Evaluating Creativity Support Tools in HCI Research

Christian Remy, Lindsay MacDonald Vermeulen, Jonas Frich, Michael Mose Biskjaer, Peter Dalsgaard
Aarhus University, Centre for Digital Creativity
Helsingforsgade 14, 8200 Aarhus, Denmark
{remy, lindsay.macdonald, frich, mmb, dalsgaard}@cc.au.dk

ABSTRACT
The design and development of Creativity Support Tools (CSTs) is of growing interest in research at the intersection of creativity and Human-Computer Interaction, and has been identified as a 'grand challenge for HCI'. While creativity research and HCI each have had long-standing discussions about—and rich toolboxes of—evaluation methodologies, the nascent field of CST evaluation has so far received little attention. We contribute a survey of 113 research papers that present and evaluate CSTs, and we offer recommendations for future CST evaluation. We center our discussion around six major points that researchers might consider: 1) Clearly define the goal of the CST; 2) link to theory to further understanding of usage of CSTs; 3) recruit domain experts, if applicable and feasible; 4) consider longitudinal, in-situ studies; 5) distinguish and decide whether to evaluate usability or creativity; and 6) as a community, help develop a toolbox for CST evaluation.

Author Keywords
Creativity; Creativity Support Tools; CSTs; Creativity Metrics; Evaluation; Literature Survey

CCS Concepts
•Human-centered computing → HCI design and evaluation methods;

INTRODUCTION
The design, development and use of Creativity Support Tools (CSTs) has grown rapidly since Shneiderman [137] and Fischer [47] called attention to the potential for digital systems in Human-Computer Interaction to better support and augment human creativity. As documented by a recent large-scale survey [49], research into CSTs is characterized by a high degree of diversity in terms of use domains and development approaches, ranging from real-time crowd-sourced ideation support systems employing wall-size displays [5] to desktop-based interfaces that support the exploration and development of joinery in laser-cutting [164]. The survey also highlights, however, that HCI is characterized by remarkably few shared theories and analytical frameworks for examining and understanding the role of CSTs in creative processes, as well as a lack of shared basic methodologies for designing, developing, and, not least, evaluating CSTs. While this might be expected for a nascent field, it is problematic for *HCI researchers* who need sturdy, common ground to share, compare, and discuss findings, and, by extension, for *HCI practitioners* who lack approaches for designing and evaluating CSTs in a way that integrates key research findings and insights from examples of best practice. In order to increase current understanding of the role, nature, and further potential of CSTs, we focus in this paper on *how HCI researchers evaluate CSTs*. This is motivated by the fact that evaluation is a critical component in HCI research, in the sense that it is the process whereby researchers determine the potentials and limitations of interactive systems, which, in turn, sets the course for the development of future systems. By exploring the criteria by which HCI researchers evaluate the appropriateness and success of CSTs, we can gain a richer understanding of how this research community perceives creative practice in general, the role that CSTs can (come to) play in it, and the theoretical foundation on which the CSTs are built.

To explore in depth how HCI researchers evaluate CSTs, we present a comprehensive survey of CST contributions to the ACM Digital Library (1999–2018). We systematically sample and analyze 113 publications that present and evaluate CSTs. We examine the evaluation methodologies and criteria, as well as the underlying frameworks and theories in these sampled papers. The intended audience is HCI researchers and practitioners who design, develop, and study CSTs. As shown by [48, 50], this constitutes a growing part of the HCI community that engages with creativity practices.

The paper is structured as follows. We first frame the survey and analysis based on related work from HCI and creativity research, underlining how the two well-established research communities approach evaluation of CSTs. We then present our methodology and main findings from the survey before discussing some of the salient strengths and shortcomings of the current state of CST evaluation in HCI. Central insights from the survey are that there is a *variety of different goals, theories, and methods used in evaluating CSTs*, with no outstanding trends or dominance, highlighting both the diversity of research in this field and the challenges faced. We con-

DIS '20, July 6–10, 2020, Eindhoven, Netherlands.

ACM ISBN 978-1-4503-6974-9/20/07 ...$15.00.
http://dx.doi.org/10.1145/3357236.3395474

tribute *a set of recommendations for how to establish coherent CST evaluation strategies*, and we outline avenues for future research to develop better means of evaluating CSTs in HCI.

BACKGROUND AND RELATED WORK

Evaluation of research is an important, long-standing topic in many different research fields, including HCI and creativity research. As a basic overview, we revisit the history and most relevant recent development of both fields. We highlight the need for, and difficulty in, finding solutions to the problem of how to consider the validity of CST-focused HCI research in terms of assessing the actual appropriateness and success of a given CST. This serves to frame the third and final part of this related work section where we discuss the intersection of the two—evaluation methods for CSTs in HCI research.

Evaluation in Human-Computer Interaction

In the early years of HCI, such as the mid-1960s, most evaluation strategies focused on assessing the performance by measuring the computer's capabilities of a developed system [20]. Although the most influential research already concerned improving the user experience, such as most notably Engelbart's vision to augment human intellect [44], there was limited knowledge on how to measure said user experience besides measuring command line input or programming knowledge [10]. As research progressed and the field evolved, the combination of more advanced interfaces and methods borrowed from related disciplines such as psychology, sociology, and anthropology led to the appropriation and discovery of new evaluation techniques for HCI [10]. Those efforts culminated in two evaluation strategies for assessing HCI research: usability heuristics [116] and cognitive walkthroughs [153].

Nearly three decades later, those two methodologies are well-established and industry standard, but as Barkhuus and Rode observe in their survey, those methods seem not sufficient for research [10]. In fact, those evaluation strategies were discussed vigorously from the start [60], as the "discount usability" [116] title evoked debate whether it would oversimplify the complex issue of assessing how well systems met the user's needs and requirements. The debate continued and shifted over the years into territory that saw HCI researchers even questioning the usefulness of usability evaluation, as it could potentially cause harm to the innovative aspect of HCI [61]. While the suggestion was not to abandon evaluation altogether, but rather discuss new means of validating research output, it led to some more technology-focused sub-disciplines, suggesting focusing more on technical contribution rather than the proven validity based on user evaluation [129].

Most recent efforts picked up this thread to highlight that usability evaluation is just one aspect of evaluation, and while it might be critical when conducting HCI research, additional validation might be necessary. Special interest groups discussed how to shift from task-focused to experience-focused assessment [88] and the problem of evaluation beyond usability [128]. Several disciplines within or related to HCI research have discussed the challenges and complexities of evaluation, e.g., information visualization [15], prototyping toolkits [101], or sustainability [127].

All these historical developments and recent publications can be summarized to draw three insights: First, *evaluation strategies emerge and evolve over time*, often with engaged debates within the community, before any methodology becomes established—and even then, discussion does not subside, but rather shifts onto how to improve evaluation. Second, a successful strategy for discovering appropriate evaluation strategies is to *look into other disciplines* in an effort to adapt existing methods. Third, there is hardly a one-size-fits-all evaluation method; rather, a discipline should *build a toolbox comprising a variety of methods* that can be adapted, appropriated, or combined to fit the research that is to be evaluated.

Evaluation Methodologies in Creativity Research

In general creativity research, evaluation is often construed as various types of *assessment*; i.e., specifically designed methodologies for measuring if and to what extent someone or something can be considered creative. One premise for these approaches is an understanding of creativity informed by the 'standard definition' of creativity. This definition states that creativity requires both *originality* (sometimes referred to as novelty, surprise, or uniqueness) and *effectiveness* (or value, relevance, or appropriateness) [134]. Measuring creativity—how novel and useful someone or something is—has been the centrepiece of one of the most passionate discussions among creativity researchers since Guilford's [63] presidential address to the American Psychological Association (APA) in 1950, the starting point of modern creativity research. In the dawning age of this discipline psychometrics attracted much attention. Specifically, *divergent thinking* exemplified as number of ideas produced per time unit was considered a proxy of a person's creative potential. Based on cognitive psychology, Guilford [64] developed a number of such divergent thinking tests, including the Alternative Uses Tests (AUT), colloquially known as the 'brick test.' Around the same time, Torrance introduced his influential, eponymous tests of creative thinking (TTCT) with additional assessment parameters such as fluency, originality, flexibility, and elaboration of the creative ideas being generated [133]. Both the AUT and the TTCT, therefore, originate from cognitive psychology and the emergence of creativity research as a modern, specialist discipline.

Given that creativity is generally seen as "precisely the kind of problem which eludes explanation within one discipline," [53, p. 22], many conceptual models have been proposed in order to grasp even more of the complexity of creativity than cognitive psychological aspects. Chief among these models is Rhodes' [130] seminal *four p model* of creativity, which offers a simple, but useful distinction between person, product, process, and press. Differentiating in this way makes it possible to focus more clearly on which aspect of creativity that is being evaluated. As shown by Kaufman, Plucker, and Baer [87] in their comprehensive overview of creativity assessment, such conceptual scaffolding helps to convey a more operationable outline of evaluation techniques in creativity research. More detailed categorizations, on the other hand, remain controversial [133], e.g., Hocevar's [75] proposed quartered typology of creativity tests, which includes divergent thinking tests, attitude and interest inventories, personality inventories, and biographical inventories.

As Kaufman [86] has underlined, creativity measurement is the Achilles heel of the field. To some degree, this stems from the field's long-lived reliance on divergent thinking as the (cognitive) unit of measure. In response to this complexity, other types of evaluation have emerged, notably among them Amabile's [3, 72] influential Consensual Assessment Technique (CAT), which is based on combined assessments of experts within a specific creative domain. As opposed to said creativity assessment tests based on divergent thinking, the CAT does not rely on a theory of creativity, and this ensures its methodological validity as proven by several empirical studies [7]. Other evaluation methods that share some features with the CAT are *peer assessment*, *expert assessment*, and *self-assessment*. The latter can occasionally be the only viable option, however, its validity remains more contested [87].

Although many types of measurement and assessment have emerged in creativity research since the 1950s, it is important to stress a key difference between these types of creativity evaluation methodologies and the ones observed in HCI. The vast majority of the types of evaluation studies include *the researchers themselves* designing a specific (digital) tool to support creativity, after which the tool is evaluated in a concrete creative process involving specifically recruited users. Although a few, specialized types of creativity training studies may bear a little resemblance to this methodological approach [147], it is very rare to see creativity researchers evaluating the effect of CSTs that have been developed in their own research labs. This approach to evaluation of creativity, on the other hand, is quite customary in HCI.

Evaluation of Creativity Support Tools (CSTs)

Since the HCI research community has emphasized the importance of research concerning CSTs [137], the evaluation thereof becomes an urgent matter as well, as it is needed to determine and distinguish advancements in the field. In a call to action for researchers pointing out avenues for future development of CSTs, Shneiderman [138] also emphasized the shifting nature of evaluating tools in general and its critical nature for establishing knowledge of the usefulness of tools. The community of researchers engaged at the intersection of creativity and interaction design have repeatedly alluded to evaluation challenges as well (e.g., [1, 91, 93, 34]).

One contribution that stands out is the Creativity Support Index (CSI) [23, 22]. Conceptually based on the NASA Task Load Index [70], it allows for quick adaptation by researchers to evaluate CSTs, as it features an easy-to-implement general-purpose survey investigating the effectiveness of a newly developed tool. Its survey questions are based on creativity research theories and allow for quantifiable and comparable results. A follow-up publication explained the CSI in more detail, including an elaborate case study [27]. While we consider the CSI to be an established, useful tool, similar to what has been observed in HCI and creativity research there is need for developing multiple evaluation strategies of which there is an apparent lack. Such candidates could be based on the metrics of evaluation put forth by Kerne et al. [94], or a proposed tool for evaluating CSTs as informed by Warr and O'Neill [151].

Year	Publications
2018	[105, 142, 150, 30, 57, 29, 85, 115, 146, 66]
2017	[90, 5, 2, 152, 45, 164, 58, 79, 123, 120, 136, 99, 109, 36, 95, 139]
2016	[119, 59, 160, 78, 82, 107, 143, 76, 39, 98, 6, 108, 25, 35, 121]
2015	[51, 38, 159, 126, 140, 144, 97, 84, 112]
2014	[37, 156, 111, 96, 113, 163, 155, 52, 80, 94, 13, 83]
2013	[158, 81, 104, 14, 157, 162, 9, 114, 16, 62, 40]
2012	[118, 65, 55, 24, 135]
2011	[161, 103, 149, 141, 41, 89, 69, 56, 28]
2010	[12, 21, 154, 148, 132, 8, 11, 26, 77, 110, 106]
2009	[131, 42, 102, 145, 32]
2008	[92]
2007	[67, 18, 74, 46, 91]
2006	[68]
2005	[no publications]
2004	[17]
2003	[no publications]
2002	[125]

Table 1. List of surveyed papers, sorted by year.

We hope that the insights from our survey and the recommendations based on our discussion will inspire the community to strengthen their efforts in addressing this challenge and thus identify new and established ways to evaluate CSTs.

METHODOLOGY

To gain an understanding of current evaluation practices in CST research, we conducted a literature survey and an iterative coding exercise. In the following, we elaborate on our approach, first by describing our sampling methodology, as well as how we derived and iterated on our coding scheme.

Sampling

As starting point for our paper, we used the recent work of Frich et al. [50], who conducted an in-depth literature review of CSTs in HCI research. They conducted a keyword search for "creativity" or any occurrence of the word "creativity support tool" in the ACM Digital Library, and then reduced the sample size by selecting the papers with above average citations per year (0.669). For papers after 2016, the citation count was found too low and too volatile to be a reliable metric. This is why the average download count (per year) in the ACM Digital Library was used as cut-off measurement for 2016-2018, resulting in a final corpus of 143 papers [50]. We took their corpus and selected all papers in this sample that contained an evaluation of a CST, leading to a final count of 113 papers for our analysis. The same sampling was used to survey 2019 publications, but no papers above the average download count provided an evaluation of a CST.

Since our survey's goal is to advance the evaluation strategies for new CST development in HCI research, this sample is an optimal starting point for our inquiry. We consciously decided against adding keywords such as "assessment" or "evaluation" to the sampling method, as this might have skewed the sample towards papers focusing on the discussion of evaluation within HCI. This would defeat the purpose of a survey aiming to summarize the existing de-facto standard practices in evaluating CSTs in HCI research. We are aware of past and ongoing discussions of evaluating HCI research and will address this topic in the Discussion. Similarly, we did not manually include papers or CSTs from other sources that did not make the cut, as that would bias the sample. Our survey aims to provide an objective description of how the most prominent CSTs in

HCI research have been evaluated, as those publications are set up to shape the future of CST research. In doing so, we hope that the analysis of this sample will provide new insights that can inform the future of evaluating such tools. A list of all sampled papers by year can be found in Table 1.

Analysis

The development of the coding scheme for our survey took multiple iterations and highlighted the complexity and challenges of the evaluation question in CST research. As a first, naïve approach we added two open-ended code fields, "method" and "criterion", and three researchers independently started coding the publications. We quickly realized that the variety of different evaluation strategies employed in CST research led to an equally varying degree of descriptions for the methods and criteria, forming no basis for any comparative means required in a comprehensive literature survey. We also noticed that even seemingly straightforward codes such as "number of participants" caused issues, as some evaluations comprised multiple steps with largely varying evaluator numbers, several iterations of one CST with different evaluation techniques, or even multiple CSTs in one publication.

To resolve the problem of varying iterations in CST evaluations, we added a code listing the "number of iterations", as well as a checkbox for whether or not the paper included a preliminary or formative study that gained in-depth insights into the target audience prior to developing the CST, eliciting requirements. We also added binary codes for the evaluation using "teams", "facilitators", and whether or not they recruited "experts" (whereas we define "experts" here as evaluators who match their described target audience). The numeric code for number of participants was then streamlined into an integer, listing one number for the main evaluation (if there were several) and the total number of participants (e.g., number of groups times number of participants). Similarly, we added a code for the duration of the study, which ranged from a few minutes to several years listed as total number of hours, rounded up to the nearest integer or approximated to an even number (e.g., 1,000 hours for a six-week study).

Addressing the issue of the open code fields, with which we aimed to investigate details about the methodology, proved to be more difficult. We turned our attention to recent publications discussing the evaluation question in HCI research at large to gain further insights [122, 101, 127]. We sought to arrive at an overview of the different prevalent categories in CST evaluation similar to Ledo et al. [101] or Pettersson et al. [122], but upon re-iterating on our coding scheme it became apparent we needed a more suitable starting point. Inspired by the "evaluation recipe" recommended for sustainability research, we derived new coding categories based on the five "ingredients" presented in the paper [127]: goal, mechanisms, metrics, methods, and scope. In several iterations of independent coding and comparison by three researchers, we adjusted those five points until they proved to be a fit for our analysis.

Goal was kept as a category with the restriction of it being as concise as possible, and close to the author's description, i.e., derived from a line in the title or abstract rather than our interpretation of what the authors wanted to achieve. We initially aimed to survey the goal of the evaluation; however, many papers did not state a specific goal of the evaluation but rather the goal of the CST, which is why the goal code refers to the former. Identifying the relevant mechanisms proved to be difficult, which is why we went back to the original concept as proposed by Dix [43]: "If you understand the mechanism, the steps and phases of the interaction, you can choose finer measurements and use these". To this end, we substituted this code with listing the "theory or background" that the paper in our survey listed. For the metrics category we observed a similar varying degree as in our initial "criterion" code; to solve the confusion and arrive at a more streamlined code, we decided to limit this category to "creativity metrics", and only those explicitly mentioned by researchers in the paper.

To analyze the coding results of the "goal" category in our survey, we started by conducting an inductive card sorting activity, discussing potential themes, groups, or dimensions emerging from the goals. All 113 goals were written on sticky notes and put on a whiteboard, and over time after several rounds of reorganization, discussion, and reflection, two dimensions emerged: *domain specificity* and *level of intrusion*. Domain specificity refers to whether the goal seems to be oriented towards rather generic processes (e.g., "multi-idea design support" [67]), or describes a narrow target audience or application context (e.g., "automatic assessment of song mashability" [37]). Level of intrusion describes how much the CST potentially changes the creative process itself, from tools that only add or offer technical support (e.g., "social media content summarization" [98]) to those that directly influence and alter the activity (e.g., "manipulate video conference facial expression" [113]). For an initial calibration of the various degrees of intrusion, we compiled a list of 30 verbs used in the goals (from "support" and "assist" to "edit" and "enhance").

The fourth category, methods, was already in our initial exercise, but needed further specification to be applicable in our coding exercise. We split the code into three smaller codes: "generic methods", which are well-established evaluation techniques that do not require a reference (i.e., interview, workshop, or questionnaire). Contrary to that, "specific methods" describes evaluation techniques that either require a link to a publication presenting an explanation of the methodology or even a newly invented evaluation technique by the researchers themselves. Lastly, we added a code for "analysis", with the restriction of only coding for explicit mentions in the paper.

Once the coding was finished for all 113 papers, one researcher went over all the codes to sanitize the input from five researchers from varying backgrounds. The "methods" field saw several repetitions, which allowed it to further split into nine binary codes (for methods that were mentioned at least eight times) and an open-ended field for "other methods". A final overview of all the codes can be found in Table 2.

RESULTS AND OBSERVATIONS

In this section, we describe the results from our coding exercise as well as outline prescriptive findings. For our interpretation of the insights considering the bigger picture of CST research and recommendation for future evaluation strategies, see the Discussion section following thereafter. An overview of the

Type	Codes (O = open-ended, N = numeric, B = binary)
O	Goal
O	Creativity metrics
O	Theory/background
B	Formative pre-CST study
N	Design cycles
N	Duration (in hours)
N	Number of participants
B	Teams
B	Facilitator
B	Expert evaluators
B	Method: interview
B	Method: survey/questionnaire
B	Method: observation
B	Method: logs/measurement
B	Method: video recordings
B	Method: Mechanical Turk
B	Method: field study
B	Method: usability study
B	Method: workshop
O	Method: other
O	Analysis

Table 2. Final codes used in our analysis.

results for binary codes can be found in the Appendix on page 10 of this paper.

Goals of the Developed CSTs

Upon organizing all goals along the two dimensions of domain specificity and level of intrusion, we observed that the spectrum is only seemingly continuous, but we noticed the formation of groups along discrete levels. Therefore, we decided to split the dimensions up into three levels each: *low, medium*, and *high domain specificity* and *level of intrusion*, respectively. The final distribution can be seen in Figure 1. One observation that matched our expectation is that there is a variety of generic tools that support general design processes such as ideation, but also CSTs that are built for very specific purposes such as knitting [132]. This also aligns with the findings of Frich et al. [50], who observe that there are special-purpose CSTs as well as tools that can be appropriated to more generic tasks, such as general idea generation sessions. For the intrusiveness of the CSTs, we noted the opposite result. While in the beginning clusters of notes in the corners seemed to form, towards the end of the card sorting and after several rounds of discussion and double-checking, we noticed that our assessment shifted towards seeing most goals being rather in the medium section of the spectrum (see Figure 1). This might be due to the subjectivity and interpretability of the verbs we analyzed, but it also matches our observation that most goals were vague and generic. For the purpose of evaluation, this is an interesting finding that we discuss later in this paper.

	Low < domain specificity > high			∑
High	10	9	18	37
Level of intrusion	17	16	16	49
Low	12	5	10	27
∑	39	30	44	

Figure 1. Breakdown of the card sorting results for the "goal" code.

We want to reiterate that the goal phrase that was noted on the sticky notes in our card sorting exercise is an abbreviated form of the goal stated by the authors in the title or abstract of the paper as it was understood by the researchers. As such, there is a considerable level of subjectivity and room for interpretation in this activity. Also, our survey is based on a sample of papers—above the cut-off point for average citations or downloads. Our results should thus not be taken as a quantitative measure. Nevertheless, we found that the outcome aligned with our subjective perceptions during the coding as well as with our experience from working in the field of CST development.

Theory/Background Sections and Creativity Metrics

For this category, we only coded papers that had an explicit section on non-trivial theory or elaborating in depth about background literature that strongly influenced the design of the CST—and outside the related work section. We ended up with 52 codes, such as the "socio-cognitive model of group brainstorming" [149] or the "paths model" [163]. Unlike for the goal category, a card sorting exercise is unlikely to yield surprising insights, and the particular choice of how we selected the code (by considering explicit mentions of a theory) explains why more than half of our sample draws a blank. It is worth noting that of the 52 codes, most of them can be attributed to creativity research in some way; either by being creativity models themselves (e.g., "little-c creativity" [82, 105]), or closely related (e.g., "storytelling patterns" [97]).

This is particularly interesting when compared to the results of our "creativity metrics" code, for which we saw only 26 papers directly quoting creativity metrics (e.g., fluency, originality, novelty, flexibility). Of those 26 papers, 17 were also coded as having a dedicated theory or background section, although those do not necessarily match, but might build on different theoretical foundations. One noteworthy observation is that out of our reviewed 113 papers, 61 have either some theoretical backing and/or a clear link to creativity metrics, but 52 engage with neither. Since most, but not all, theories can be attributed to some level of creativity research or related insights, we can conclude that about *half of our sample engages with creativity at least for inspirational intents and purposes*, and roughly *a quarter relies on explicit creativity research metrics to assess the creative impact of the CST* on the creative process.

Formative Studies and Iterations

Looking at the number of publications that were coded for having a formative study, we observe that 36 papers (approx. 32 %) employed research methodology to derive guidelines for the design before developing the CST, with no historic

trend apparent when plotted over the years of our sample. There were outliers in 2009 and 2011 that saw 60 % and 67 %, respectively, of the papers conducting a formative study prior to the tool development. However, in 2009 there is a notable drop in sample size (five papers), and 2011 is braced by two years of rather low figures (18 % and 20 %, respectively).

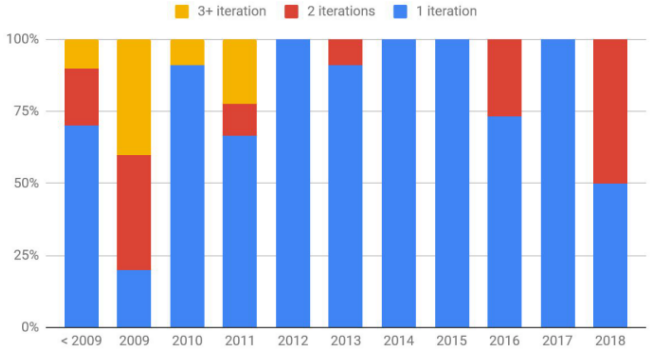


Figure 2. Number of design iterations of the CST.

We excluded the formative study aspect from the coding for number of iterations, i.e., we only counted iterations of the prototype and not iterations of the study design as such (though those might correlate). Keeping that in mind, it is interesting to see that while the formative study code did not yield any historical trend, the number of iterations does. As can be observed in Figure 2, in the early years of CST research papers seemed to have more iterations, as there was no paper since 2011 in our sample that saw three or more iterations (and a total of six in 2011 and before). In 2018 we saw the highest number of papers (five out of ten) having more than one iteration in almost a decade, raising the question whether this will be an outlier or a historical shift.

Evaluation Details: Duration and Participants

The duration of the evaluation sessions varied considerably, between a few minutes for short brainstorm sessions up to several years for long-term deployments, with the majority of evaluations being short-term (46 papers or 41 % saw an evaluation lasting less than or up to one hour); only 21 evaluations (18 %) lasted for more than a day. We saw no signs of any correlation between a certain study type and long-term studies, as the studies that lasted for more than a week had a similar distribution than the overall corpus. This may seem surprising, but looking at the data reveals that the prevalent evaluation methodology (interviewing or observing participants as well as measuring tool usage) remains the same, whether the exposure to the tool was limited to a few minutes in an ideation session, or lasted over several weeks in a long-term deployment.

The codes for evaluation methods that use facilitators (16 in total, 14 %) and specifically recruit domain experts (27, 24 %) experience several spikes suggesting a high standard deviation. Overall, however, there is no evidence to make any statistical claim for a trend in CST evaluation noticeable, and is, especially for the facilitators, subject to a low sample size. This is different for the number of evaluations that uses teams (45, 40 %), which indicates a historical trend that is declining,

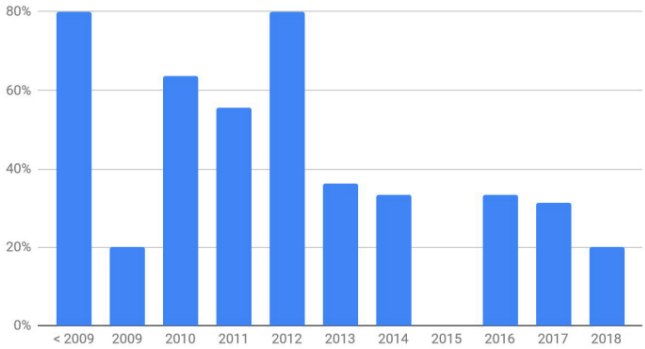


Figure 3. Percentage of evaluations using teams as evaluators.

from 80 % in papers before 2009 to 20 % in papers published in 2018 (see Figure 3).

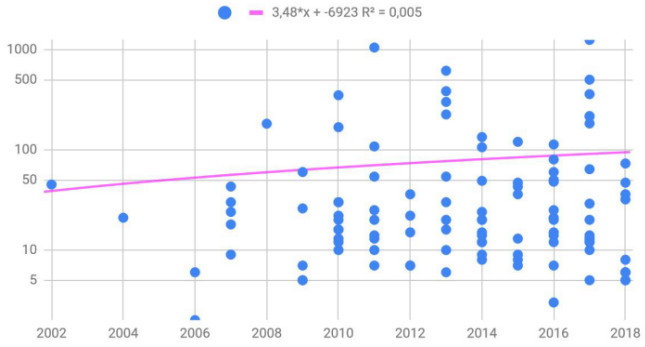


Figure 4. Number of participants in the evaluation.

The analysis of the participant numbers reveals a trend of growing average number of participants for the evaluation of CSTs, as shown in Figure 4. The trendline is only added (programmatically) for visual aid and matches our subjective perception from look at the data. Two important points to mention are that it appears less strong than it is due to the logarithmic scale of the vertical axis, but that its coefficient of determination is extremely low at $R^2 = 0.005$ due to the clearly visible spread of the data points. Even discounting Mechanical Turk studies that have a higher number of participants (average of 285 in our sample), this does not change the picture; especially since one of the outliers with over 1000 participants is from 2011 [161]. We also excluded two data points, which made the chart unreadable, even with its logarithmic scaling, namely two studies from 2016 that analyzed data from 173,053 [35] and 23,092 [108] participants, respectively. These studies would have caused a drastic shift in the trend from 2016 onwards.

Methods and Analysis

The three most commonly used methods were observations (38 occurrences), surveys/questionnaire (37), and interviews (33). Since many papers were coded for multiple methods, there was some overlap. Nonetheless, there were only 32 papers that used neither of those three methods as per our analysis. With considerable distance, the next categories to follow are workshop (18), logs/measurement (18), usability study (15), and video recordings (14). Last but not least, we coded eight

papers for the use of Mechanical Turk as well as eight papers conducting a field study. To our surprise, only three of the total number of surveyed papers made use of the Creativity Support Index (CSI) [100]. While the number of field studies might be considered low to some readers familiar to the realm of HCI, a field with a long-standing tradition to value field research, we reiterate that we used a deductive open coding technique from which we extracted the nine code fields. Other papers in our sample might be considered field studies as well, but during the coding activity other evaluation methods might have been appeared to be more important to the coder.

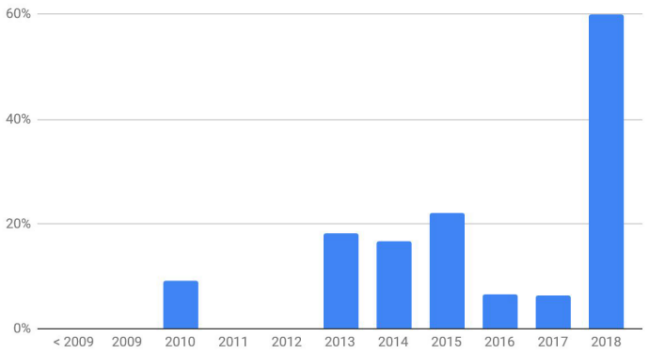


Figure 5. Breakdown of evaluations that were coded as usability study.

We additionally looked into the chronological trend of different methods to see if there were any tendencies in which methodologies have been applied historically. The full set of graphs representing this data take up too much space in this paper and provide limited insight, but will be made publicly available on an interactive website. The only trend that was significant enough to be worth mentioning is the apparent increase of usability evaluations for CSTs in recent years (see Figure 5). Of the papers in our sample published in 2018, 60 % employed a usability study in their evaluation, while only one paper prior to 2013 focused on usability evaluation methodology [12].

Our results for coding the mentioned "analysis" mostly comprised statistical means of analysis, such as ANOVA/MANOVA (mentioned twelve times) or t-test (eight times). Overall, we counted 53 occurrences of explicit means of analysis; however, as a significant portion of our sample employs qualitative methodology (62 papers), this does not come as a surprise to us. Qualitative data is often analyzed by theme elicitation through coding or close reading to identify patterns in the data, and is applied without explicit mention, hence its lack of appearance in our coding.

DISCUSSION

We surveyed 113 papers that presented the evaluation of CSTs and described the results of our coding in the previous section. While the historical account of evaluation is interesting to us as researchers to gain insight of the big picture, we are equally eager to identify avenues for future research on how CSTs can be evaluated. In the following, we connect the descriptive observations from our survey to insights from scientific literature and, also in light of our own observations from our research and the extensive discussions surrounding our literature analysis, derive recommendations for the research community.

Many Studies are Unclear about the Goal of CSTs and the Goal of Evaluations

One of the difficulties in analyzing the goals of CSTs is that they are vaguely defined in many of the surveyed papers. In particular, this made it hard to assess the level of intrusion, i.e., the extent to which the CST influences or changes the creative process. While we kept the goal description in our coding intentionally short, all researchers were familiar with the entire corpus and during the card sorting discussed the goals of the paper—only to discover levels of disagreement about what the paper's goal, and therefore the goal of the CST itself. This can contribute to the difficulty of evaluating a CST; similar to how Remy et al. [127] argued that the goal of an evaluation needs to be more narrowly defined than just "sustainability", "creativity" is an equivalently broad term that can obfuscate the view upon the real (and realistic) goal of what the CST seeks to achieve. This is critical for the evaluation of CSTs, as it is not just about clearly defining the goal of what the tool wants to achieve, but it can also help to distinguish the difference between the research purpose of the CST and the potential real-world use case, and which of those we want to evaluate. Here we draw on Ledo et al. [101], who in their analysis of the evaluation of tool kits point out that "researchers often have quite different goals from commercial toolkit developers," which is analogous to CST research.

Moreover, the goal of the CST and the goal of the evaluation of the CST can be confused and conflated. For example, if a tool developed by researchers is intended to improve the creative output for a specific task in a certain domain, the goal of the evaluation might vary significantly: one could evaluate the productivity of the process affected by the CST, the creativity of the outcome, or the usability of the tool itself. In fact, we found all those three different aspects to be present in our survey, though often not clearly spelled out and distinguished. Based on our discussions around the difficulty of coding the corpus surveyed in this paper, we hypothesize that being more specific about the goal of your CST can help sharpen the goal of the evaluation of the CST as well.

As we tried to formulate the goal for each paper ourselves preceding the card sorting activity, we realized that this is not always a simple task. Especially, if the goal truly is to support a wide range of creative processes and being as non-intrusive as possible, the goal may be vague. In that case, it can help to investigate the mechanisms affected by the tool, as suggested for the second ingredient of the evaluation recipe [127]. Mechanisms in HCI research evaluation describe the "details of what goes on, whether in terms of user actions, perception, cognition, or social interactions" [43], and have been discussed in various other disciplines such as philosophy of science [124] or social science [71]. When evaluating CSTs, a useful starting point to investigate the mechanisms that concern the tool use and how it affects creative practice can be to consider creativity research or other theoretical foundations that describe the interaction between the tool and its appropriation in the creativity process, e.g., as part of the background

theory or creativity metrics that we coded for in our survey and found in roughly half the papers (61 out of 113). We believe that thinking about mechanisms of a CST in projected use cases can help to sharpen and clarify the goal, thus further strengthening the focus of the evaluation.

Evaluations of CSTs Lack Theoretical Grounding

In line with the discussion on the scarcity of creativity metrics in the surveyed papers, we also see that less than half of the papers build on identifiable theoretical foundations for addressing creativity (52 out of 113). This is in line with the wider field of creativity research within HCI [49], but nonetheless unfortunate for several reasons. For the individual study, it makes it harder for readers to position the work within the field, and to assess which aspects of creativity the CST addresses. Seen in a wider perspective, it makes it difficult to compare research insights and outcomes across cases and studies, and by extension it limits the collective efforts to build up a body of knowledge about CSTs.

To be clear, we are not proposing that authors should be constrained to fixed definitions or sets of theories for studying and developing CSTs, especially given the multi-faceted nature of creativity and the plethora of domains and use practices in which CSTs occur. Rather, we argue for the need for greater clarity in terms of which theories underpin specific studies, and how they address certain aspects of creativity. This is, in our view, especially pertinent exactly because creativity is a concept that can be understood and approached from a range of perspectives. Moreover, a clearer theoretical foundation for a CST study makes it easier to establish criteria by which to evaluate the CST. For example, if a study is theoretically grounded in theories on combinatorial creativity, this can inform criteria for assessing the ways in which the CST supports users in exploring, combining, and synthesizing ideas.

Evaluations of CSTs Lack Expert Participants

We found that only 24 % (27) of papers in our sample recruit domain expert participants to evaluate their CST intended for that particular area or task, as mentioned in the Results section. This is in our view a surprisingly low number, and we find good reason to argue that evaluations of tools intended for experts or with a specific user group in mind ought to include experts or users from these specific groups as participants. Otherwise, the evaluation results can be questionable, since novices and experts may have very different perspectives and frames of reference for assessing a tool. It may require more effort to find and recruit experts and participants from a specific user group than e.g. recruiting students with little or no experience in the intended use domain. However, the results from these evaluations will be more likely to offer ecologically valid evaluations of the CST, which in turn may not only enrich our research community's understanding of CSTs but also help developers create better tools for these communities. As a community of researchers interested in supporting creative practice, we find this to be a serious concern.

This being said, there can be good reasons for not recruiting domain experts as participants in some phases of CST evaluations, particularly in when it comes to domain specific CSTs with high levels of intrusion. We also want to point out that this discussion only pertains to tools that are aimed at specific target audiences and/or expert practitioners. If, for example, the goal is to increase creativity among novices, or if the tool is aimed at the general public, there might not be a need for expert evaluation.

If expert participants offer insight and feedback on the usefulness or effectiveness of a CST during evaluation, i.e., through interviews or workshops, the data collected can be used to improve the CST in further iterations. A complication to this, however, is that expert participants may be accustomed to working within a particular ecosystem of tools, particularly in the case of evaluations of domain-specific CSTs that intrude in or disrupt existing workflows. This means that the expert is less flexible or receptive towards the idea of changing their workflow with the incorporation of a new CST.

Short-Term Controlled Evaluations of CSTs Are Prioritised over Longitudinal In-Situ Studies

A clear and noteworthy finding from the survey is that there are few long-term studies that evaluate the use and impact of CSTs over the course of time. Most of the studies focus on the immediate use of one system, often developed by the researchers themselves. On the one hand, this is unsurprising in that the majority of studies in the wider field of HCI also focus on immediate or short-term studies, and in that it is, in most cases, more time-consuming to carry out longitudinal studies. On the other hand, it is clearly problematic if we wish to get an understanding of the role and nature of CSTs in real-life creative practices. It can take time for creative professionals to master tools, determine in which situations they are beneficial, and integrate them into their workflows. Longitudinal studies are required to gain insight into these processes, but such studies are to a large extent lacking from current CST research. As a result, there is a prominent and problematic gap in our collective body of knowledge. This is compounded by the prior issue of the relative lack of evaluations that include expert practitioners, meaning that we in many cases only see short-term evaluations with novice users, often in controlled setups, rather than long-term evaluations of tools used by experts in real-life use situations. We must stress that there are exceptions to this trend. One such example, which might inspire other CST researchers, is the INJECT project, a CST to augment journalists' creativity when researching and developing news angles. The INJECT services has been presented to the research community throughout its development and has been deployed and studied in real-life use with journalists at three newspapers over the course of two months [105]. This echoes developments in other HCI research communities such as persuasive technology [19], calling for longitudinal studies as well, to better understand the long-term impact of research deployments.

We consider this problematic, but if we are to frame it in a positive light, this is a highly promising avenue for future research, and we hope that this survey can spark discussions in the CST research community about undertaking longitudinal studies. We acknowledge that longitudinal studies can be at odds with

the nature of research projects that often have shorter timespans, and in which there can be pressure to publish results within the project's funding period. Therefore, in addition to a call to action for researchers to consider longitudinal projects, it is also upon our research community to reconsider publication strategies and venues. A starting point could be to take inspiration from the recently emerging discussion about preregistration [31]—for instance, an early submission can present the CST itself and the preregistered strategy for evaluation, while a subsequent submission can report on the results of the study. Preregistration is still rare in HCI and interaction design research, but could offer a complimentary step towards increasing the validity of evaluations.

Many CST Evaluations Focus on Usability, not Creativity

There is a clear trend among the most recent papers in the survey to include usability testing as a means for evaluating CSTs. For example, this goes for 60 % of the papers from 2018. In comparison, only three of the total number of surveyed papers made use of the Creativity Support Index (CSI) [100], the most well-established standardized means of evaluating the creativity-supporting aspects of digital tools. This indicates the potential, or even need, for more standardized forms of CST evaluation, such as those developed in usability engineering. Moreover, it raises the question of the relationship between evaluating usability and creativity. The development of usability principles and evaluation methods has been highly influential in HCI and has furthered the development of tools to automate and support routine work. However, we do not know whether usability principles and evaluation criteria can help us assess how well a tool supports creative work. For instance, traditional usability methods emphasize aspects such as efficiency, precision, error prevention, and adherence to standards [117, 54], but do not address core dynamics of creative work, such as exploration, experimentation, and deliberate transgression of standards [33, 109]. We support the growing endeavors to ensure that CSTs are usable, but we raise this point to emphasize that the usability of a system does not necessarily correlate with its creativity—supporting or—augmenting properties. Based on the seemingly contradictory principles, e.g., adherence to standards vs. transgression of standards, it is possible that they may even run counter to them in some cases; however, we are not in a position to determine if and how this is the case based on the survey at hand.

CST Research Lacks a Toolbox for Evaluation

We started this paper by a historical overview of early usability and HCI methods in the 80s and how they were referred to as "discount usability" [116]. While discount denotes something of a lesser quality, it also highlights widespread access and low "cost". Considering the current state of evaluations of CSTs, the field could perhaps use a discount CST evaluation to really advance the development of new tools for digital support. A bid for this might be the already available and thoroughly tested CSI [27], as some of the creativity metrics surveyed in this paper [139, 82, 45, 123] could be at least partly covered using the CSI. As indicated by this review, the CSI is rarely used at the at the expense of other evaluations. One reason might be that similar to the quest for evaluation methods in HCI in general, the use of conceptual models and psychometric surveys have fallen out of vogue, in favor of mixed-methods evaluations and qualitative methods as highlighted in our study.

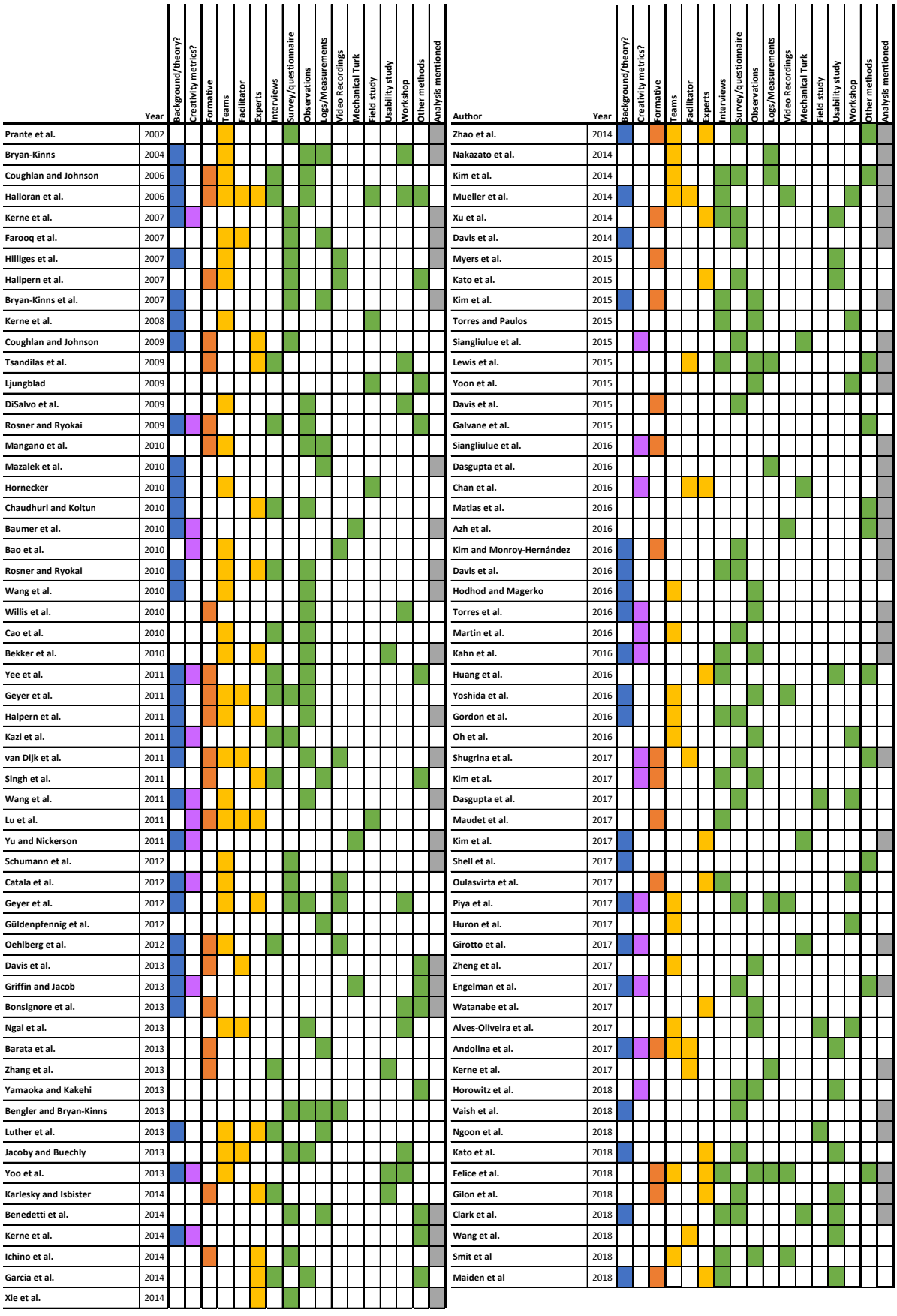
Maybe a suitable question is whether evaluations of CSTs necessarily require some sort of creativity metric? In the case of Davis et al. [40], the system clearly supports the user in adhering to the domain-specific requirements of machinima. Consequently, the tool supports creativity by enhancing creativity-relevant skills, which is on out of three essential foundations for creativity in Amabile's Componential Theory of Creativity [4]. Furthermore, if we consider the work by Ngoon et al. on improving critique [115], the improvements to creativity might eventually be realized in the long-term, as the authors hint at themselves. In this case, evaluating the critique delivered might serve as a proxy for potential long-term improvements. On the other hand, if we consider evaluations of CSTs for more generic domains such as brainstorming, it might be easier to spot or identify possible avenues for the evaluations of those CSTs, as they naturally come closer to the canonical ways of evaluating creativity, e.g., by using fluency and elaboration as metrics. The development of a "one-size-fits-all" evaluation of CSTs might not be appropriate; something that was already pointed out in 2005 in the NSF report's [73] chapter on evaluating CSTs. Even so, we believe the best way to solve the challenge of evaluating CSTs is following in the footsteps of *usability evaluation*. That is, develop a *toolbox* (i.e., multiple options) *of evaluation techniques* and tools that are *accessible* and *feasible* (i.e., "discount"), and establish *community consensus* on their usefulness through a *proven track record* of CST evaluations.

CONCLUSION

We have reported on the results of a survey of 113 papers from HCI research that design and evaluate CSTs. We have presented the results of our extensive coding analysis and, based on these findings, discussed our insights for future evaluation of CSTs. We have centered our discussion around six major points that researchers developing CSTs should consider in their evaluation: 1) Clearly define the goal of the CST; 2) link to theory to further the understanding of the CST's use and how to evaluate it; 3) recruit domain experts, if applicable and feasible; 4) consider longitudinal, in-situ studies; 5) distinguish and decide whether to evaluate usability or creativity; and 6) as a community, help develop a toolbox for CST evaluation. Our survey aims to provide researchers looking for an appropriate evaluation method for their own CST with inspiration and guidance from past examples of research. First and foremost, however, we see this work as a starting point for the community to arrive at an *established toolbox for evaluation methods*; be that by creating and establishing new techniques, or by reaffirming and consolidating existing ones.

ACKNOWLEDGEMENTS

This research has been funded by The Velux Foundations grant: Digital Tools in Collaborative Creativity (grant no. 00013140), and the Aarhus University Research Foundation grant: Creative Tools.


Author
Year
Background/theory?
Creativity metrics?
Formative
Teams
Facilitator
Experts
Interviews
Survey/questionnaire
Observations
Logs/Measurements
Video Recordings
Mechanical Turk
Field study
Usability study
Workshop
Other methods
Analysis mentioned
Prante et al. 2002
Bryan-Kinns 2004
Coughlan and Johnson 2006
Halloran et al. 2006
Kerne et al. 2007
Farooq et al. 2007
Hilliges et al. 2007
Hailpern et al. 2007
Bryan-Kinns et al. 2007
Kerne et al. 2008
Coughlan and Johnson 2009
Tsandilas et al. 2009
Ljungblad 2009
DiSalvo et al. 2009
Rosner and Ryokai 2009
Mangano et al. 2010
Mazalek et al. 2010
Hornecker 2010
Chaudhuri and Koltun 2010
Baumer et al. 2010
Bao et al. 2010
Rosner and Ryokai 2010
Wang et al. 2010
Willis et al. 2010
Cao et al. 2010
Bekker et al. 2010
Yee et al. 2011
Geyer et al. 2011
Halpern et al. 2011
Kazi et al. 2011
van Dijk et al. 2011
Singh et al. 2011
Wang et al. 2011
Lu et al. 2011
Yu and Nickerson 2011
Schumann et al. 2012
Catala et al. 2012
Geyer et al. 2012
Güldenpfennig et al. 2012
Oehlberg et al. 2012
Davis et al. 2013
Griffin and Jacob 2013
Bonsignore et al. 2013
Ngai et al. 2013
Barata et al. 2013
Zhang et al. 2013
Yamaoka and Kakehi 2013
Bengler and Bryan-Kinns 2013
Luther et al. 2013
Jacoby and Buechly 2013
Yoo et al. 2013
Karlesky and Isbister 2014
Benedetti et al. 2014
Kerne et al. 2014
Ichino et al. 2014
Garcia et al. 2014
Xie et al. 2014
Zhao et al. 2014
Nakazato et al. 2014
Kim et al. 2014
Mueller et al. 2014
Xu et al. 2014
Davis et al. 2014
Myers et al. 2015
Kato et al. 2015
Kim et al. 2015
Torres and Paulos 2015
Siangliulue et al. 2015
Lewis et al. 2015
Yoon et al. 2015
Davis et al. 2015
Galvane et al. 2015
Siangliulue et al. 2016
Dasgupta et al. 2016
Chan et al. 2016
Matias et al. 2016
Azh et al. 2016
Kim and Monroy-Hernández 2016
Davis et al. 2016
Hodhod and Magerko 2016
Torres et al. 2016
Martin et al. 2016
Kahn et al. 2016
Huang et al. 2016
Yoshida et al. 2016
Gordon et al. 2016
Oh et al. 2016
Shugrina et al. 2017
Kim et al. 2017
Dasgupta et al. 2017
Maudet et al. 2017
Kim et al. 2017
Shell et al. 2017
Oulasvirta et al. 2017
Piya et al. 2017
Huron et al. 2017
Girotto et al. 2017
Zheng et al. 2017
Engelman et al. 2017
Watanabe et al. 2017
Alves-Oliveira et al. 2017
Andolina et al. 2017
Kerne et al. 2017
Horowitz et al. 2018
Vaish et al. 2018
Ngoon et al. 2018
Kato et al. 2018
Felice et al. 2018
Gilon et al. 2018
Clark et al. 2018
Wang et al. 2018
Smit et al 2018
Maiden et al 2018

REFERENCES

[1] Piotr D. Adamczyk, Kevin Hamilton, Michael B. Twidale, and Brian P. Bailey. 2007. HCI and new media arts: methodology and evaluation. In *Extended Abstracts Proceedings of the 2007 Conference on Human Factors in Computing Systems, CHI 2007, San Jose, California, USA, April 28 - May 3, 2007*. 2813–2816. DOI: http://dx.doi.org/10.1145/1240866.1241084

[2] PatrÃ■cia Alves-Oliveira, PatrÃ■cia Arriaga, Ana Paiva, and Guy Hoffman. 2017. YOLO, a Robot for Creativity: A Co-Design Study with Children. In *Proceedings of the 2017 Conference on Interaction Design and Children - IDC '17*. ACM Press, Stanford, California, USA, 423–429. DOI: http://dx.doi.org/10.1145/3078072.3084304

[3] Teresa M Amabile. 1983. The social psychology of creativity: A componential conceptualization. *Journal of personality and social psychology* 45, 2 (1983), 357.

[4] Teresa M Amabile. 1996. *Creativity in context: Update to "The social psychology of creativity"* (2 ed.). Westview Press, Boulder, CO.

[5] Salvatore Andolina, Hendrik Schneider, Joel Chan, Khalil Klouche, Giulio Jacucci, and Steven Dow. 2017. Crowdboard: Augmenting In-Person Idea Generation with Real-Time Crowds. In *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition - C&C '17*. ACM Press, Singapore, Singapore, 106–118. DOI:http://dx.doi.org/10.1145/3059454.3059477

[6] Maryam Azh, Shengdong Zhao, and Sriram Subramanian. 2016. Investigating Expressive Tactile Interaction Design in Artistic Graphical Representations. *ACM Trans. Comput.-Hum. Interact.* 23, 5 (Oct. 2016), 32:1–32:47. DOI: http://dx.doi.org/10.1145/2957756

[7] John Baer and Sharon S McKool. 2009. *Assessing Creativity Using the Consensual Assessment Technique*. Hershey, 65âĂŞ77.

[8] Patti Bao, Elizabeth Gerber, Darren Gergle, and David Hoffman. 2010. Momentum: Getting and Staying on Topic During a Brainstorm. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1233–1236. DOI: http://dx.doi.org/10.1145/1753326.1753511

[9] Gabriel Barata, Sandra Gama, Manuel J. Fonseca, and Daniel GonÃğalves. 2013. Improving Student Creativity with Gamification and Virtual Worlds. In *Proceedings of the First International Conference on Gameful Design, Research, and Applications*. ACM, New York, NY, USA, 95–98. DOI: http://dx.doi.org/10.1145/2583008.2583023

[10] Louise Barkhuus and Jennifer A. Rode. 2007. From Mice to Men - 24 Years of Evaluation in CHI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, USA. DOI: http://dx.doi.org/10.1145/1240624.2180963

[11] Eric P.S. Baumer, Jordan Sinclair, and Bill Tomlinson. 2010. America is Like Metamucil: Fostering Critical and Creative Thinking About Metaphor in Political Blogs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1437–1446. DOI: http://dx.doi.org/10.1145/1753326.1753541

[12] Tilde Bekker, Janienke Sturm, and Berry Eggen. 2010. Designing Playful Interactions for Social Interaction and Physical Play. *Personal Ubiquitous Comput.* 14, 5 (July 2010), 385–396. DOI: http://dx.doi.org/10.1007/s00779-009-0264-1

[13] Luca Benedetti, Holger WinnemÃűller, Massimiliano Corsini, and Roberto Scopigno. 2014. Painting with Bob: Assisted Creativity for Novices. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*. ACM, New York, NY, USA, 419–428. DOI: http://dx.doi.org/10.1145/2642918.2647415

[14] Ben Bengler and Nick Bryan-Kinns. 2013. Designing Collaborative Musical Experiences for Broad Audiences. In *Proceedings of the 9th ACM Conference on Creativity & Cognition*. ACM, New York, NY, USA, 234–242. DOI: http://dx.doi.org/10.1145/2466627.2466633

[15] Enrico Bertini, Catherine Plaisant, and Giuseppe Santucci. 2007. BELIV'06: Beyond Time and Errors; Novel Evaluation Methods for Information Visualization. *interactions* 14, 3 (May 2007), 59–60. DOI:http://dx.doi.org/10.1145/1242421.1242460

[16] Elizabeth Bonsignore, Alexander J. Quinn, Allison Druin, and Benjamin B. Bederson. 2013. Sharing Stories âĂIJin the WildâĂİ: A Mobile Storytelling Case Study Using StoryKit. *ACM Trans. Comput.-Hum. Interact.* 20, 3 (July 2013), 18:1–18:38. DOI: http://dx.doi.org/10.1145/2491500.2491506

[17] N. Bryan-Kinns. 2004. Daisyphone: The Design and Impact of a Novel Environment for Remote Group Music Improvisation. In *Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*. ACM, New York, NY, USA, 135–144. DOI: http://dx.doi.org/10.1145/1013115.1013135

[18] Nick Bryan-Kinns, Patrick G. T. Healey, and Joe Leach. 2007. Exploring Mutual Engagement in Creative Collaborations. In *Proceedings of the 6th ACM SIGCHI Conference on Creativity & Cognition*. ACM, New York, NY, USA, 223–232. DOI: http://dx.doi.org/10.1145/1254960.1254991

[19] Hronn Brynjarsdottir, Maria HÃěkansson, James Pierce, Eric Baumer, Carl DiSalvo, and Phoebe Sengers. 2012. Sustainably unpersuaded: how persuasion narrows our vision of sustainability. In

Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems (CHI '12). ACM, New York, NY, USA, 947–956. DOI: http://dx.doi.org/10.1145/2208516.2208539

[20] Peter Calingaert. 1967. System Performance Evaluation: Survey and Appraisal. *Commun. ACM* 10, 1 (Jan. 1967), 12–18. DOI: http://dx.doi.org/10.1145/363018.363040

[21] Xiang Cao, SiÃćn E. Lindley, John Helmes, and Abigail Sellen. 2010. Telling the Whole Story: Anticipation, Inspiration and Reputation in a Field Deployment of TellTable. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*. ACM, New York, NY, USA, 251–260. DOI: http://dx.doi.org/10.1145/1718918.1718967

[22] Erin A. Carroll and Celine Latulipe. 2009. The Creativity Support Index. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems (CHI EA '09)*. ACM, New York, NY, USA, 4009–4014. DOI:http://dx.doi.org/10.1145/1520340.1520609

[23] Erin A. Carroll, Celine Latulipe, Richard Fung, and Michael Terry. 2009. Creativity Factor Evaluation: Towards a Standardized Survey Metric for Creativity Support. In *Proceedings of the Seventh ACM Conference on Creativity and Cognition*. ACM, New York, NY, USA, 127–136. DOI: http://dx.doi.org/10.1145/1640233.1640255

[24] Alejandro Catala, Javier Jaen, Betsy van Dijk, and Sergi JordÃă. 2012. Exploring Tabletops As an Effective Tool to Foster Creativity Traits. In *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction*. ACM, New York, NY, USA, 143–150. DOI: http://dx.doi.org/10.1145/2148131.2148163

[25] Joel Chan, Steven Dang, and Steven P Dow. 2016. Improving Crowd Innovation with Expert Facilitation. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW '16*. ACM Press, San Francisco, California, USA, 1221–1233. DOI: http://dx.doi.org/10.1145/2818048.2820023

[26] Siddhartha Chaudhuri and Vladlen Koltun. 2010. Data-driven Suggestions for Creativity Support in 3D Modeling. In *ACM SIGGRAPH Asia 2010 Papers*. ACM, New York, NY, USA, 183:1–183:10. DOI: http://dx.doi.org/10.1145/1866158.1866205

[27] Erin Cherry and Celine Latulipe. 2014. Quantifying the Creativity Support of Digital Tools Through the Creativity Support Index. *ACM Trans. Comput.-Hum. Interact.* 21, 4 (June 2014), 21:1–21:25. DOI: http://dx.doi.org/10.1145/2617588

[28] Sharon Lynn Chu Yew Yee, Francis K.H. Quek, and Lin Xiao. 2011. Studying Medium Effects on Children's Creative Processes. In *Proceedings of the 8th ACM Conference on Creativity and Cognition*. ACM, New York, NY, USA, 3–12. DOI: http://dx.doi.org/10.1145/2069618.2069622

[29] Marianela Ciolfi Felice, Sarah Fdili Alaoui, and Wendy E. Mackay. 2018. Knotation: Exploring and Documenting Choreographic Processes. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 448:1–448:12. DOI: http://dx.doi.org/10.1145/3173574.3174022

[30] Elizabeth Clark, Anne Spencer Ross, Chenhao Tan, Yangfeng Ji, and Noah A. Smith. 2018. Creative Writing with a Machine in the Loop: Case Studies on Slogans and Stories. In *Proceedings of the 2018 Conference on Human Information Interaction&Retrieval - IUI '18*. ACM Press, Tokyo, Japan, 329–340. DOI: http://dx.doi.org/10.1145/3172944.3172983

[31] Andy Cockburn, Carl Gutwin, and Alan Dix. 2018. HARK No More: On the Preregistration of CHI Experiments. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 141, 12 pages. DOI: http://dx.doi.org/10.1145/3173574.3173715

[32] Tim Coughlan and Peter Johnson. 2009. Understanding Productive, Structural and Longitudinal Interactions in the Design of Tools for Creative Activities. In *Proceedings of the Seventh ACM Conference on Creativity and Cognition*. ACM, New York, NY, USA, 155–164. DOI: http://dx.doi.org/10.1145/1640233.1640258

[33] Peter Dalsgaard. 2014. Pragmatism and Design Thinking. 8, 1 (2014), 13.

[34] Peter Dalsgaard, Christian Remy, Jonas Frich Pedersen, Lindsay MacDonald Vermeulen, and Michael Mose Biskjaer. 2018. Digital Tools in Collaborative Creative Work. In *Proceedings of the 10th Nordic Conference on Human-Computer Interaction (NordiCHI '18)*. ACM, New York, NY, USA, 964–967. DOI: http://dx.doi.org/10.1145/3240167.3240262

[35] Sayamindu Dasgupta, William Hale, AndrÃl's Monroy-HernÃąndez, and Benjamin Mako Hill. 2016. Remixing as a Pathway to Computational Thinking. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW '16*. ACM Press, San Francisco, California, USA, 1436–1447. DOI: http://dx.doi.org/10.1145/2818048.2819984

[36] Sayamindu Dasgupta and Benjamin Mako Hill. 2017. Scratch Community Blocks: Supporting Children as Data Scientists. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*. ACM Press, Denver, Colorado, USA, 3620–3631. DOI: http://dx.doi.org/10.1145/3025453.3025847

[37] Matthew E. P. Davies, Philippe Hamel, Kazuyoshi Yoshii, and Masataka Goto. 2014. AutoMashUpper: Automatic Creation of Multi-song Music Mashups. *IEEE/ACM Trans. Audio, Speech and Lang. Proc.* 22, 12 (Dec. 2014), 1726–1737. DOI: http://dx.doi.org/10.1109/TASLP.2014.2347135

[38] Nicholas Davis, Chih-PIn Hsiao, Kunwar Yashraj Singh, Lisa Li, Sanat Moningi, and Brian Magerko. 2015. Drawing Apprentice: An Enactive Co-Creative Agent for Artistic Collaboration. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition*. ACM, New York, NY, USA, 185–186. DOI: http://dx.doi.org/10.1145/2757226.2764555

[39] Nicholas Davis, Chih-PIn Hsiao, Kunwar Yashraj Singh, Lisa Li, and Brian Magerko. 2016. Empirically Studying Participatory Sense-Making in Abstract Drawing with a Co-Creative Cognitive Agent. In *Proceedings of the 21st International Conference on Intelligent User Interfaces - IUI '16*. ACM Press, Sonoma, California, USA, 196–207. DOI: http://dx.doi.org/10.1145/2856767.2856795

[40] Nicholas Davis, Alexander Zook, Brian O'Neill, Brandon Headrick, Mark Riedl, Ashton Grosz, and Michael Nitsche. 2013. Creativity Support for Novice Digital Filmmaking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 651–660. DOI: http://dx.doi.org/10.1145/2470654.2470747

[41] Jelle van Dijk, Jirka van der Roest, Remko van der Lugt, and Kees C.J. Overbeeke. 2011. NOOT: A Tool for Sharing Moments of Reflection During Creative Meetings. In *Proceedings of the 8th ACM Conference on Creativity and Cognition*. ACM, New York, NY, USA, 157–164. DOI: http://dx.doi.org/10.1145/2069618.2069646

[42] Carl DiSalvo, Marti Louw, Julina Coupland, and MaryAnn Steiner. 2009. Local Issues, Local Uses: Tools for Robotics and Sensing in Community Contexts. In *Proceedings of the Seventh ACM Conference on Creativity and Cognition*. ACM, New York, NY, USA, 245–254. DOI: http://dx.doi.org/10.1145/1640233.1640271

[43] Alan Dix. 2008. Theoretical analysis and theory creation. In *Research Methods for Human-Computer Interaction*. Cambridge University Press.

[44] Douglas Engelbart. 1962. *Augmenting Human Intellect: A Conceptual Framework*. Summary Report AFOSR-3233. Washington, DC, USA. https://web.stanford.edu/dept/SUL/library/extra4/sloan/mousesite/EngelbartPapers/B5_F18_ConceptFrameworkInd.html

[45] Shelly Engelman, Brian Magerko, Tom McKlin, Morgan Miller, Doug Edwards, and Jason Freeman. 2017. Creativity in Authentic STEAM Education with EarSketch. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education - SIGCSE '17*. ACM Press, Seattle, Washington, USA, 183–188. DOI: http://dx.doi.org/10.1145/3017680.3017763

[46] Umer Farooq, John M. Carroll, and Craig H. Ganoe. 2007. Supporting Creativity with Awareness in Distributed Collaboration. In *Proceedings of the 2007 International ACM Conference on Supporting Group Work*. ACM, New York, NY, USA, 31–40. DOI: http://dx.doi.org/10.1145/1316624.1316630

[47] Gerhard Fischer. 2004. Social Creativity: Turning Barriers into Opportunities for Collaborative Design. In *Proceedings of the Eighth Conference on Participatory Design: Artful Integration: Interweaving Media, Materials and Practices - Volume 1*. ACM, New York, NY, USA, 152–161. DOI: http://dx.doi.org/10.1145/1011870.1011889

[48] Jonas Frich, Michael Mose Biskjaer, and Peter Dalsgaard. 2018a. Why HCI and Creativity Research Must Collaborate to Develop New Creativity Support Tools. In *Proceedings of the Technology, Mind, and Society*. ACM Press, Washington, DC, USA, 1–6. DOI: http://dx.doi.org/10.1145/3183654.3183678

[49] Jonas Frich, Michael Mose Biskjaer, and Peter Dalsgaard. 2018b. Twenty Years of Creativity Research in Human-Computer Interaction: Current State and Future Directions. In *Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18)*. ACM, New York, NY, USA, 1235–1257. DOI: http://dx.doi.org/10.1145/3196709.3196732

[50] Jonas Frich, Lindsay MacDonald Vermeulen, Christian Remy, Michael Mose Biskjaer, and Peter Dalsgaard. 2019. Mapping the Landscape of Creativity Support Tools in HCI. 18.

[51] Quentin Galvane, Marc Christie, Chrsitophe Lino, and RÃl'mi Ronfard. 2015. Camera-on-rails: Automated Computation of Constrained Camera Paths. In *Proceedings of the 8th ACM SIGGRAPH Conference on Motion in Games*. ACM, New York, NY, USA, 151–157. DOI: http://dx.doi.org/10.1145/2822013.2822025

[52] JÃl'rÃl'mie Garcia, Theophanis Tsandilas, Carlos Agon, and Wendy E. Mackay. 2014. Structured Observation with Polyphony: A Multifaceted Tool for Studying Music Composition. In *Proceedings of the 2014 Conference on Designing Interactive Systems*. ACM, New York, NY, USA, 199–208. DOI: http://dx.doi.org/10.1145/2598510.2598512

[53] Howard Gardner. 1988. Creativity: An interdisciplinary perspective. *Creativity Research Journal* 1, 1 (Dec. 1988), 8–26. DOI: http://dx.doi.org/10.1080/10400418809534284

[54] Jill GerhardtâĂŘPowals. 1996. Cognitive engineering principles for enhancing humanâĂŘcomputer performance. *International Journal of Human-Computer Interaction* 8, 2 (Apr 1996),

189âĂŞ211. DOI: http://dx.doi.org/10.1080/10447319609526147

[55] Florian Geyer, Jochen Budzinski, and Harald Reiterer. 2012. IdeaVis: A Hybrid Workspace and Interactive Visualization for Paper-based Collaborative Sketching Sessions. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design*. ACM, New York, NY, USA, 331–340. DOI:http://dx.doi.org/10.1145/2399016.2399069

[56] Florian Geyer, Ulrike Pfeil, Anita HÃűchtl, Jochen Budzinski, and Harald Reiterer. 2011. Designing Reality-based Interfaces for Creative Group Work. In *Proceedings of the 8th ACM Conference on Creativity and Cognition*. ACM, New York, NY, USA, 165–174. DOI:http://dx.doi.org/10.1145/2069618.2069647

[57] Karni Gilon, Joel Chan, Felicia Y. Ng, Hila Liifshitz-Assaf, Aniket Kittur, and Dafna Shahaf. 2018. Analogy Mining for Specific Design Needs. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. ACM Press, Montreal QC, Canada, 1–11. DOI: http://dx.doi.org/10.1145/3173574.3173695

[58] Victor Girotto, Erin Walker, and Winslow Burleson. 2017. The Effect of Peripheral Micro-tasks on Crowd Ideation. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*. ACM Press, Denver, Colorado, USA, 1843–1854. DOI: http://dx.doi.org/10.1145/3025453.3025464

[59] Eric Gordon, Becky E. Michelson, and Jason Haas. 2016. @Stake: A Game to Facilitate the Process of Deliberative Democracy. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion - CSCW '16 Companion*. ACM Press, San Francisco, California, USA, 269–272. DOI: http://dx.doi.org/10.1145/2818052.2869125

[60] Wayne D. Gray. 1995. Discount or Disservice?: Discount Usability AnalysisâĂŞevaluation at a Bargain Price or Simply Damaged Merchandise? (Panel Session). In *Conference Companion on Human Factors in Computing Systems (CHI '95)*. ACM, New York, NY, USA, 176–177. DOI: http://dx.doi.org/10.1145/223355.223488

[61] Saul Greenberg and Bill Buxton. 2008. Usability Evaluation Considered Harmful (Some of the Time). In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*. ACM, New York, NY, USA, 111–120. DOI: http://dx.doi.org/10.1145/1357054.1357074

[62] Garth Griffin and Robert Jacob. 2013. Priming Creativity Through Improvisation on an Adaptive Musical Instrument. In *Proceedings of the 9th ACM Conference on Creativity & Cognition*. ACM, New York, NY, USA, 146–155. DOI: http://dx.doi.org/10.1145/2466627.2466630

[63] J. P. Guilford. 1950. Creativity. *American Psychologist* 5, 9 (1950), 444–454. DOI: http://dx.doi.org/10.1037/h0063487

[64] Joy Paul Guilford. 1967. *The nature of human intelligence*. McGraw-Hill, New York.

[65] Florian GÃijldenpfennig, Wolfgang Reitberger, and Geraldine Fitzpatrick. 2012. Capturing Rich Media Through Media Objects on Smartphones. In *Proceedings of the 24th Australian Computer-Human Interaction Conference*. ACM, New York, NY, USA, 180–183. DOI: http://dx.doi.org/10.1145/2414536.2414569

[66] Adam Haar Horowitz, Ishaan Grover, Pedro Reynolds-CuÃľllar, Cynthia Breazeal, and Pattie Maes. 2018. Dormio: Interfacing with Dreams. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (CHI EA '18)*. ACM, New York, NY, USA, alt10:1–alt10:10. DOI: http://dx.doi.org/10.1145/3170427.3188403

[67] Joshua Hailpern, Erik Hinterbichler, Caryn Leppert, Damon Cook, and Brian P. Bailey. 2007. TEAM STORM: Demonstrating an Interaction Model for Working with Multiple Ideas During Creative Group Work. In *Proceedings of the 6th ACM SIGCHI Conference on Creativity & Cognition*. ACM, New York, NY, USA, 193–202. DOI: http://dx.doi.org/10.1145/1254960.1254987

[68] John Halloran, Eva Hornecker, Geraldine Fitzpatrick, Mark Weal, David Millard, Danius Michaelides, Don Cruickshank, and David De Roure. 2006. The Literacy Fieldtrip: Using UbiComp to Support Children's Creative Writing. In *Proceedings of the 2006 Conference on Interaction Design and Children*. ACM, New York, NY, USA, 17–24. DOI: http://dx.doi.org/10.1145/1139073.1139083

[69] Megan K. Halpern, Jakob Tholander, Max Evjen, Stuart Davis, Andrew Ehrlich, Kyle Schustak, Eric P.S. Baumer, and Geri Gay. 2011. MoBoogie: Creative Expression Through Whole Body Musical Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 557–560. DOI: http://dx.doi.org/10.1145/1978942.1979020

[70] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Human mental workload*. North-Holland, Oxford, England, 139–183. DOI: http://dx.doi.org/10.1016/S0166-4115(08)62386-9

[71] Peter HedstrÃűm and Petri Ylikoski. 2010. Causal Mechanisms in the Social Sciences. *Annual Review of Sociology* 36, 1 (June 2010), 49–67. DOI: http://dx.doi.org/10.1146/annurev.soc.012809.102632

[72] Beth A Hennessey and Teresa M Amabile. 1999. Consensual assessment. *Encyclopedia of creativity* 1 (1999), 347–359.

[73] Tom Hewett, Mary Czerwinski, Michael Terry, Jay Nunamaker, Linda Candy, Bill Kules, and Elisabeth Sylvan. 2005. Creativity support tool evaluation methods and metrics. *Creativity Support Tools* (2005), 10–24.

[74] Otmar Hilliges, Lucia Terrenghi, Sebastian Boring, David Kim, Hendrik Richter, and Andreas Butz. 2007. Designing for Collaborative Creative Problem Solving. In *Proceedings of the 6th ACM SIGCHI Conference on Creativity & Cognition*. ACM, New York, NY, USA, 137–146. DOI: http://dx.doi.org/10.1145/1254960.1254980

[75] D. Hocevar. 1981. Measurement of creativity: review and critique. *Journal of Personality Assessment* 45, 5 (Oct. 1981), 450–464. DOI: http://dx.doi.org/10.1207/s15327752jpa4505_1

[76] Rania Hodhod and Brian Magerko. 2016. Closing the Cognitive Gap between Humans and Interactive Narrative Agents Using Shared Mental Models. In *Proceedings of the 21st International Conference on Intelligent User Interfaces - IUI '16*. ACM Press, Sonoma, California, USA, 135–146. DOI: http://dx.doi.org/10.1145/2856767.2856774

[77] Eva Hornecker. 2010. Creative Idea Exploration Within the Structure of a Guiding Framework: The Card Brainstorming Game. In *Proceedings of the Fourth International Conference on Tangible, Embedded, and Embodied Interaction*. ACM, New York, NY, USA, 101–108. DOI: http://dx.doi.org/10.1145/1709886.1709905

[78] Cheng-Zhi Anna Huang, David Duvenaud, and Krzysztof Z. Gajos. 2016. ChordRipple: Recommending Chords to Help Novice Composers Go Beyond the Ordinary. In *Proceedings of the 21st International Conference on Intelligent User Interfaces - IUI '16*. ACM Press, Sonoma, California, USA, 241–250. DOI: http://dx.doi.org/10.1145/2856767.2856792

[79] Samuel Huron, Pauline Gourlet, Uta Hinrichs, Trevor Hogan, and Yvonne Jansen. 2017. Let's Get Physical: Promoting Data Physicalization in Workshop Formats. In *Proceedings of the 2017 Conference on Designing Interactive Systems - DIS '17*. ACM Press, Edinburgh, United Kingdom, 1409–1422. DOI: http://dx.doi.org/10.1145/3064663.3064798

[80] Junko Ichino, Aura Pon, Ehud Sharlin, David Eagle, and Sheelagh Carpendale. 2014. Vuzik: The Effect of Large Gesture Interaction on Children's Creative Musical Expression. In *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: The Future of Design*. ACM, New York, NY, USA, 240–249. DOI: http://dx.doi.org/10.1145/2686612.2686649

[81] Sam Jacoby and Leah Buechley. 2013. Drawing the Electric: Storytelling with Conductive Ink. In *Proceedings of the 12th International Conference on Interaction Design and Children*. ACM, New York, NY, USA, 265–268. DOI: http://dx.doi.org/10.1145/2485760.2485790

[82] Peter H. Kahn, Takayuki Kanda, Hiroshi Ishiguro, Brian T. Gill, Solace Shen, Jolina H. Ruckert, and Heather E. Gary. 2016. Human creativity can be facilitated through interacting with a social robot. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, Christchurch, New Zealand, 173–180. DOI: http://dx.doi.org/10.1109/HRI.2016.7451749

[83] Michael Karlesky and Katherine Isbister. 2014. Designing for the Physical Margins of Digital Workspaces: Fidget Widgets in Support of Productivity and Creativity. In *Proceedings of the 8th International Conference on Tangible, Embedded and Embodied Interaction*. ACM, New York, NY, USA, 13–20. DOI: http://dx.doi.org/10.1145/2540930.2540978

[84] Jun Kato, Tomoyasu Nakano, and Masataka Goto. 2015. TextAlive: Integrated Design Environment for Kinetic Typography. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 3403–3412. DOI: http://dx.doi.org/10.1145/2702123.2702140

[85] Natsumi Kato, Hiroyuki Osone, Daitetsu Sato, Naoya Muramatsu, and Yoichi Ochiai. 2018. DeepWear: a Case Study of Collaborative Design between Human and Artificial Intelligence. In *Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction - TEI '18*. ACM Press, Stockholm, Sweden, 529–536. DOI: http://dx.doi.org/10.1145/3173225.3173302

[86] James C. Kaufman. 2014. Measuring Creativity - the Last Windmill? | The Creativity Post. (2014). http://www.creativitypost.com/psychology/measuring_creativity_the_last_windmill

[87] James C. Kaufman, Jonathan A. Plucker, and John Baer. 2008. *Essentials of creativity assessment*. John Wiley & Sons Inc.

[88] Joseph 'Jofish' Kaye. 2007. Evaluating Experience-focused HCI. In *CHI '07 Extended Abstracts on Human Factors in Computing Systems (CHI EA '07)*. ACM, New York, NY, USA, 1661–1664. DOI:http://dx.doi.org/10.1145/1240866.1240877

[89] Rubaiat Habib Kazi, Kien Chuan Chua, Shengdong Zhao, Richard Davis, and Kok-Lim Low. 2011. SandCanvas: A Multi-touch Art Medium Inspired by Sand Animation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1283–1292. DOI: http://dx.doi.org/10.1145/1978942.1979133

[90] Andruid Kerne, Andrew Billingsley, Nic Lupfer, Rhema Linder, Yin Qu, Alyssa Valdez, Ajit Jain, Kade Keith, Matthew Carrasco, and Jorge Vanegas. 2017. 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 - C&C '17*. ACM Press, Singapore, Singapore, 380–392. DOI: http://dx.doi.org/10.1145/3059454.3059471

[91] Andruid Kerne, Eunyee Koh, Steven Smith, Hyun Choi, Ross Graeber, and Andrew Webb. 2007. Promoting Emergence in Information Discovery by Representing Collections with Composition. In *Proceedings of the 6th ACM SIGCHI Conference on Creativity & Cognition*. ACM, New York, NY, USA, 117–126. DOI: http://dx.doi.org/10.1145/1254960.1254977

[92] Andruid Kerne, Eunyee Koh, Steven M. Smith, Andrew Webb, and Blake Dworaczyk. 2008. combinFormation: Mixed-initiative Composition of Image and Text Surrogates Promotes Information Discovery. *ACM Trans. Inf. Syst.* 27, 1 (Dec. 2008), 5:1–5:45. DOI: http://dx.doi.org/10.1145/1416950.1416955

[93] Andruid Kerne, Andrew M. Webb, Celine Latulipe, Erin Carroll, Steven M. Drucker, Linda Candy, and Kristina HÃűÃűk. 2013. Evaluation Methods for Creativity Support Environments. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems (CHI EA '13)*. ACM, New York, NY, USA, 3295–3298. DOI:http://dx.doi.org/10.1145/2468356.2479670

[94] Andruid Kerne, Andrew M. Webb, Steven M. Smith, Rhema Linder, Nic Lupfer, Yin Qu, Jon Moeller, and Sashikanth Damaraju. 2014. Using Metrics of Curation to Evaluate Information-Based Ideation. *ACM Trans. Comput.-Hum. Interact.* 21, 3 (June 2014), 14:1–14:48. DOI:http://dx.doi.org/10.1145/2591677

[95] Joy Kim, Maneesh Agrawala, and Michael S. Bernstein. 2017. 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 - CSCW '17*. ACM Press, Portland, Oregon, USA, 246–258. DOI: http://dx.doi.org/10.1145/2998181.2998195

[96] Joy Kim, Justin Cheng, and Michael S. Bernstein. 2014. Ensemble: Exploring Complementary Strengths of Leaders and Crowds in Creative Collaboration. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, New York, NY, USA, 745–755. DOI: http://dx.doi.org/10.1145/2531602.2531638

[97] Joy Kim, Mira Dontcheva, Wilmot Li, Michael S. Bernstein, and Daniela Steinsapir. 2015. Motif: Supporting Novice Creativity Through Expert Patterns. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1211–1220. DOI: http://dx.doi.org/10.1145/2702123.2702507

[98] Joy Kim and Andres Monroy-Hernandez. 2016. Storia: Summarizing Social Media Content Based on Narrative Theory Using Crowdsourcing. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. ACM, New York, NY, USA, 1018–1027. DOI: http://dx.doi.org/10.1145/2818048.2820072

[99] Joy Kim, Sarah Sterman, Allegra Argent Beal Cohen, and Michael S. Bernstein. 2017. Mechanical Novel: Crowdsourcing Complex Work through Reflection and Revision. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing - CSCW '17*. ACM Press, Portland, Oregon, USA, 233–245. DOI: http://dx.doi.org/10.1145/2998181.2998196

[100] Celine Latulipe, Erin A. Carroll, and Danielle Lottridge. 2011. Evaluating Longitudinal Projects Combining Technology with Temporal Arts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1835–1844. DOI: http://dx.doi.org/10.1145/1978942.1979209

[101] David Ledo, Steven Houben, Jo Vermeulen, Nicolai Marquardt, Lora Oehlberg, and Saul Greenberg. 2018. Evaluation Strategies for HCI Toolkit Research. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 36:1–36:17. DOI: http://dx.doi.org/10.1145/3173574.3173610

[102] Sara Ljungblad. 2009. Passive Photography from a Creative Perspective: "If I Would Just Shoot the Same Thing for Seven Days, It's Like... What's the Point?". In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 829–838. DOI: http://dx.doi.org/10.1145/1518701.1518828

[103] Fei Lu, Feng Tian, Yingying Jiang, Xiang Cao, Wencan Luo, Guang Li, Xiaolong Zhang, Guozhong Dai, and Hongan Wang. 2011. ShadowStory: Creative and Collaborative Digital Storytelling Inspired by Cultural Heritage. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1919–1928. DOI: http://dx.doi.org/10.1145/1978942.1979221

[104] Kurt Luther, Casey Fiesler, and Amy Bruckman. 2013. Redistributing Leadership in Online Creative Collaboration. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*. ACM, New York, NY, USA, 1007–1022. DOI: http://dx.doi.org/10.1145/2441776.2441891

[105] Neil Maiden, Konstantinos Zachos, Amanda Brown, George Brock, Lars Nyre, Aleksander NygÃěrd Tonheim, Dimitris Apsotolou, and Jeremy Evans. 2018. Making the News: Digital Creativity Support for Journalists. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. ACM Press, Montreal QC, Canada, 1–11. DOI:http://dx.doi.org/10.1145/3173574.3174049

[106] Nicolas Mangano, Alex Baker, Mitch Dempsey, Emily Navarro, and AndrÃl' van der Hoek. 2010. Software Design Sketching with Calico. In *Proceedings of the IEEE/ACM International Conference on Automated Software Engineering*. ACM, New York, NY, USA, 23–32. DOI: http://dx.doi.org/10.1145/1858996.1859003

[107] Charles Martin, Henry Gardner, Ben Swift, and Michael Martin. 2016. Intelligent Agents and Networked Buttons Improve Free-Improvised Ensemble Music-Making on Touch-Screens. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*. ACM Press, Santa Clara, California, USA, 2295–2306. DOI: http://dx.doi.org/10.1145/2858036.2858269

[108] J. Nathan Matias, Sayamindu Dasgupta, and Benjamin Mako Hill. 2016. Skill Progression in Scratch Revisited. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*. ACM Press, Santa Clara, California, USA, 1486–1490. DOI: http://dx.doi.org/10.1145/2858036.2858349

[109] Nolwenn Maudet, Ghita Jalal, Philip Tchernavskij, Michel Beaudouin-Lafon, and Wendy E. Mackay. 2017. Beyond Grids: Interactive Graphical Substrates to Structure Digital Layout. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*. ACM Press, Denver, Colorado, USA, 5053–5064. DOI: http://dx.doi.org/10.1145/3025453.3025718

[110] Ali Mazalek, Sanjay Chandrasekharan, Michael Nitsche, Tim Welsh, Paul Clifton, Andrew Quitmeyer, Firaz Peer, Friedrich Kirschner, and Dilip Athreya. 2011. I'M in the Game: Embodied Puppet Interface Improves Avatar Control. In *Proceedings of the Fifth International Conference on Tangible, Embedded, and Embodied Interaction*. ACM, New York, NY, USA, 129–136. DOI: http://dx.doi.org/10.1145/1935701.1935727

[111] Florian Mueller, Martin R. Gibbs, Frank Vetere, and Darren Edge. 2014. Supporting the Creative Game Design Process with Exertion Cards. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 2211–2220. DOI: http://dx.doi.org/10.1145/2556288.2557272

[112] Brad A. Myers, Ashley Lai, Tam Minh Le, YoungSeok Yoon, Andrew Faulring, and Joel Brandt. 2015. Selective Undo Support for Painting Applications. ACM Press, 4227–4236. DOI: http://dx.doi.org/10.1145/2702123.2702543

[113] Naoto Nakazato, Shigeo Yoshida, Sho Sakurai, Takuji Narumi, Tomohiro Tanikawa, and Michitaka Hirose. 2014. Smart Face: Enhancing Creativity During Video Conferences Using Real-time Facial Deformation. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, New York, NY, USA, 75–83. DOI: http://dx.doi.org/10.1145/2531602.2531637

[114] Grace Ngai, Stephen C.F. Chan, Hong Va Leong, and Vincent T.Y. Ng. 2013. Designing I*CATch: A Multipurpose, Education-friendly Construction Kit for Physical and Wearable Computing. *Trans. Comput. Educ.* 13, 2 (July 2013), 7:1–7:30. DOI: http://dx.doi.org/10.1145/2483710.2483712

[115] Tricia J. Ngoon, C. Ailie Fraser, Ariel S. Weingarten, Mira Dontcheva, and Scott Klemmer. 2018. Interactive Guidance Techniques for Improving Creative Feedback. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. ACM Press, Montreal QC, Canada, 1–11. DOI: http://dx.doi.org/10.1145/3173574.3173629

[116] Jakob Nielsen. 1989. Usability Engineering at a Discount. In *Proceedings of the Third International Conference on Human-computer Interaction on Designing and Using Human-computer Interfaces and Knowledge Based Systems (2Nd Ed.)*. Elsevier Science Inc., New York, NY, USA, 394–401. http://dl.acm.org/citation.cfm?id=92449.92499

[117] Jakob Nielsen. 1994. *Usability engineering*. Morgan Kaufmann Publishers, San Francisco, Calif.

[118] Lora Oehlberg, Kyu Simm, Jasmine Jones, Alice Agogino, and BjÃűrn Hartmann. 2012. Showing is Sharing: Building Shared Understanding in Human-centered Design Teams with Dazzle. In *Proceedings of the Designing Interactive Systems Conference*. ACM, New York, NY, USA, 669–678. DOI:http://dx.doi.org/10.1145/2317956.2318057

[119] Hyunjoo Oh, Jiffer Harriman, Abhishek Narula, Mark D. Gross, Michael Eisenberg, and Sherry Hsi. 2016. Crafting Mechatronic Percussion with Everyday Materials. In *Proceedings of the TEI '16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction - TEI '16*. ACM Press, Eindhoven, Netherlands, 340–348. DOI: http://dx.doi.org/10.1145/2839462.2839474

[120] Antti Oulasvirta, Anna Feit, Perttu LÃd'hteenlahti, and Andreas Karrenbauer. 2017. Computational Support for Functionality Selection in Interaction Design. *ACM Transactions on Computer-Human Interaction* 24, 5 (Oct. 2017), 1–30. DOI: http://dx.doi.org/10.1145/3131608

[121] Ingrid Pettersson, Anna-Katharina Frison, Florian Lachner, Andreas Riener, and Jesper Nolhage. 2017. Triangulation in UX Studies: Learning from Experience. In *Proceedings of the 2017 ACM Conference Companion Publication on Designing Interactive Systems (DIS '17 Companion)*. ACM, New York, NY, USA, 341–344. DOI: http://dx.doi.org/10.1145/3064857.3064858

[122] Ingrid Pettersson, Florian Lachner, Anna-Katharina Frison, Andreas Riener, and Andreas Butz. 2018. A Bermuda Triangle?: A Review of Method Application and Triangulation in User Experience Evaluation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 461:1–461:16. DOI: http://dx.doi.org/10.1145/3173574.3174035

[123] Cecil Piya, Vinayak , Senthil Chandrasegaran, Niklas Elmqvist, and Karthik Ramani. 2017. Co-3Deator: A Team-First Collaborative 3D Design Ideation Tool. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*. ACM Press, Denver, Colorado, USA, 6581–6592. DOI: http://dx.doi.org/10.1145/3025453.3025825

[124] Karl R. Popper. 1962. *Conjectures and Refutations: The Growth of Scientific Knowledge* (first edition edition ed.). Basic Books.

[125] Thorsten Prante, Carsten Magerkurth, and Norbert Streitz. 2002. Developing CSCW Tools for Idea Finding -: Empirical Results and Implications for Design. In *Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work*. ACM, New York, NY, USA, 106–115. DOI: http://dx.doi.org/10.1145/587078.587094

[126] Daniel Rees Lewis, Emily Harburg, Elizabeth Gerber, and Matthew Easterday. 2015. Building Support Tools to Connect Novice Designers with Professional Coaches. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition*. ACM, New York, NY, USA, 43–52. DOI: http://dx.doi.org/10.1145/2757226.2757248

[127] Christian Remy, Oliver Bates, Alan Dix, Vanessa Thomas, Mike Hazas, Adrian Friday, and Elaine M. Huang. 2018a. Evaluation beyond Usability: Validating Sustainable HCI Research. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*.

[128] Christian Remy, Oliver Bates, Jennifer Mankoff, and Adrian Friday. 2018b. Evaluating HCI Research beyond Usability. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. Montreal, Canada.

[129] Christian Remy, Oliver Bates, Vanessa Thomas, and Elaine May Huang. 2017. The Limits of Evaluating Sustainability. In *Proceedings of the Third Workshop on Computing within Limits*. ACM Press, Santa Barbara, California, USA.

[130] Mel Rhodes. 1961. An Analysis of Creativity. *The Phi Delta Kappan* 42, 7 (1961), 305–310. https://www.jstor.org/stable/20342603

[131] Daniela Rosner and Jonathan Bean. 2009. Learning from IKEA Hacking: I'M Not One to Decoupage a Tabletop and Call It a Day.. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 419–422. DOI: http://dx.doi.org/10.1145/1518701.1518768

[132] Daniela K. Rosner and Kimiko Ryokai. 2010. Spyn: Augmenting the Creative and Communicative Potential of Craft. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 2407–2416. DOI: http://dx.doi.org/10.1145/1753326.1753691

[133] Mark A. Runco and Selcuk Acar. 2012. Divergent Thinking as an Indicator of Creative Potential. *Creativity Research Journal* 24, 1 (2012), 66–75. DOI: http://dx.doi.org/10.1080/10400419.2012.652929

[134] Mark A. Runco and Garrett J. Jaeger. 2012. The Standard Definition of Creativity. *Creativity Research Journal* 24, 1 (Jan. 2012), 92–96. DOI: http://dx.doi.org/10.1080/10400419.2012.650092

[135] Jana Schumann, Patrick C. Shih, David F. Redmiles, and Graham Horton. 2012. Supporting Initial Trust in Distributed Idea Generation and Idea Evaluation. In *Proceedings of the 17th ACM International Conference on Supporting Group Work*. ACM, New York, NY, USA, 199–208. DOI: http://dx.doi.org/10.1145/2389176.2389207

[136] Duane F. Shell, Leen-Kiat Soh, Abraham E. Flanigan, Markeya S. Peteranetz, and Elizabeth Ingraham. 2017. Improving Students' Learning and Achievement in CS Classrooms through Computational Creativity Exercises that Integrate Computational and Creative Thinking. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education - SIGCSE '17*. ACM Press, Seattle, Washington, USA, 543–548. DOI: http://dx.doi.org/10.1145/3017680.3017718

[137] Ben Shneiderman. 2000. Creating Creativity: User Interfaces for Supporting Innovation. *ACM Trans. Comput.-Hum. Interact.* 7, 1 (March 2000), 114–138. DOI:http://dx.doi.org/10.1145/344949.345077

[138] Ben Shneiderman. 2007. Creativity Support Tools: Accelerating Discovery and Innovation. *Commun. ACM* 50, 12 (Dec. 2007), 20–32. DOI: http://dx.doi.org/10.1145/1323688.1323689

[139] Maria Shugrina, Jingwan Lu, and Stephen Diverdi. 2017. Playful palette: an interactive parametric color mixer for artists. *ACM Transactions on Graphics* 36, 4 (July 2017), 1–10. DOI: http://dx.doi.org/10.1145/3072959.3073690

[140] Pao Siangliulue, Joel Chan, Krzysztof Z. Gajos, and Steven P. Dow. 2015. Providing Timely Examples Improves the Quantity and Quality of Generated Ideas. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition*. ACM, New York, NY, USA, 83–92. DOI: http://dx.doi.org/10.1145/2757226.2757230

[141] Vikash Singh, Celine Latulipe, Erin Carroll, and Danielle Lottridge. 2011. The Choreographer's Notebook: A Video Annotation System for Dancers and Choreographers. In *Proceedings of the 8th ACM Conference on Creativity and Cognition*. ACM, New York, NY, USA, 197–206. DOI: http://dx.doi.org/10.1145/2069618.2069653

[142] DorothÃl' Smit, Thomas Grah, Martin Murer, Vincent van Rheden, and Manfred Tscheligi. 2018. MacroScope: First-Person Perspective in Physical Scale Models. In *Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction - TEI '18*. ACM Press, Stockholm, Sweden, 253–259. DOI: http://dx.doi.org/10.1145/3173225.3173276

[143] Cesar Torres, Wilmot Li, and Eric Paulos. 2016. ProxyPrint: Supporting Crafting Practice through Physical Computational Proxies. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems - DIS '16*. ACM Press, Brisbane, QLD, Australia, 158–169. DOI: http://dx.doi.org/10.1145/2901790.2901828

[144] Cesar Torres and Eric Paulos. 2015. MetaMorphe: Designing Expressive 3D Models for Digital Fabrication. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition*. ACM, New York, NY, USA, 73–82. DOI: http://dx.doi.org/10.1145/2757226.2757235

[145] Theophanis Tsandilas, Catherine Letondal, and Wendy E. Mackay. 2009. Musink: Composing Music Through Augmented Drawing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 819–828. DOI: http://dx.doi.org/10.1145/1518701.1518827

[146] Rajan Vaish, Shirish Goyal, Amin Saberi, and Sharad Goel. 2018. Creating Crowdsourced Research Talks at Scale. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web - WWW '18*. ACM Press, Lyon, France, 1–11. DOI: http://dx.doi.org/10.1145/3178876.3186031

[147] Dagny Valgeirsdottir and Balder Onarheim. 2017. Studying creativity training programs: A methodological analysis. *Creativity and Innovation Management* 26, 4 (2017), 430âĂŞ439. DOI: http://dx.doi.org/10.1111/caim.12245

[148] Hao-Chuan Wang, Dan Cosley, and Susan R. Fussell. 2010. Idea Expander: Supporting Group Brainstorming with Conversationally Triggered Visual Thinking Stimuli. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*. ACM, New York, NY, USA, 103–106. DOI: http://dx.doi.org/10.1145/1718918.1718938

[149] Hao-Chuan Wang, Susan R. Fussell, and Dan Cosley. 2011. From Diversity to Creativity: Stimulating Group Brainstorming with Cultural Differences and Conversationally-retrieved Pictures. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*. ACM, New York, NY, USA, 265–274. DOI: http://dx.doi.org/10.1145/1958824.1958864

[150] Tianyi Wang, Ke Huo, Pratik Chawla, Guiming Chen, Siddharth Banerjee, and Karthik Ramani. 2018. Plain2Fun: Augmenting Ordinary Objects with Interactive Functions by Auto-Fabricating Surface Painted Circuits. In *Proceedings of the 2018 on Designing Interactive Systems Conference 2018 - DIS '18*. ACM Press, Hong Kong, China, 1095–1106. DOI: http://dx.doi.org/10.1145/3196709.3196791

[151] Andrew Warr and Eamonn O'Neill. 2007. Tool Support for Creativity Using Externalizations. In *Proceedings of the 6th ACM SIGCHI Conference on Creativity & Cognition (C&C '07)*. ACM, New York, NY, USA, 127–136. DOI: http://dx.doi.org/10.1145/1254960.1254979

[152] Kento Watanabe, Yuichiroh Matsubayashi, Kentaro Inui, Tomoyasu Nakano, Satoru Fukayama, and Masataka Goto. 2017. LyriSys: An Interactive Support System for Writing Lyrics Based on Topic Transition. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces - IUI '17*. ACM Press, Limassol, Cyprus, 559–563. DOI: http://dx.doi.org/10.1145/3025171.3025194

[153] Cathleen Wharton, Janice Bradford, Robin Jeffries, and Marita Franzke. 1992. Applying Cognitive Walkthroughs to More Complex User Interfaces: Experiences, Issues, and Recommendations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '92)*. ACM, New York, NY, USA, 381–388. DOI: http://dx.doi.org/10.1145/142750.142864

[154] Karl D.D. Willis, Juncong Lin, Jun Mitani, and Takeo Igarashi. 2010. Spatial Sketch: Bridging Between Movement & Fabrication. In *Proceedings of the Fourth International Conference on Tangible, Embedded, and Embodied Interaction*. ACM, New York, NY, USA, 5–12. DOI:http://dx.doi.org/10.1145/1709886.1709890

[155] Jun Xie, Aaron Hertzmann, Wilmot Li, and Holger WinnemÃűller. 2014. PortraitSketch: Face Sketching Assistance for Novices. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*. ACM, New York, NY, USA, 407–417. DOI:http://dx.doi.org/10.1145/2642918.2647399

[156] Anbang Xu, Shih-Wen Huang, and Brian Bailey. 2014. Voyant: Generating Structured Feedback on Visual Designs Using a Crowd of Non-experts. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, New York, NY, USA, 1433–1444. DOI: http://dx.doi.org/10.1145/2531602.2531604

[157] Junichi Yamaoka and Yasuaki Kakehi. 2013. dePENd: Augmented Handwriting System Using Ferromagnetism of a Ballpoint Pen. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*. ACM, New York, NY, USA, 203–210. DOI: http://dx.doi.org/10.1145/2501988.2502017

[158] Daisy Yoo, Alina Huldtgren, Jill Palzkill Woelfer, David G. Hendry, and Batya Friedman. 2013. A Value Sensitive Action-reflection Model: Evolving a Co-design Space with Stakeholder and Designer Prompts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 419–428. DOI: http://dx.doi.org/10.1145/2470654.2470715

[159] Sang Ho Yoon, Ansh Verma, Kylie Peppler, and Karthik Ramani. 2015. HandiMate: Exploring a Modular Robotics Kit for Animating Crafted Toys. In *Proceedings of the 14th International Conference on Interaction Design and Children*. ACM, New York, NY, USA, 11–20. DOI: http://dx.doi.org/10.1145/2771839.2771841

[160] Natsuko Yoshida, Shogo Fukushima, Daiya Aida, and Takeshi Naemura. 2016. Practical Study of Positive-feedback Button for Brainstorming with Interjection Sound Effects. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '16*. ACM Press, Santa Clara, California, USA, 1322–1328. DOI: http://dx.doi.org/10.1145/2851581.2892418

[161] Lixiu Yu and Jeffrey V. Nickerson. 2011. Cooks or Cobblers?: Crowd Creativity Through Combination. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1393–1402. DOI: http://dx.doi.org/10.1145/1978942.1979147

[162] Yupeng Zhang, Teng Han, Zhimin Ren, Nobuyuki Umetani, Xin Tong, Yang Liu, Takaaki Shiratori, and Xiang Cao. 2013. BodyAvatar: Creating Freeform 3D Avatars Using First-person Body Gestures. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*. ACM, New York, NY, USA, 387–396. DOI: http://dx.doi.org/10.1145/2501988.2502015

[163] Zhenpeng Zhao, Sriram Karthik Badam, Senthil Chandrasegaran, Deok Gun Park, Niklas L.E. Elmqvist, Lorraine Kisselburgh, and Karthik Ramani. 2014. skWiki: A Multimedia Sketching System for Collaborative Creativity. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1235–1244. DOI: http://dx.doi.org/10.1145/2556288.2557394

[164] Clement Zheng, Ellen Yi-Luen Do, and Jim Budd. 2017. Joinery: Parametric Joint Generation for Laser Cut Assemblies. In *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition - C&C '17*. ACM Press, Singapore, Singapore, 63–74. DOI:http://dx.doi.org/10.1145/3059454.3059459