," in *Proceedings of the 8th iternational conference on Computer supported collaborative learning*, pp. 2–11, International Society of the Learning Sciences, 2007.

[219] D. Cai, S. Yu, J.-R. Wen, and W.-Y. Ma, "Vips: a vision-based page segmentation algorithm," tech. rep., Microsoft, November 2003.

[220] C. Eames and R. Eames, "Powers of ten," *Pyramid Films*, 1968.

[221] K. Perlin and D. Fox, "Pad: an alternative approach to the computer interface," in *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, pp. 57–64, 1993.

[222] B. B. Bederson, "The promise of zoomable user interfaces," *Behaviour & Information Technology*, vol. 30, no. 6, pp. 853–866, 2011.

[223] Y. Bar-Yam, "Engineering complex systems: multiscale analysis and evolutionary engineering," in *Complex engineered systems*, pp. 22–39, Springer, 2006.

[224] J. D. Summers and J. J. Shah, "Mechanical engineering design complexity metrics: size, coupling, and solvability," *Journal of Mechanical Design*, vol. 132, no. 2, 2010.

[225] J. S. Gero and M. L. Maher, *Modeling creativity and knowledge-based creative design*. Psychology Press, 1993.

[226] J. Koch and A. Oulasvirta, "Computational layout perception using gestalt laws," in *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 1423–1429, 2016.

[227] Q. Liu, M. Deng, Y. Shi, and J. Wang, "A density-based spatial clustering algorithm considering both spatial proximity and attribute similarity," *Computers & Geosciences*, vol. 46, pp. 296–309, 2012.

[228] M. Deng, Q. Liu, T. Cheng, and Y. Shi, "An adaptive spatial clustering algorithm based on delaunay triangulation," *Computers, Environment and Urban Systems*, vol. 35, no. 4, pp. 320–332, 2011.

[229] V. Estivill-Castro and I. Lee, "Autoclust: Automatic clustering via boundary extraction for mining massive point-data sets," in *In Proceedings of the 5th International Conference on Geocomputation*, 2000.

[230] K. E. Arnold and M. D. Pistilli, "Course signals at Purdue: Using learning analytics to increase student success," in *Proceedings of the 2nd international conference on learning analytics and knowledge*, pp. 267–270, ACM, 2012.

[231] T. H. Grubesic, R. Wei, and A. T. Murray, "Spatial clustering overview and comparison: Accuracy, sensitivity, and computational expense," *Annals of the Association of American Geographers*, vol. 104, no. 6, pp. 1134–1156, 2014.

[232] A. Aksaç, T. Özyer, and R. Alhajj, "Cutesc: cutting edge spatial clustering technique based on proximity graphs," *Pattern Recognition*, vol. 96, p. 106948, 2019.

[233] G. Karypis, E.-H. Han, and V. Kumar, "Chameleon: Hierarchical clustering using dynamic modeling," *Computer*, vol. 32, no. 8, pp. 68–75, 1999.

[234] M. L. McHugh, "Interrater reliability: the kappa statistic," *Biochemia medica: Biochemia medica*, vol. 22, no. 3, pp. 276–282, 2012.

[235] C. D. Manning, H. Schütze, and P. Raghavan, *Introduction to information retrieval*. Cambridge university press, 2008.

[236] Y. Zhao and G. Karypis, "Evaluation of hierarchical clustering algorithms for document datasets," in *Proceedings of the eleventh international conference on Information and knowledge management*, pp. 515–524, 2002.

[237] D. Doran, S. Schulz, and T. R. Besold, "What does explainable AI really mean? A new conceptualization of perspectives," *arXiv preprint arXiv:1710.00794*, 2017.

[238] D. Gunning, "Explainable artificial intelligence (xai)," *Defense Advanced Research Projects Agency (DARPA)*, 2017.

[239] W. Samek, T. Wiegand, and K.-R. Müller, "Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models," *arXiv preprint arXiv:1708.08296*, 2017.

[240] E. Balcetis and D. Dunning, "See what you want to see: motivational influences on visual perception.," *Journal of personality and social psychology*, vol. 91, no. 4, p. 612, 2006.

[241] J. Wagemans, J. H. Elder, M. Kubovy, S. E. Palmer, M. A. Peterson, M. Singh, and R. von der Heydt, "A century of gestalt psychology in visual perception: I. perceptual grouping and figure–ground organization.," *Psychological bulletin*, vol. 138, no. 6, p. 1172, 2012.

[242] J. Beck, "Effect of orientation and of shape similarity on perceptual grouping," *Perception & Psychophysics*, vol. 1, no. 5, pp. 300–302, 1966.

[243] M. Yin, J. Wortman Vaughan, and H. Wallach, "Understanding the effect of accuracy on trust in machine learning models," in *Proceedings of the 2019 chi conference on human factors in computing systems*, pp. 1–12, 2019.

[244] W. W. Gaver, J. Beaver, and S. Benford, "Ambiguity as a resource for design," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 233–240, 2003.

[245] S. Amershi, D. Weld, M. Vorvoreanu, A. Fourney, B. Nushi, P. Collisson, J. Suh, S. Iqbal, P. N. Bennett, K. Inkpen, *et al.*, "Guidelines for human-ai interaction," in *Proceedings of the 2019 chi conference on human factors in computing systems*, pp. 1–13, 2019.

[246] M. O. Gurel and I. Basa, "The status of graphical presentation in interior/architectural design education," *International Journal of Art & Design Education*, vol. 23, no. 2, pp. 192–206, 2004.

[247] S. Ruder, "Nlp-progress," *London (UK): Sebastian Ruder (accessed 2020-09-01). https://nlpprogress. com*, 2020.

[248] M. B. Ben-Av and D. Sagi, "Perceptual grouping by similarity and proximity: Experimental results can be predicted by intensity autocorrelations," *Vision research*, vol. 35, no. 6, pp. 853–866, 1995.

[249] J. S. Gero, "Computational models of creative design processes," in *Artificial intelligence and creativity*, pp. 269–281, Springer, 1994.

[250] B. S. Bloom *et al.*, "Taxonomy of educational objectives. vol. 1: Cognitive domain," *New York: McKay*, vol. 20, p. 24, 1956.

[251] N. E. Adams, "Bloom's taxonomy of cognitive learning objectives," *Journal of the Medical Library Association: JMLA*, vol. 103, no. 3, p. 152, 2015.

[252] J. Davies, E. de Graaff, and A. Kolmos, "Pbl across the disciplines: research into best practice," in *The 3rd International Research Symposium on PBL. Aalborg: Aalborg Universitetsforlag*, 2011.

[253] A. Abdul, J. Vermeulen, D. Wang, B. Y. Lim, and M. Kankanhalli, "Trends and trajectories for explainable, accountable and intelligible systems: An hci research agenda," in *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1–18, 2018.

[254] B. Tversky, J. B. Morrison, and M. Betrancourt, "Animation: can it facilitate?," *International journal of human-computer studies*, vol. 57, no. 4, pp. 247–262, 2002.

[255] R. E. Mayer and R. Moreno, "Animation as an aid to multimedia learning," *Educational psychology review*, vol. 14, no. 1, pp. 87–99, 2002.

[256] B. Liz, T. Dreyfus, J. Mason, P. Tsamir, A. Watson, and O. Zaslavsky, "Exemplification in mathematics education," in *Proceedings of the 30th Conference of the International Group for the Psychology of Mathematics Education*, vol. 1, pp. 126–154, ERIC, 2006.

[257] K. S. Taber, "Challenging gifted learners: General principles for science educators; and exemplification in the context of teaching chemistry.," *Science Education International*, vol. 21, no. 1, pp. 5–30, 2010.

[258] V. L. Uskov, J. P. Bakken, A. Shah, N. Hancher, C. McPartlin, and K. Gayke, "Innovative interlabs system for smart learning analytics in engineering education," in *2019 IEEE Global Engineering Education Conference (EDUCON)*, pp. 1363–1369, IEEE, 2019.

[259] D. Wang, Q. Yang, A. Abdul, and B. Y. Lim, "Designing theory-driven user-centric explainable ai," in *Proceedings of the 2019 CHI conference on human factors in computing systems*, pp. 1–15, 2019.

[260] A. Bussone, S. Stumpf, and D. O'Sullivan, "The role of explanations on trust and reliance in clinical decision support systems," in *2015 International Conference on Healthcare Informatics*, pp. 160–169, IEEE, 2015.

[261] A. Deeks, "The judicial demand for explainable artificial intelligence," *Columbia Law Review*, vol. 119, no. 7, pp. 1829–1850, 2019.

[262] National Academy of Engineering, *Rising Above the Gathering Storm, Revisited: Rapidly Approaching Category 5*. The National Academies Press, 2010.

[263] "Artificial Intelligence (AI) at National Science Foundation (NSF)." https://www.nsf.gov/cise/ai.jsp. Last accessed: 2020-10-11.

[264] J. Heer, "Agency plus automation: Designing artificial intelligence into interactive systems," *Proceedings of the National Academy of Sciences*, vol. 116, no. 6, pp. 1844–1850, 2019.

[265] J. Mongan and M. Kohli, "Artificial intelligence and human life: Five lessons for radiology from the 737 max disasters," *Radiology: Artificial Intelligence*, vol. 2, no. 2, p. e190111, 2020.

APPENDIX A

CO-DESIGN: DISCUSSION TOPICS AND QUESTIONS

We engaged design instructors in an ongoing discourse—spanning a period of more than one year—discussing their course learning objective, assignment specifications, assessment and feedback practices. We discussed whether and how computational means could support them in their assessment and feedback. We term these as discussions because we shared perspectives that we gained from prior work and by engaging with other instructors. In addition to below, emergent topics included students' use of sketching, instructors' creativity assessment, and students' considering instructor feedback beyond a specific deliverable.

A.1 Individual Discussion

- What design courses do you teach?
- How do you define learning objectives for each course? What are the guiding principles?
- As a design instructor, why do you think these specific objectives are important?
- How do you prioritize?
- What assignments do you define for meeting the objectives?
- Do the assignments vary at times? How? Why?
- How do you formulate the grading rubric for an assignment?
- Are there weights for the design product and process?
- What is given more weight in assignments: understanding course material, application of learned principles, visual design, creative ideas, other factors?
- Do weights vary across assignments? How?

- How are grading rubrics applied? Are there macro and micro level guidelines based on the learning objectives of that assignment?
- What instructions are TAs / graders given toward applying the rubric? How is consistency ensured?
- Do you use peer review for grading? Do the guidelines differ? How?
- How do you understand the picture of a class? for e.g., How is a class doing overall? What are common and specific problems?
- How do you address the common and specific problems?
- Do you think that the design process / addressing of problems could be better supported by seeing students' progress more regularly?
- What do you think about our intention to design computer algorithms that measure aspects of design work that are representative of the criteria in your grading rubrics?
- Would some kind of analytics based on the grading rubric be helpful?
- Would some representations of their work be useful, e.g. conceptual themes, assignment sections?
- Would a dashboard interface help in this regard (show a mockup)?
- What else could be useful?
- What features would you like to see? Examples, sorting, mean, median, outliers?
- What other features do you think would be useful?

A.2 Workshop Discussion

For workshop, while we had put together a list of topics, the discussion emerged on its own after instructors finished presenting their design ideation assignments, along with assessment approach. Topics that emerged included:

- the sizes of courses, teaching teams, and student teams;
- the level of details in assignment and rubric specifications;
- challenges in tracking students' progress and incorporation of feedback; and
- difficulty assessing contributions in team assignments.

APPENDIX B

USE OF DESIGN CHARACTERISTICS ACROSS FIELDS

In Section 3.4.2, we developed a set of situated design creativity analytics—Fluency, Flexibility, Visual Consistency, Multiscale Organization, and Legible Contrast—that can provide AI-based, on demand, actionable assessment to design students. Here, we enumerate additional characteristics identified through co-design engagements with design instructors across fields (Table B.1).

Analytic	**Landscape Architecture & Urban Planning**	**Computer Science and Engineering**	**Mechanical Engineering**	**Interactive Art & Design**	**Architecture**
Fluency			•	•	
Flexibility		•	•	•	
Visual Consistency	•	•			•
Multiscale Organization	•	•			
Legible Contrast	•	•			
Team Members' Contribution	•	•	•	•	
Novelty		•	•		
Relative Object Sizes	•				•
Number of Iterations	•		•		
Use of Small Multiples [178]	•				
Range of Colors	•				
Coverage of Assigned Sources		•			
Edits during Practice	•				

Table B.1: Learning analytics relevant to design education. In the top part, we list analytics for which we developed AI-based approaches in Chapter 3. In the bottom part, we list additional characteristics that instructors in our study emphasize. We prioritized developing analytics based on: a) the number of contexts which emphasize a design characteristic and b) the extent of conceptual and empirical evidence we identified in regard to developing an effective AI approach for the corresponding analytic.

APPENDIX C

ADAPTED AMOEBA: ORIGINAL ALGORITHM COMBINED WITH THE CHANGES

In Section 4.2.2, we described changes for adapting AMOEBA for multiscale clustering of dimensional design elements. We include the adapted algorithm below, i.e., original algorithm combined with the changes. We reproduce the visual representation of the changes here (Figure C.1), to make it convenient to connect it with corresponding textual annotations on the original algorithm.

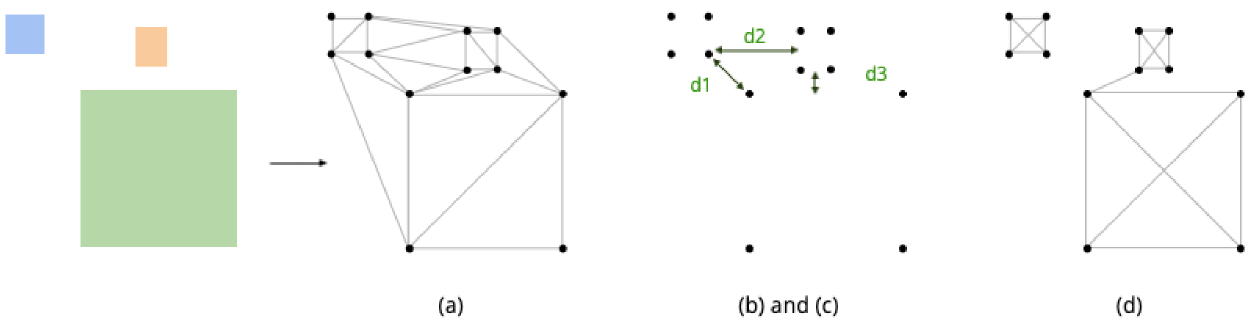

Figure C.1: The adapted algorithm adds four aspects to the original algorithm: (a) compute Delaunay graph using all vertices of an element, (b) instead of using Delaunay edges directly, find the shortest distance between elements represented by the edge, (c) relatedly, ignore any edges between vertices of the same element, and (d) when calculating subgraphs, add edges among all vertices of an element. The below algorithmic representation highlights these added aspects.

```
procedure ADAPTED_AMOEBA(Graph)
  begin
  Graph = CreateDelaunayGraph(ElementVertices); // (a): Add each vertex
      of the element
  WriteCluster(Graph); // Write the Graph as a cluster
  // (b): Find shortest distance between elements represented by each
      edge in the graph
```

```
// In the process, add edges connecting vertices of the same element
   to IntraElementEdges
DistanceMap = FindShortestDistance(Graph, IntraElementEdges);
// (c): Delete edges connecting vertices of the same element
Graph.del_edges(IntraElementEdges);
// Use the shortest distance when computing global mean and standard
   deviation
NumberOfEdges = CalculateMeanandStDev(Graph, GlobalMean, GlobalStDev,
   DistanceMap)
if (NumberOfEdges <= 1)
  return;
for each node v in Graph do {
  EdgeList = Graph.adjacent_edges(v); // Extract edges incident to
     node v
  // Use the shortest distance when computing local mean
  LocalMean = CalculateLocalMean(EdgeList, DistanceMap);
  for each edge e in EdgeList do {
    ToleranceValue = ALPHA * GlobalStDev * GlobalMean / LocalMean;
    if (e.distance() >= (GlobalMean + ToleranceValue))
      DeleteEdgeList.append(e);
  }
}
Graph.del_edges(DeleteEdgeList); // Eliminate passive edges
if (Graph.degree(v) == 0)
  Graph.delete_node(v);
AddIntraElementEdges(Graph); // (d): Add edges among all vertices of
   the same element
ConnectedComponents(Graph, ComponentNumber); // Calculate connected
   components
for each connected component c do {
  SubGraph = ConstructSubGraph(Graph, ComponentNumber, c);
```

```
    if (NumberOfEdges != SubGraph.number_of_nodes());
      ADPATED_AMOEBA(SubGraph);
  }
end
```

APPENDIX D

MULTISCALE ANALYTICS STUDY: INTERVIEW QUESTIONS

We used the following questions to guide our semi-structured interviews:

- Please briefly describe your experiences with the courses dashboard.
- Do you think the class would be different with and without the dashboard? If so, how?
- How does how you use the courses dashboard compare with other learning management systems and environments? What is similar? Is anything different?
- Has using the dashboard shown you anything new or unexpected about your students' learning? If yes, what?
- What do you understand about the analytics presented on the dashboard with submissions?
- Do you utilize analytics? If so, do they support in monitoring and intervening? Assessment and feedback? How?
- If the answer to *'Do you utilize analytics'* is *'No'*: Do you think these analytics have the potential to become a part of the assessment and feedback that you provide to the students? If so, how?
- What do you think about showing these analytics to students on-demand?
- Did you click on 'Scales' analytics? How did seeing its relationship with the actual design work affect your utilization (or potential utilization) for assessment and feedback?
- Did you click on 'Clusters' analytics? How did seeing its relationship with the actual design work affect your utilization (or potential utilization) for assessment and feedback?
- Has using the dashboard to follow and track student design work changed how you teach or interact with the students? If so, how?

- What would you do different, if anything, next time you teach the class?
- What are your suggestions for making the dashboard more suited for your teaching and assessment practices? Or for design education in general?