Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Computer Science

Marte Hoff Hagen

Creativity as a Quality Factor in Software for Children

Master's thesis in Master of Science in Informatics
Supervisor: Maria Letizia Jaccheri
Co-supervisor: Daniela Soares Cruzes
June 2021

NTNU
Norwegian University of Science and Technology

ablenagic

Marte Hoff Hagen

Creativity as a Quality Factor in Software for Children

Master's thesis in Master of Science in Informatics
Supervisor: Maria Letizia Jaccheri
Co-supervisor: Daniela Soares Cruzes
June 2021

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Computer Science

NTNU
Norwegian University of
Science and Technology

Abstract

Motivation: Creativity is linked to innovation and growth in society and is considered one of the best abilities for children. Thus, any software development team targeting children should be aware of the need to foster children's creativity. However, this aspect is often lacking in the current software development processes.

Objective: This research aims to explore how to improve the development process of software for children by including creativity as a quality factor.

Methods: A systematic literature review was performed to investigate the research literature of Creativity Evaluation Methods of children's digital Creativity Support Tools from the last 10 years. The findings of the literature review were combined with insight into the software company Ablemagic to develop and evaluate a tailored Creativity Evaluation Method.

Results: The systematic literature review identified 81 peer-reviewed articles from the last 10 years showing the wide variations of Creativity Evaluation Methods. The insight in Ablemagic indicated they already had a good Creativity Evaluation Method in the company, and the tailored Creativity Evaluation Method of Ablemagic had good usefulness, intuitiveness, and compatibility rate. The tester in Ablemagic integrated the Creativity Evaluation Method into the company's evaluation routines.

Contribution: This research contributes to researchers and practitioners in the choice of Creativity Evaluation Method to investigate creativity as a quality factor in software targeting children.

Limitations: The limitations of this research comprise the choice of methods, the possibility of generalizing the findings, and the COVID-19 restrictions in the research period.

Conclusions: Based on the findings from the systematic literature review and the evaluation of the tailored Creativity Evaluation Method of Ablemagic, recommendations were made for Creativity Evaluation Methods of digital Creativity Support Tools for children and for integrating evaluation of creativity in the software development process. Future research can use these recommendations to provide more knowledge regarding integrating creativity as a quality factor in agile software development.

Key words: *creativity*, *creativity support tool*, *creativity evaluation method*, *children*, *child-computer interaction*, *quality factor*, *software engineering*, *systematic literature review*.

Sammendrag

Motivasjon: Kreativitet er forbundet med innovasjon og sammfunnsvekst og regnes som en av de viktigste egenskapene hos barn. Derfor bør alle som utvikler programvare rettet mot barn være klar over behovet for å forbedre barns kreativitet. Dette mangler imidlertid ofte i dagens programvareutviklingsprosesser.

Målsetninger: Denne forskningen ønsker å undersøke hvordan utviklingsprosessen av programvare for barn kan forbedres ved å inkludere kreativitet som en kvalitetsfaktor.

Metode: Det ble gjennomført et systematisk litteratursøk av forskningslitteratur fra de siste 10 årene som benyttet metoder for evaluering av kreativitet i digitale kreativitetsstøtteverktøy beregnet for barn. Funn i litteraturgjennomgangen ble kombinert med innsikt i programvareselskapet Ablemagic for å utvikle og evaluere en skreddersydd kreativitetsevaluaeringsmetode.

Resultater: Det systematiske litteratursøket identifiserte 81 fagfellevurderte artikler fra de siste 10 årene som viste en stor variasjon av kreativitetsevaluaringsmetoder. Innsikten i Ablemagic indikerte at de hadde en allerede god kreativitetsevalueringsmetode i selskapet, og den skreddersydde kreativitetsevalueringsmetoden for Ablemagic fikk god nytte-, intuitivitets-, og kompabilitetsvurdering. Testeren i Ablemagic integrerte kreativitetsevalueringsmetoden i selskapets evalueringsrutiner.

Bidrag: Denne forskningen hjelper forskere og programvareutviklere i valg av kreativitetsevalueringsmetode for å undersøkelsen kreativitet som en kvalitetsfaktor i programvare rettet mot barn.

Begrensninger: Begrensningene til denne forskningen er knyttet til valg av metode, muligheten til å generalisere resultatene og koronarestriksjonene i forskningsperioden.

Konklusjon: Basert på funnene fra det systematiske litteratursøket og evalueringen av den skreddersydde kreativitetsevalueringsmetoden for Ablemagic ble det lagd anbefalinger for kreativitetsevalueringsmetoder for digitale kreativitetsstøtteverktøy for barn og for å integrere kreativitetsevaluering i programvareutviklingsprosessen. Videre forskning kan bygge på disse anbefalingene for å gi mer kunnskap om integrering av kreativitet som en kvalitetsfaktor i smidig programvareutvikling.

Nøkkelord: *kreativitet*, *kreativitetsstøtteverktøy*, *kreativitetsevalueringsmetode*, *barn*, *barn-maskin-interaksjon*, *kvalitetsfaktor*, *programvareutvikling*, *systematisk litteratursøk*.

Acknowledgment

This master thesis was written at the Department of Computer Science at the Norwegian University of Science and Technology (NTNU) from September 2020 to June 2021, as part of the course *IT3906 - Informatics Postgraduate Thesis: Interaction Design, Game and Learning Technology*.

Firstly, I want to thank my supervisor, Professor Letizia Jaccheri, and co-supervisor, Professor Daniela Soares Cruzes, for their great guidance and engagement throughout this year. This master thesis had not been possible without them.

Further, I would like to express my gratitude to the rest of the master group in the Children-ByDesign project for discussing and sharing their reflections regarding this master thesis in our weekly meetings. This includes Post Doctoral Fellow Juan Carlos Torrado Vidal and the master students Anna Kristine Roarsen, Andreas Kristiansen Melzer, and Betina Pedersen Bye.

This master project was done in cooperation with the software company Ablemagic. I want to extend my greatest appreciation to this collaboration, especially Simone Bak Laursen and Hallgeir Løkken who invited me to relevant meetings and other events in Ablemagic that were relevant for this thesis.

Last but not least, I am grateful to Ph.D. student Farzana Quayyum for her inspiration in the systematic literature review included in this thesis, Ph.D. student Nektaria Kaloudi for helping me submitting the systematic literature review to the journal *Computing Surveys*, and family and friends contributing with proof-reading and support.

Table of Contents

List of Figures

List of Tables

Abbreviations

AUT Alternative Uses Task. 11, 26, 37

C exclusion criteria. 19

CAD computer aided design. 33

CAQ Creativity Achievement Questionnaire. 12

CCI child-computer interaction. 2, 7, 8, 19, 22, 62

CEM Creativity Evaluation Method. 2–9, 11–15, 17–19, 23, 26, 27, 29–32, 34, 36–43, 47, 50, 51, 53–60, 62, 63, 65–70, 73, 74, 95

CRC United Nations Convention on the Rights of the Child. 2, 14

CSI Creativity Support Index. 13, 15, 29, 37

CST Creativity Support Tool. 2–10, 12–15, 17, 18, 27–29, 31, 37–51, 55–63, 65–69, 73, 95

GQM Goal Question Metric. 53, 56, 62

HCI human-computer interaction. 2, 13, 19, 22

IMC ideation metrics of curation. 13, 37

IT information technology. 4, 5, 53, 54, 62

K-DOCS Kaufman Domains of Creativity Scale. 12

MPCU Model of Personal Computer Utilization. 54, 63

MSFM Multidimensional Stimulus Fluency Measure. 12, 15

NSD Norwegian Centre for Research Data. 6, 8

NTNU Norwegian University of Science and Technology. III, 6, 7

PCI Perceived Characteristics of Innovating. 53, 54

QA quality assessment. 20, 21

R recommendation. 14, 38–40

RAT Remote Associates Test. 12, 37

REC recommendation. 40, 51, 66, 73

RECO recommendation. 67, 74

RQ research question. 3–5, 8, 18, 19, 21, 22, 40, 41, 65, 68–70, 73, 74

SE software engineering. 1, 2, 7, 8, 19, 22, 42, 62, 70, 74

SLR systematic literature review. 4, 8, 17–19, 38, 40, 42, 43, 50, 51, 62, 65–70, 73

TAM Technology Acceptance Model. 53, 54, 63

TIPS theory of inventive problem solving. 13

TPB Theory of Planned Behavior. 54, 63

TTCT Torrance Tests of Creative Thinking. 11, 12, 15

VR virtual reality. 33

Chapter 1

Introduction

This master thesis is a part of a project called ChildrenByDesign, led by Letizia Jaccheri. The project aims to help software companies make better software by creating design principles for developing software targeting children's rights and needs. Jaccheri suggests, together with Morasca, a quality model for children as software engineering (SE) stakeholders comprising the four qualities *fun*, *well-being*, *creativity*, and *security* [1], illustrated in figure 1.1. Each of these qualities could be divided into sub-characteristics. Creativity could be divided into the process of creating products and the solving of problems by exploring solutions. The focus of this thesis is the process part of creativity as a quality factor in software for children.

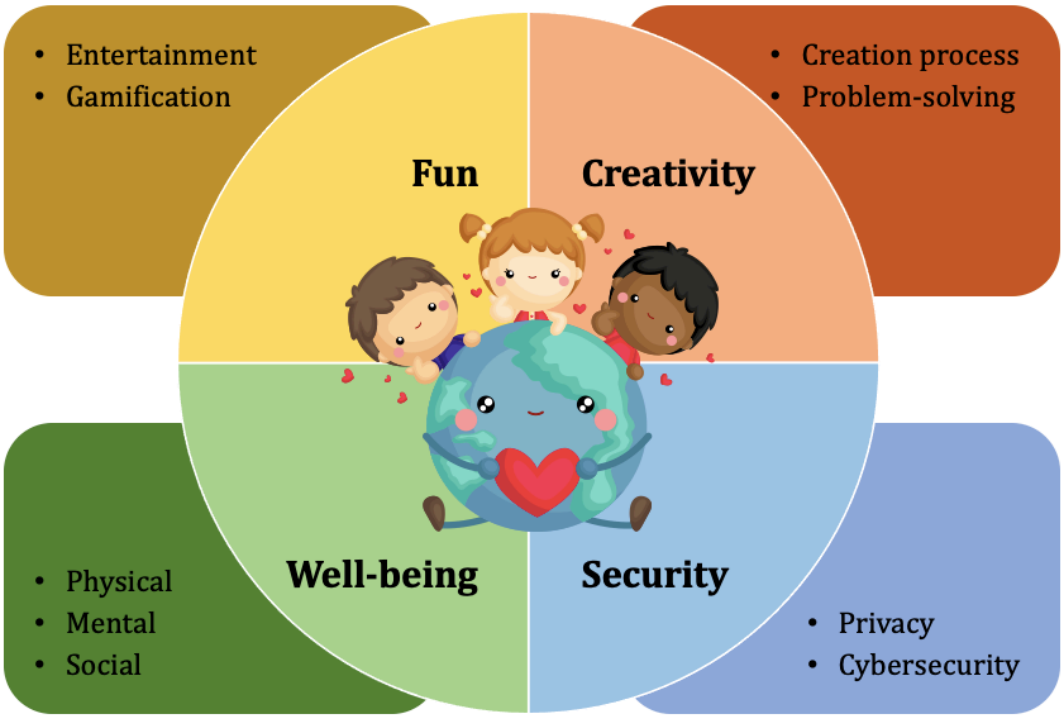


Figure 1.1: Quality model for children as software engineering stakeholders, adapted from Jaccheri and Morasca [1].

1.1 Motivation

Creativity is connected to growth and innovation in society [3] and is considered one of the best cognitive abilities for humans [4]. Creativity is a rather beneficial quality for children, contributing to individual personal development and improves their computer science, programming, language, play, storytelling, and problem-solving skills [5; 6; 7; 8]. It is central in United Nations Convention on the Rights of the Child (CRC) Article 31 regarding play and one of the seven core digital skills for students of the 21st century [9; 10].

According to Jaccheri and Morasca, an increasing amount of software are available for children [1]. Any software developers targeting children should be aware of the need to foster their creativity. To accomplish this, one could start investigating digital Creativity Support Tools (CSTs), which is the software supposed to improve children's creativity [11]. CSTs are difficult to evaluate since there is no single Creativity Evaluation Method (CEM) (i.e., method that evaluates their creativity support) [12]. As claimed by Resnick [13], the CEMs of CSTs have unusual challenges for the human-computer interaction (HCI) field because creativity depends strongly on human factors. Besides, technology for children is often intended for entertaining rather than creative support [14].

Child-computer interaction (CCI) is a subfield of HCI and, in similarity to SE, a computer science discipline. The discipline strives to involve children actively in developing technologies contributing to their creative, social, and intellectual growth [15]. Involving children in SE processes gives new perspectives [16], but collaboration with children can be challenging [17]. Thus, this thesis is motivated by the thought it is better to improve the SE process by making the software developers aware of creativity as a quality factor for children.

Current practice in the software development process with the software product quality model ISO/IEC 25010:2011 [18] does not include specific needs to ensure the quality of child-specific software. Creativity is not a quality of the software product the development team is aware of during the process. The research of creativity in the software development process targets the creative ability of the developers [19]. With the current development pace of software features, the whole software development team should be aware and working actively on certain that the final product addresses the children's need to improve their creativity.

1.2 Research Objective and Questions

This research aims to explore how to improve the software development process of software for children by including creativity as a quality factor. This information will be found by answering the following research questions (RQs):

- **RQ1:** How to evaluate digital Creativity Support Tools for children?
  - **RQ1.1:** What are the current Creativity Evaluation Methods of digital Creativity Support Tools for children?
    - **RQ1.1.1:** Which factors have been considered in the Creativity Evaluation Methods?
    - **RQ1.1.2:** In which context are the Creativity Evaluation Methods performed?
    - **RQ1.1.3:** How does the age of the target group impact the Creativity Evaluation Methods?
    - **RQ1.1.4:** How effective are the Creativity Evaluation Methods?
  - **RQ1.2:** Which changes are needed in the Creativity Evaluation Methods to evaluate digital Creativity Support Tools for children better?
- **RQ2:** How to integrate evaluation of creativity in the software development process targeting children?

1.3 Context

A specific case of a CEM of digital CSTs for children is the method used in Ablemagic[1], a Norwegian software company with long experience producing digital CSTs for children with a playful approach. The interdisciplinary software development team in Ablemagic comprises 10 different employees divided into developers, designers, concept developers, and one business administrator. One of the designers also has the role of a tester. A previous master thesis regarding Ablemagic by Jan-Magnus Neverdal stated creativity should be a starting point in designing digital play [20]. Thus, Ablemagic is aware of creativity as a quality factor in software for children, but they still want to improve their products. Together with the subsidiary company Munti Magic[2], Ablemagic wants to improve the software development process to target their audience more accurately by evaluating, analyzing, and logging. One possibility to improve this is by looking into the creative aspect of the software.

[1] ablemagic.no
[2] muntimagic.com

1.4 Research Method

Figure 1.2 presents the overview of the research method highlighted in bold outlines. To address **RQ1.1**, a systematic literature review (SLR) was performed to get an overview of the current CEMs of digital CSTs for children. An SLR is an auditable, trustworthy, and rigorous methodology identifying, summarizing, interpreting, and evaluating all available relevant research in the focus area [21]. Chapter 3 describes the SLR, which also helped decide a conceptual framework for further research and, together with experience and motivation, select the RQs. The next section describes the research strategies, while details regarding the data collection methods and data analysis are presented in chapters 4 and 5.

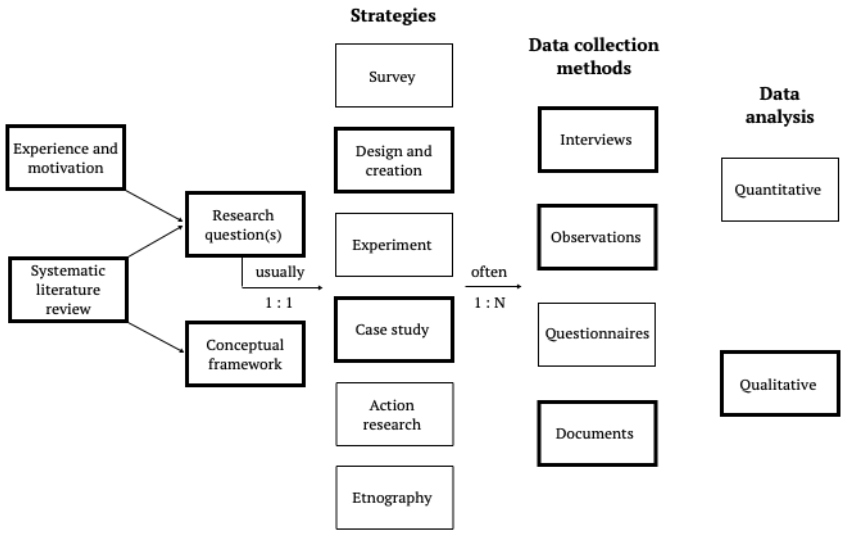


Figure 1.2: Overview of the research method, adapted from Oates [2].

1.4.1 Research Strategies

To address the remainder of the RQs, the research strategy design and creation was utilized. The strategy focuses on developing new information technology (IT) artifacts, using a fluid, iterative process of the steps awareness, suggestion, development, evaluation, and conclusion [2]. In this instance, the new IT artifact is a CEM of digital CSTs for children, illustrated in figure 1.3. Firstly, the development process went from awareness of the problem to unspecific suggestions for CEMs based on the SLR. Further, a more specific tailored CEM for the software company Ablemagic was developed from these recommendations and insight into the company. This development addressed **RQ1.2** regarding needed changes in the current CEMs and, together with the SLR, **RQ1** regarding the evaluation of digital CSTs for children. More details regarding the development of the tailored CEM are described in chapter 4.

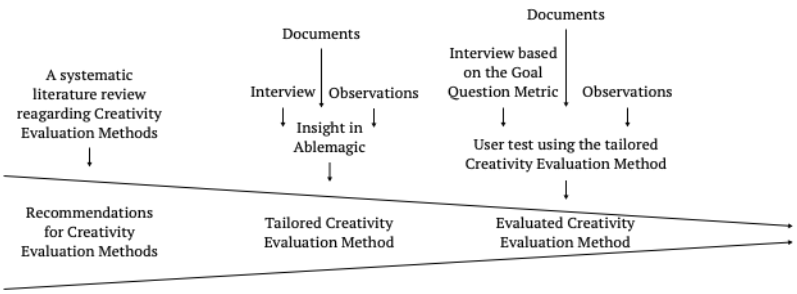


Figure 1.3: Overview of the development of the Creativity Evaluation Method.

Moreover, the research strategy case study was utilized to evaluate the tailored CEM, so the CEM becomes even more specific, and the process concludes. It is normal to combine a design and creation study with a case study evaluating the IT artifact in a real-world situation instead of an artificial university situation [2]. A case study focuses on one instance of the specific CEM of digital CSTs for children [2]. This thesis utilized a longitudinal case study focusing on obtaining a rich, detailed insight into the CEM of the software company Ablemagic to answer **RQ2** regarding integrating evaluation of creativity in the software development process. More details regarding the evaluation of the tailored CEM are described in chapter 5.

Another relevant research strategy was action research. This strategy focuses on doing research into action by conducting cycles of plan-act-reflect in a real-world situation [2]. This research aimed to improve the CEM of CSTs for children of the participating software company Ablemagic, but it was not conducted plan-act-reflect cycles.

1.4.2 Research Paradigm

The interpretivism paradigm is most suited for this research since this master thesis tries to identify, explain, and explore the case of the CEM of digital CSTs for children for a specific software company. The focus in this paradigm is understanding the social context [2]. On the other hand, the positivism paradigm regarding objectively investigating the regular and ordered world [2] is unsuitable because it is no hypothesis. The critical research paradigm is inappropriate as it focuses on identifying and reducing conflicts, contradictions, and power relations [2].

1.4.3 Participants

Software engineers in Ablemagic participated in this research, both directly and indirectly. After the data analysis, they got access to the relevant parts to check if the information regarding

Ablemagic was correct and could be published. Besides, children indirectly participated in this research since Ablemagic includes them as the target group in their CEM of digital CSTs for children. It was not directly collected any data of the children in this master thesis due to the ethical issues for research with children as participants regarding confidentiality, consent, and protection [22]. The focus of the observations of the tailored CEM of Ablemagic was to observe how the software engineers in Ablemagic used the CEM. Information regarding the children was only mentioned in the thesis in cases they were relevant for the CEM. Consciously, the pictures taken during the user tests were not including any persons.

1.4.4 Ethics

It is important to consider ethical issues in empirical research [2]. Since this thesis mainly collected qualitative data of the participants' words and actions, protecting the participants' rights was important. Thus, all participants in this master thesis were anonymized. Furthermore, it was followed NTNU's guidelines [23] and sent a notification form to the Norwegian Centre for Research Data (NSD) [24] to get an ethical assessment of the project because it processed digital media files by interview recording. The NSD application addressed many of the ethical issues regarding collecting, storing, and processing data. However, the background information mentioned in the application was decided not to collect. The application was approved by NSD on the 29th of October 2020 and could be found in Appendix A.

The project obtained voluntary, informed, specific, documented, and unambiguous consent from the interviewee to process personal data, following the requirements in the Personal Data Act Articles 4 and 7 [25]. The consent form attached in Appendix B informed the interviewee of the nature of the interviews and the possibility of withdrawing from the research. The consent was a legal basis for processing, cf. the Personal Data Act Article 6 no. 1 letter a [25]. The processing met the law's requirements for form and content, cf. Article 12.1 and 13 [25], as well as the requirements for accuracy (Article 5.1 d [25]), integrity and confidentiality (Article 5.1 f [25]), and security (Article 32 [25]).

The processing of personal data followed the principles of the Personal Data Act [25] in:

- Legality, fairness, and transparency (Article 5.1 a), because the data subjects receive satisfactory information about and consent to the processing;
- Purpose limitation (Article 5.1 b) since personal data is collected for specific, explicitly stated, and justified purposes and not processed for new, incompatible purposes;
- Data minimization (Article 5.1 c) as the only information that is adequate, relevant, and necessary for the project is processed;

- Storage restriction (Article 5.1 e) seeing that the personal data is not stored longer than necessary to fulfill the purpose.

As long as the participants could be identified in the data material, they had the rights to transparency (Article 12 [25]), information (Article 13 [25]), access (Article 15 [25]), correction (Article 16 [25]), deletion (Article 17 [25]), restriction (Article 18 [25]), notification (Article 19 [25]), and data portability (Article 20 [25]).

1.4.5 Storing of Data

Microsoft OneDrive was used to store the collected personal data confidentiality, as NTNU has a data processor agreement with Microsoft [26]. This met the requirements for using a data processor, cf. the Personal Data Act's Articles 28 and 29 [25]. As I was the person that conducted the research and analyzed the data, I was the only one with access to the recording of the interviews. Every service in Microsoft OneDrive was protected with a password. The personal data will be deleted after this master thesis is submitted.

1.5 Contribution

There exists numerous research on creativity, CCI, and SE, but the intersection of including creativity as a quality factor in software for children is scarce. This thesis investigates this knowledge gap, aiming to produce valuable information regarding the intersection between creativity, CCI, and SE.

The possible impact of the research will be as following:

- Starting the investigation of creativity as a quality factor;
- A new CEM assisting practitioners in software companies improving their software developing process of digital CSTs for children;
- Software improving children's creative skills;
- Two published articles.

1.6 Outline

This thesis comprises seven chapters. Chapter 1 has introduced the motivation, the objectives, the context, the method, and the contribution of the research. The thesis proceeds with the background regarding creativity, children, and software quality factors in chapter 2. Further, chapter 3 describes the specific method, results, and discussion of the SLR summarizing the state-of-the-art CEMs of digital CSTs for children. Similarly, chapters 4 and 5 present the specific method, results, and discussion of the development and the evaluation of the tailored CEM of Ablemagic, respectively. Moreover, chapter 6 discusses the thesis in its entirety in the form of the RQs, the implications for research and practice, the limitations, and lessons learned. In the end, chapter 7 concludes the thesis by answering the RQs, summarizing the limitations and contribution, and presents suggestions for future work.

Furthermore, the appendices comprise the mentioned NSD application and consent form in Appendix A and B, respectively. The interview guides of the interviews conducted during the development and evaluating of the tailored CEM could be found in Appendix C. A manuscript regarding the SLR of CEMs of digital CSTs for children was submitted to *Computing Surveys* in Mars. This journal has an impact factor of 8.96 [27]. The article belongs to the intersection between creativity and CCI and could be found in Appendix D. A conference paper summarizing the evaluation of the tailored CEM of Ablemagic was submitted to the *14th International Conference on the Quality of Information and Communications Technology* at the end of April. The article belongs to the intersection between creativity and SE and could be found in Appendix E. Right before the deadline of this master thesis, the paper was unfortunately not accepted. Further, the paper will be resubmitted to the *44'th International Conference on Software Engineering: Software Engineering in Society*, which has the best conference ranking (A^*) [28].

Chapter 2

Background

This chapter will first introduce the term creativity, including CSTs and CEMs. Further, multiple CEMs are presented before CEMs used for CSTs are discussed. Moreover, the chapter connects these theories to children. Lastly, theories regarding software quality factors are introduced.

2.1 Creativity

As stated by Ritter and Rietzschel [29], the English word *creativity* originates from the Latin word *creō*, meaning "to create, to make" [29, p. 97]. The term evolving from psychology theory is defined in various ways [30]. Oxford Dictionary defines it as "the use of skill and imagination to produce something new or to produce art" [31]. This definition resembles the common definition of creativity as the ability to foster problem solutions or ideas that are both *valuable* (i.e., perceived functional or useful) and *novel* (i.e., perceived original or unique) [32; 33; 34]. By analyzing over 160 definitions of creativity, Sarkar and Chakrabarti proposed the common definition: "Creativity occurs through a process by which an agent uses its ability to generate ideas, solutions or products that are novel and valuable" [35, p. 11].

Jordanous and Keller [36] expanded the components *valuable* and *novel* to 14 components by examining the meaning of the word creativity. The remaining components were *active involvement*, *collaboration*, *decision-making*, *dealing with uncertainty*, *domain-specific competence*, *freedom*, *general intellect*, *generation of results*, *immersion*, *progression*, *spontaneity*, and *variety*. Weisberg suggested defining creativity as "intentional novelty" [37, p. 119] because he did not saw *valuable* as appropriate since what society thinks is *valuable* changes over generations. Another approach of creativity by Vygotsky [38] focuses on the process of revising old memories to construct these novel behaviors or ideas.

2.1.1 Creativity Support Tool

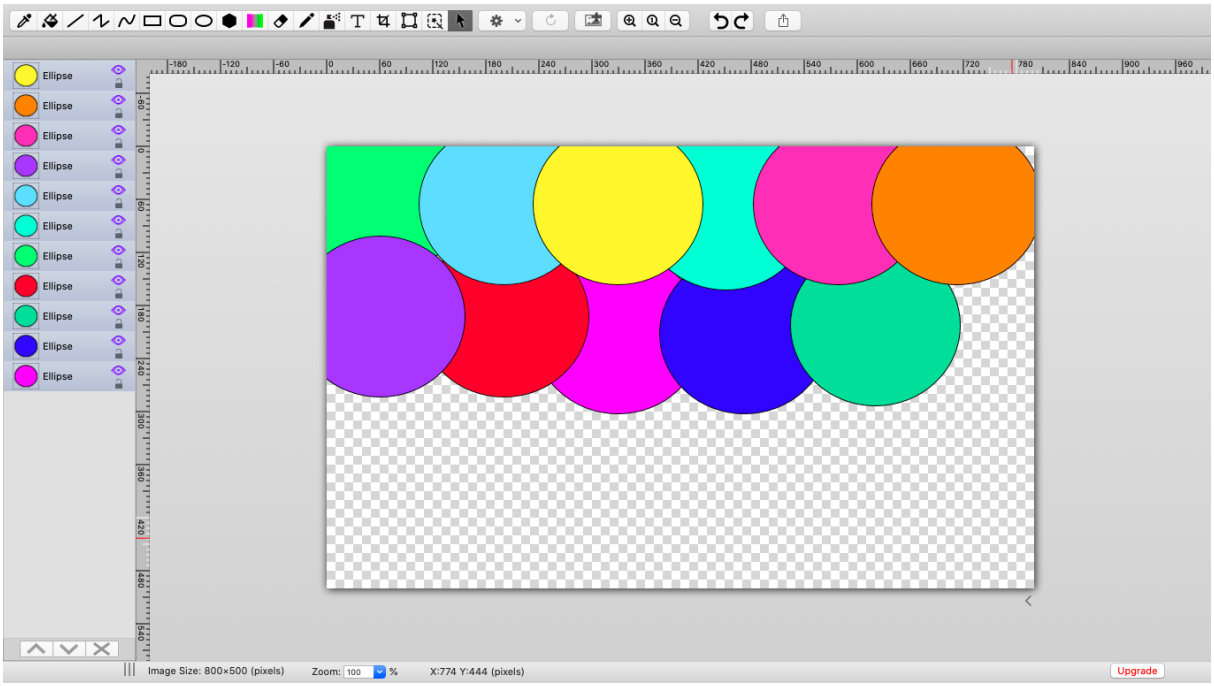


Figure 2.1: Screenshot of the digital Creativity Support Tool Paint.

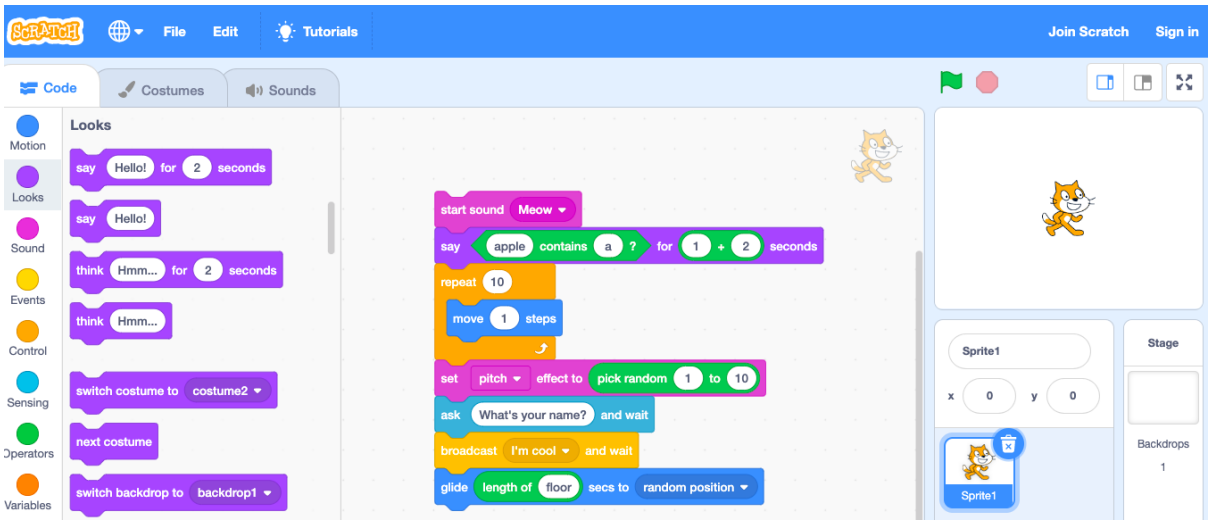


Figure 2.2: Screenshot of the digital Creativity Support Tool Scratch.

Cherry and Latulipe define a CST as "any tool that can be used by people in the open-ended creation of new artifacts" [39, p. 2]. CSTs are often apps used to create or complete digital artifacts in computer science [39]. Tack and Jasperneite define an app as "a small piece of software, that provides a specific functionality and is able to use device features. Furthermore, the app is totally integrated into a distribution system, that provides benefits for device suppliers, developers and customers" [40, p. 2]. The focus on CSTs in this thesis is on apps. There exist a

multitude of such apps. Examples are the drawing program Paint and the visual programming language Scratch, as pictured in figures 2.1 and 2.2, respectively.

2.1.2 Creativity Evaluation Methods

There exist a wide specter of CEMs. As stated by Carroll and Latulipe [41], the CEMs could be grouped into three approaches: *psychophysical measurement* using biometrics, *self-reporting* by users, and *external judges*. Self-reporting is a subjective CEM because the user does the evaluation. The CEM external judges is more objective, utilizing others to evaluate. Psychophysical measurement is the most objective CEM observing biometrics like the user's behavior with the product. Table 2.1 presents an overview of the approaches to the CEMs that will be discussed later in this section.

Table 2.1: An overview of the approaches of the existing Creativity Evaluation Methods.

Psychophysical measurement	Self-reporting	External judges
• Alternative Uses Task • Torrance Tests of Creative Thinking • Abbreviated Torrance test for adults • Multidimensional Stimulus Fluency Measure • Remote Associate Test	• Likert scale • Creativity Achievement Questionnaire • Kaufman Domains of Creative Scale • AttrakDiff • This or That • Smileyometer • Fun-Sorter • Memoline	• Theory of inventive problem solving • Consensual Assessment Technique • Creative Product Semantic Scale
• Ideation metrics of curation • Creativity Support Index		

Psychophysical measurement

The oldest CEM is the psychophysical measure Alternative Uses Task (AUT). Guilford established modern creativity research in 1950, developing the divergent thinking test AUT based on cognitive psychology [42]. The AUT reveals how many alternative uses a participant managed to mention for an object within a given time [43].

Torrance built on evaluating creativity centered on the ability of divergent thinking. He proposed the psychophysical measurement Torrance Tests of Creative Thinking (TTCT) based on the following factors of problem responses by humans: *elaboration* (i.e., the number of details in the answers), *flexibility* (i.e., the number of different categories in the answers), *fluency* (i.e., the number of relevant answers), and *originality* (i.e., the statistical rarity of the answers) [44]. TTCT is a standardized evaluation of children's creativity [45], comprising a verbal test of writing and a figural test of drawings [46]. Based on the principles of TTCT, it was made the

abbreviated Torrance test for adults [47]. However, TTCT has been criticized for being uncertain by overestimating the factor *elaboration* [48].

The Multidimensional Stimulus Fluency Measure (MSFM) is a similar psychophysical measurement, evaluating the factors *originality* and *fluency* [49]. Since the test uses visual and tactile stimuli and requires verbal responses, it can be applied to young children [49]. In contrast, the psychophysical measurement Remote Associates Test (RAT) is a convergent creativity test where the participants should find the fourth word all three given words are semantically related to [50].

Self-reporting

The Creativity Achievement Questionnaire (CAQ) is a self-reporting questionnaire of creativity scores across the domains of architecture, culinary arts, dance, humor, inventions, music, scientific discovery, theater, visual arts, and writing [51]. The five-point Likert scale Kaufman Domains of Creativity Scale (K-DOCS) evaluating the domains academic, artistic, everyday, mechanical, performance, and science is relatable to CAQ [52]. A Likert scale is a self-reporting questionnaire where the participant evaluates the level of agreement of a set of statements on a scale ranging from strongly disagree to strongly agree [53]. For domain-specific creativity, K-DOCS has been proved to be a reliable and valid CEM [54]. On the other hand, AttrakDiff is an acceptance self-reporting CEM for products, evaluating the factors *aesthetic*, *emotional experience*, and *functionality* on a seven-point scale [55]. In contrast to the Likert scale, the user evaluates how the product corresponds between two opposites (e.g., “confusing-clear”) instead of statements.

There exist several self-reporting CEMs designed for children. An example is the self-reporting preference CEM This or That where the child compares two different options and selects the preferred one [56; 57]. Another established CEM for children is the self-reporting acceptance scale Smileyometer [58], a five-point Likert scale of smileys representing the perceived level of *enjoyment* [59]. Figure 5.3 illustrates examples of the CEMs Smileyometer and This or That. The Smileyometer is similar to Fun-Sorter, which ranks statements of connected activities based on the perceived level of *fun* [60]. Another child-friendly self-reporting CEM is the MemoLine, evaluating the long-term user experience by letting the child color a timeline based on the experience with a CST [61].

External judges

The theory of inventive problem solving (TIPS) uses external judges to compare the process with a systematic human-oriented knowledge-based method of inventive problem-solving [62]. On the other hand, the Consensual Assessment Technique utilizes external expert judges to evaluate the creativity of the products based on their subjective perception [32]. Another CEM using external judges is the Creative Product Semantic Scale, which evaluates the factors *elaboration and synthesis* (i.e., aesthetic and beauty), *resolution* (i.e., relevance, logic, and usefulness), and *novelty* (i.e., intuitiveness and originality) [63; 64; 65]. Others have simplified these three factors to *valuable* and *novelty* [66; 67; 68].

According to Hocevar, an issue using external judges is their non-distinguishing between aesthetics, technical skills, and creativity [69]. As claimed by Glăveanu [70], this judgment depends on the press because the same item could be evaluated variously by members of different communities.

Creativity Evaluation Methods of Creativity Support Tools

Different evaluation factors are utilized in the CEMs of CSTs. Kerne et al.'s [71] self-reporting psychophysical measurement called ideation metrics of curation (IMC) combined observations and self-reporting questionnaires to evaluate the support of creative engagement by CSTs in idea development. The CEM utilized the factors *emergence*, *exposition*, *flexibility*, *fluency*, *novelty*, *relevance*, and *visual presentation* as a basis [71]. The IMC has some similarities with the standardized quantitative psychometric questionnaire Creativity Support Index (CSI) developed by Cherry and Latulipe [39], which evaluates a CST's ability to assist a user's engagement in creative work. This is evaluated through the factors *collaboration*, *immersion*, *enjoyment*, *exploration*, *expressiveness*, and *results satisfaction*. The CSI comprises a rating scale section of two agreement statements per factor and a paired-factor comparison, enabling quantifiable and comparable results.

Remy, Vermeulen, Frich, Biskjaer, and Dalsgaard recently performed a relevant in-depth literature review of CEMs of CSTs in HCI [72]. By providing an objective description of the CEMs of the most leading CSTs in HCI research, the literature review aimed to improve the strategies for developing CSTs. The literature review reused the sample of the authors' earlier in-depth literature review of CSTs in HCI research [73], which used the search string *"creativity" OR "creativity support tool"* in the *ACM Digital Library* and reduced the sample size based on the above-average citations per year. Remy et al. selected all papers in this corpus that evaluated a CST and used the same method for recent papers, where the reducing of the sample size was

based on the average download count per year [72].

Grounded on the findings, Remy et al. [72] stated six recommendations (Rs) for CEMs of CSTs:

- **R1)** Defining a clear goal of the CEM because multiple studies are unclear regarding this goal;
- **R2)** Using theoretical grounding in the CEM since this is lacking in current CEMs;
- **R3)** Deciding whether to evaluate usability or creativity as numerous of the CEMs focus on usability instead of creativity;
- **R4)** Considering utilizing longitudinal in-situ studies because short-term controlled CEMs are prioritized;
- **R5)** Recruiting domain experts if appropriate since the CEMs lack expert participants;
- **R6)** Contributing to developing a toolbox for CEMs of CSTs as it is non-existing.

Lamb, Brown, and Clarke [74] spotlighted this need for a toolbox in their interdisciplinary tutorial on evaluating computational creativity, referring to using the existing CEM standards of CSTs in co-creative system evaluations.

2.2 Children

The international legal framework CRC Article 1 defines a child as a "human being below the age of eighteen years unless under the law applicable to the child, majority is attained earlier" [9]. This framework is officially approved by 192 of the 194 member countries and comprises 54 different articles declaring all children as individual rights-holders and protecting their rights.

Jaccheri and Morasca [1] recommend software developers designing software for children to consider each of the four qualities in figure 1.1 in each phase of the software development process and include children and caregivers as much as possible in the software development process. Furthermore, Jaccheri and Morasca [1] provided five guidelines for researchers:

1. Studying from the point of view of both children and software engineers;
2. Including investigating other aspects than just technical aspects;
3. Systematically collect and analyze empirical data of software engineers who develop software for children;
4. The data collection should be implemented for specific goals for each of the qualities regarding the software, the software development process, the intention of the different types of stakeholders, and relations;
5. Defining the knowledge and education needed for children to be capable of effectively participate in software development processes.

To evaluate CSTs designed for children, it could be useful to better understand different aspects of children's perspectives [75]. Some of the CEMs in section 2.1.2, like the CSI, could be too complicated to use for children. On the other hand, TTCT and MSFM could evaluate a child's creativity skills but could be difficult to use for evaluating CSTs for children. It could be better to look at the multiple self-reporting CEMs designed for children as the preference CEM This or That. As claimed by Guinard [76], the only valid method for children between two and three years old is such paired-preference testing. For older children, it can be examined using preference ranking for more than two options and acceptance CEMs between three- to nine-point [76], like the Smileyometer.

2.3 Software Quality Factors

The multidimensional construct *quality* is reflected in *quality models*, which provide an organized view of different *software qualify factors* believed to be important in evaluating software products [77]. The quality factors could be classified into *internal qualities* (i.e., the quality of the software) and *external qualities* (i.e., the quality of the software perceived by its stakeholders) [78].

There are exclusively external quality factors in the quality model for children provided by Jaccheri and Morasca [1], comprising the four quality factors *fun*, *well-being*, *creativity*, and *security*. The latter factor is also included in the software product quality model ISO/IEC 25010:2011, compromising the following eight quality factors [18]:

- *Functional suitability* (i.e., how the product provides functions that meet the given needs);
- *Performance efficiency* (i.e., how the performance under stated conditions is relative to the number of resources);
- *Compatibility* (i.e., how the product under sharing the same hardware or software environment with other product can exchange information and perform its required functions);
- *Usability* (i.e., how the product could be used for the given purpose);
- *Reliability* (i.e., how the product performs stated functions under stated conditions for a given time);
- *Security* (i.e., how the product protects data and information);
- *Maintainability* (i.e., how the product could be modified);
- *Portability* (i.e., how the product could be transferred from one environment to another).

All these quality factors are external, but some factors like *reliability* use internal quality factors like *error occurrence rates* to measure the external quality [18; 78].

Chapter 3

Systematic Literature Review in Creativity Evaluation Methods

As the literature review by Remy et al. [72] did not examine the age of the target group, there is a need for an SLR targeting the CEMs of digital CSTs for children. This chapter describes this SLR, who also was submitted as a journal article to *Computing Surveys*, attached in Appendix D.

3.1 Method

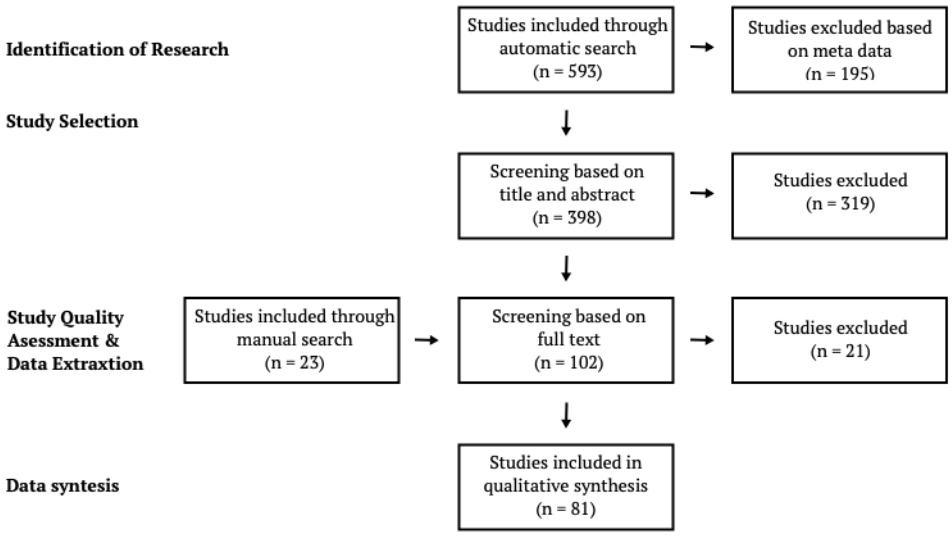


Figure 3.1: The study selection process.

This section presents the method of the SLR, the first step in figure 1.3. The overview of the method of the thesis is presented in figure 1.2. The method of the SLR was based on Kitchenham's original guidelines for performing an SLR in software engineering [21], which comprises the five steps highlighted by the bold font in figure 3.1. Each of these steps will be described in detail in the remainder of this section.

3.1.1 Protocol Development

A review protocol was developed to plan the SLR in the initial phase [21]. The protocol comprises the schedule, the rationale of the SLR, the relevant RQs, the search strategy, the study selection criteria, and the procedures of the assessment of study quality, the data extraction, and the data synthesis.

3.1.2 Identification of Research

The identification of research was the first step of the SLR [21]. To identify the most optimal databases and search string for this SLR, a pilot search was conducted. This search also discovered the related literature review by Remy et al. [72]. The collection of 113 studies in this literature review was added to the corpus.

The automatic literature search was run in October 2020. The search strategy included journal hand search in the *International Journal of Child-Computer Interaction* and the electronic bibliographic databases *IEEE Xplore*, *ACM Digital Library*, and *Scopus* to include as many relevant studies as possible. Table 3.1 presents the documenting of the search.

Table 3.1: Documenting of the automatic search.

Data Source	Search string	Search date	Hits
Recent and relevant literature review [72]	Based on *"creativity" OR "creativity support tool"* in the ACM Digital Library [73]	02.10.20 (pilot search)	113
Scopus	*(TITLE AND ABS)("creativity" AND ("measure" OR "measuring" OR "metrics" OR "evaluate" OR "evaluating") AND ("software" OR "app"))*	09.10.20	259
ACM Digital Library		10.10.20	89
IEEE Xplore		13.10.20	91
International Journal of Child-Computer Interaction	*"creativity" AND ("measure" OR "measuring" OR "metrics" OR "evaluate" OR "evaluating") AND ("software" OR "app")*	16.10.20	41
		Total:	**593**

Despite the focus on children in the RQs, it was intentionally decided not to include the term *"children"* in the search string. This explored if the CEMs of CSTs for adults could be transferable to children.

The query resulted in 593 hits. To manage this multitude of references, the bibliographic package Endnote X9.3.3[1] was utilized.

[1] endnote.com

3.1.3 Study Selection

Further, the relevance of the potential primary studies from the search was evaluated with exclusion criteria (C) [21], presented in table 3.2.

Table 3.2: The exclusion criteria (C).

Exclusion criterion		Number of excluded studies
C1	Publishing date before 2010	157
C2	Duplicated studies	37
C3	Non-English language	1
C4	No Creativity Evaluation Method	319
	Total:	**514**

C1 excluded non-recent studies because of the focus on current CEMs in **RQ1.1**. The lower time limit was set to the year 2010 since this year represented the beginning of the big growth of apps due to Apple's release of its Software Development Kit in 2009 [79]. In **C2**, duplicates were removed in Endnote. If the duplicates had different publishing dates, the most recent one was kept following Kitchenham [21]. Despite Kitchenham's [21] recommendation of avoiding exclusion based on the language, **C3** was included as a criterion since the English abstracts in non-English papers did not give as much data as needed for the analysis. Derived from the RQs, **C4** was added as a criterion. The exclusion in this criterion was determined by reading titles and abstracts.

This study selection resulted in 79 primary studies.

3.1.4 Manual Search

In November 2020, manual searches were conducted in relevant journals in the CCI, HCI, creativity, and SE field to include articles that the automatic search did not discover. Table 3.3 presents the documenting of the manual search.

The manual search added 23 primary studies, which resulted in 102 primary studies in the SLR.

Table 3.3: Documenting of the manual search.

Data Source	Search string	Search date	Hits	Relevant
International Journal of Child-Computer Interaction	*"creativity" AND PUBYEAR ≥ 2010*	16.11.20	60	2
Interaction with Computers			142	2
ACM Transactions on Computer Human Interaction			74	2
AIS Transaction on Human-Computer Interaction			33	0
Computers in Human Behavior			408	8
Computer-Aided Design and Application			155	1
International Journal of Design Creativity and Innovation	*("software" OR "app") AND PUBYEAR ≥ 2010*	18.11.20	39	0
Journal of Creativity behavior			91	2
Thinking Skills and Creativity			155	6
Digital Creativity			183	0
IEEE Transactions on Software Engineering	*"creativity" AND PUBYEAR ≥ 2010*	20.11.20	2	0
IEEE Software			11	0
Transactions on Software Engineering and Methodology			15	0
Empirical Software Engineering			34	0
Information & Software Technology			61	0
Automated Software Engineering			4	0
Requirements Engineering			29	0
Software Quality Journal			13	0
Journal of Systems and Software			96	0
Software Testing, Verification & Reliability			6	0
Journal of Software: Evolution and Process			57	0
Software: Practice & Experience			29	0
International Journal of Software Engineering and Knowledge Engineering			59	0
			Total:	**23**

3.1.5 Assessment of Study Quality

Moreover, it was made quality assessments (QAs) used to evaluate the quality of each primary study determined by full-text reading [21]. Table 3.4 presents the three QA questions each primary study had to pass regarding relevance, rigor, and credibility, adapted from Dybå and Dingsøyr's checklist [80].

Table 3.4: The quality assessments (QAs).

Quality assessment		Number of excluded studies
QA1	Do the study include a Creativity Evaluation Method of a digital Creativity Support Tool?	6
QA2	Is the research method appropriate in terms of using evaluation factors?	15
QA3	Do the presented findings have a validity aspect?	0
	Total:	**21**

To save time, this step of quality assessment was combined with the step of data extraction and

monitoring, presented in section 3.1.6. Overall, the QAs excluded 21 studies, which resulted in **81** included studies in the qualitative data synthesis, described in section 3.1.7.

3.1.6 Data Extraction

To collect all the necessary information of the primary studies to address the QAs and RQs, it was designed a predefined systematic data extraction as proposed by Kitchenham [21]. Table 3.5 presents this data extraction form.

Table 3.5: Data extraction form.
RQ = Research question, QA = Quality assessment

Focus area	Data	Mapping to QAs	Mapping to RQs
Demographics	• Publication channel • Publication year		
Factors	• Factors evaluating creativity	**QA2**	**RQ1.1.1**, **RQ1.1.3**
Products	• Main functionality • Characteristics that improve creativity • Creativity outcome • Software type • Creativity topic	**QA1**	**RQ1.1.2**

Methods	• Data gathering method
	• Data analysis
	• Participants
	  – Sample size
	  – Gender distribution
	  – Age
	• Approach
	• Validity of method
	• Study environment
		QA1, **QA2**	**RQ1.1**, **RQ1.1.2**, **RQ1.1.3**, **RQ1.1.4**

Research	• Objective • RQs • Findings • Outcome	**QA3**	**RQ1.1.4**
Creativity	• Definition of creativity • Unclear words describing creativity	**QA3**	**RQ1.1.4**

The specialist software package for qualitative analysis of textual data MAXQDA Analytics Pro 2020 (Release 20.2.1)[2] was used to extract the data. To ensure the data extraction was done systematically, it was discussed with the supervisors.

[2]maxqda.com

3.1.7 Data Synthesis

To address the RQs, data synthesis was utilized to collect and summarize the findings of the included primary studies [21]. Cruzes and Dybå's [81] five recommended steps for thematic synthesis in SE were used. Section 3.1.6 describes the first two steps related to extracting and systematically coding data from the primary studies. The next two steps translated the codes into themes of different levels and created a higher-order themes model by exploring the relationships between the themes, respectively. This was done in the mind mapping software MindManager 13.0.181[3]. It was used an inductive approach which finds themes and categories based on the data and not existing theory [2]. The last step evaluated the trustworthiness of the interpretation leading to the synthesis by searching for counter-evidence.

3.2 Results

3.2.1 Demographics

Table 3.6 presents the distribution of the publication year and channels for the studies. The full references of the included studies can be found at the end of Appendix D. The publication frequency increased from 2013 until an apex of 12 published studies in 2016 and has decreased thereafter. There were more conference papers than journal articles. Besides, **S20** and **S73** were published as a chapter in a book and a lecture note, respectively. Most of the studies were in the HCI, the creativity, or the CCI field with a total of 44, 15, and 8 studies, respectively. Only one of the publication channels, *Conference on Automated Software Engineering*, was in the SE field.

Table 3.6: Publication frequency.

Year	Journal	Conference	Book
2020	**S21**, **S48**, **S71**		
2019	**S4**, **S11**, **S22**, **S34**, **S56**, **S77**	**S12**, **S13**	
2018	**S19**, **S24**, **S55**, **S63**	**S26**, **S31**, **S43**, **S70**	
2017	**S28**, **S46**, **S58**, **S68**	**S2**, **S3**, **S76**	
2016	**S5**, **S10**, **S27**, **S47**, **S53**	**S15**, **S32**, **S33**, **S45**, **S54**, **S66**, **S80**	
2015	**S18**, **S35**, **S41**, **S62**	**S37**, **S49**, **S59**, **S60**, **S65**, **S79**	**S20**
2014	**S14**, **S36**, **S39**	**S6**, **S23**, **S51**	
2013	**S61**	**S25**, **S67**, **S75**	
2012	**S9**, **S50**, **S57**	**S1**, **S8**, **S29**	**S73**
2011	**S64**, **S74**	**S30**, **S38**, **S40**, **S42**, **S78**, **S81**	
2010	**S16**	**S7**, **S17**, **S44**, **S52**, **S69**, **S72**	
Total:	**36**	**43**	**2**

[3] mindmanager.com

3.2.2 Evaluation Factors

It was identified 53 different evaluation factors in the CEMs of the included studies. Further, these factors were categorized into the 10 most frequently occurring categories covering 90.3% of the factors. The remaining factors were put in an *Other* category. Each of these categories will be investigated in detail in the remainder of this section. The definitions of the 10 factors are presented in table 3.7, while tables 3.8 and 3.9 illustrate the distribution of these categories in the included studies.

Table 3.7: The definitions of the 10 most frequent evaluation factors.

Factor	Definition
Valuable	The user's perceived usefulness of the software.
Novelty	The extent to which the user can create something unique with the software.
Fluency	The degree the user gets relevant ideas from the software.
Enjoyment	The pleasure the user gets from the software.
User feeling	The user's perception of the software.
Collaboration	The extent to which the user works with others on the software.
Expressiveness	The degree the software expresses the user's thoughts or feelings.
Immersion	The extent to which the user is absorbed into the software.
Flexibility	The degree the software can be adapted to suit new circumstances.
Interaction	How the user's movements affect the software.

Table 3.8: Distribution of factors, products, approaches (appr.), and methods in the included studies.

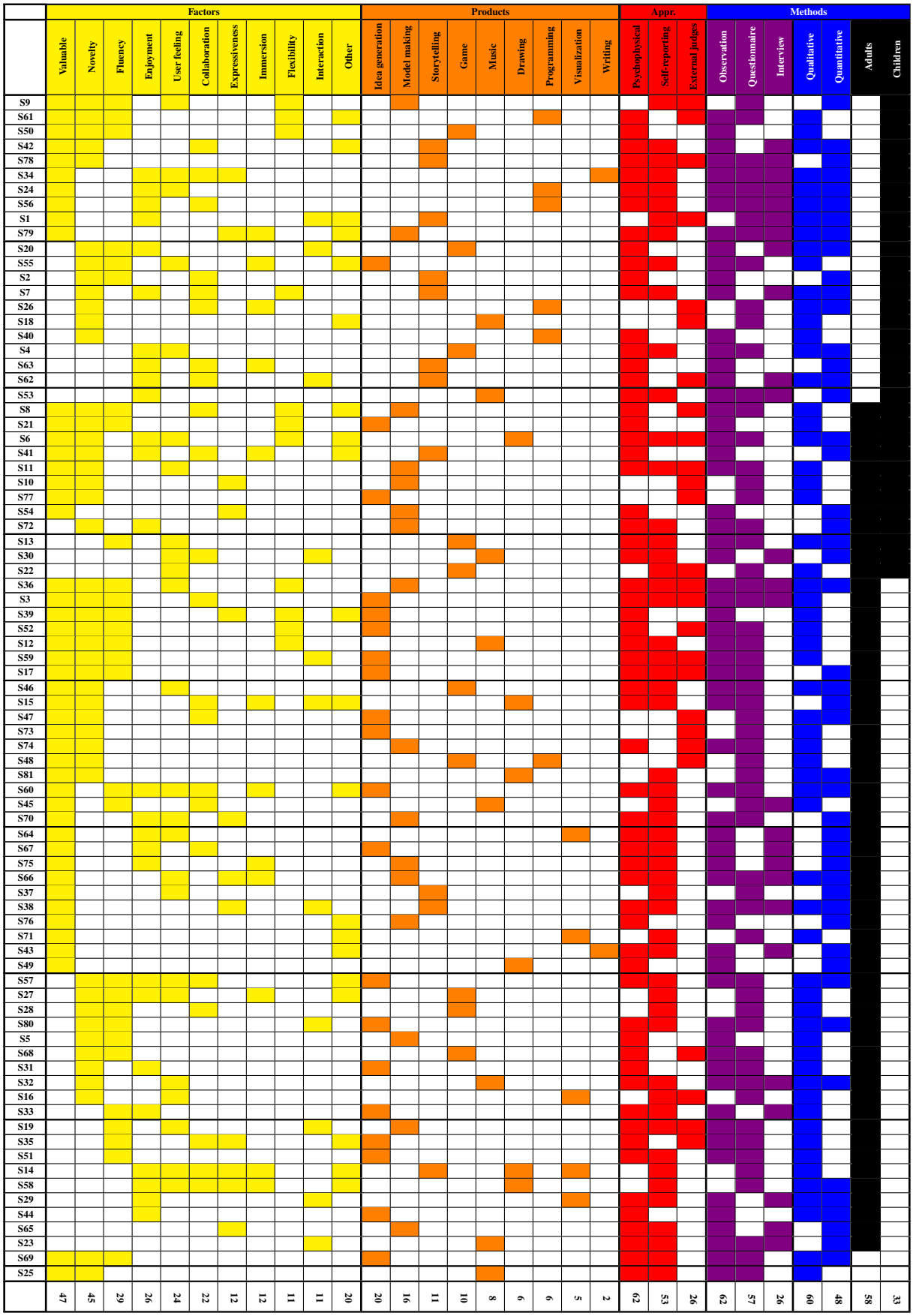
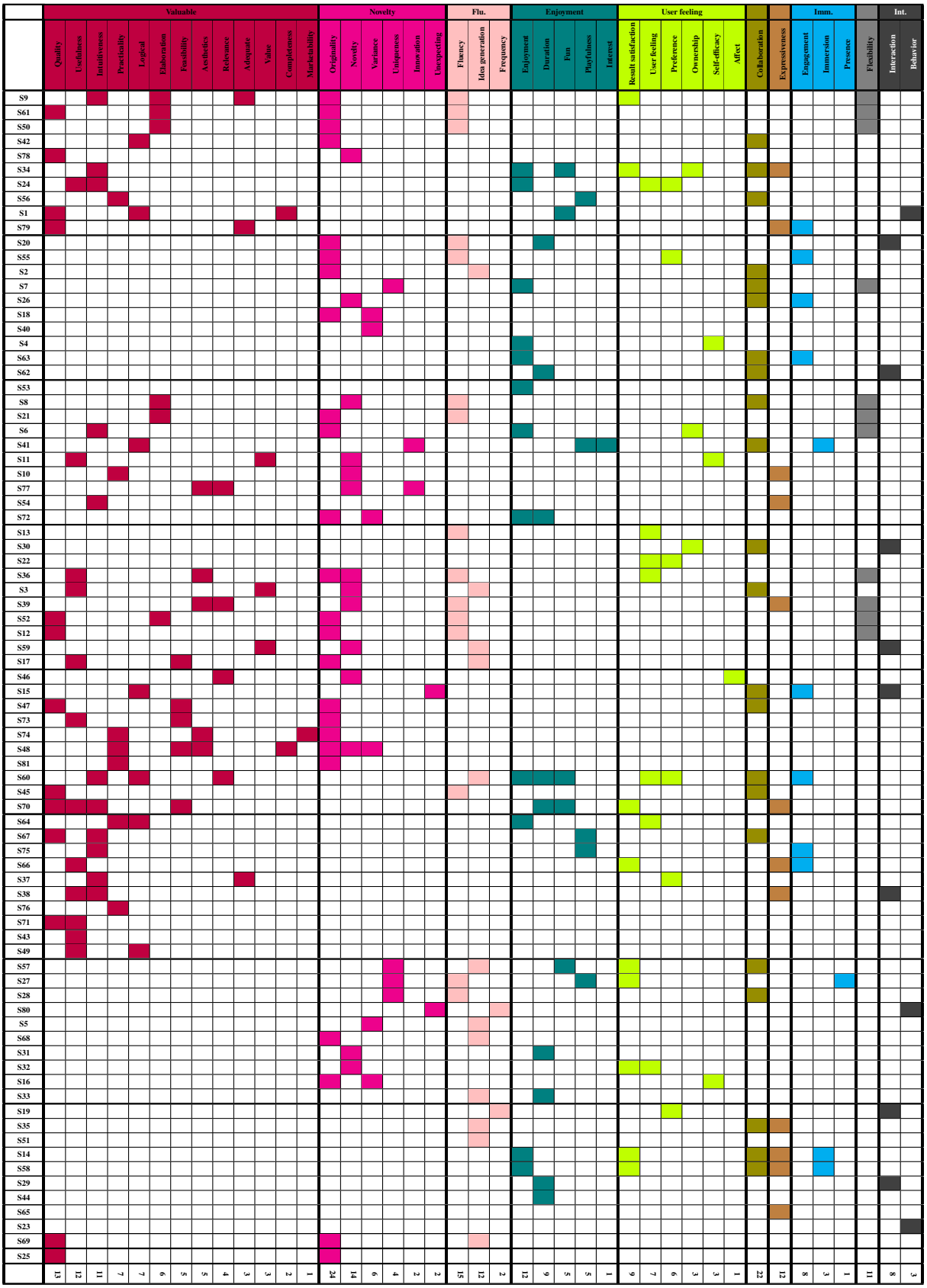


Table 3.9: Distribution of terms used to evaluate the 10 factors in the included studies. (Flu. = Fluency, Imm. = Immersion, Int. = Interaction)



Valuable

According to Oxford Dictionary [31], *valuable* is "very useful or important". The included studies used 13 different terms to evaluate *valuable*, as indicated in table 3.9. *Value* is the most related term, defined by **S3** as "the estimated likelihood that the idea would work if that it were actually implemented", **S11** as "sophistication, aesthetics, and coherence", and **S59** as "how useful the product idea is and how practical the idea sounds assuming the 'fabric display' technology is real" on p. 113, 105, and 86, respectively. However, *quality* was the most used term, included in 13 studies. Four of these studies defined *quality*, but they did it differently. For example, **S78** defined it on p. 7 as "coherence, continuation and completion" and **S25** on p. 150 as "the length of the participant response". Nevertheless, all these definitions concur with the definition provided by Oxford Dictionary [31] because the term *valuable* depends on the context.

Of the 12 studies focusing on the term *usefulness*, three defined it. An example is the definition of **S36** on p. 14: "how well the product does what it is supposed to do". Besides, four of the 11 studies utilizing the term *intuitiveness* defined it. **S6**, **S9**, and **S24** did it similarly, as "easy to use" on p. 421, "user-friendliness of product" on p. 26, and "the degree to which an individual believes that attending the activity is easy" on p. 35, respectively. **S67** elaborated on p. 8. these definitions to "the property of a tool to foster a rapid and clear understanding of the artifacts employed for idea development". Two of the seven studies utilizing the term *logical* defined it. **S1** as "the ability to provide meaningful reasoning for the actions in the story" and **S60** as "the objectives of the activity were stated clearly" on p. 29 and 109, respectively. *Aesthetic* was defined by two of the five studies using the term — **S39** as "how well the curation uses text to inform, describe, explain, and expand on constituent ideas" and **S74** as "appearance" on p. 15 and 339, respectively. The definitions of *intuitiveness*, *usefulness*, *logical*, and *aesthetic* correspond with each other and with Oxford Dictionary's definition of *valuable* [31] to a greater extent than the definitions of *quality*.

The CEMs of *quality*, *intuitiveness*, and *logical* were both objective and subjective. For example, **S25** specified an objective CEM of *quality* based on the psychophysical measurement AUT by examining the length of the participant's response, and **S45** utilized a subjective self-reporting questionnaire including questions like "How would you rate the overall quality of that performance?" on p. 2299. Examples of *intuitiveness* are **S54** conducting objective psychophysical observations of the prototyping and outcomes of the participants and **S75** performing subjective interviews of self-reporting of the participant's experience. *Logical* was also evaluated by self-reporting subjective interviews and objective psychophysical observations. Examples are **S64** asking the interviewees the question "Did you find the system easy to understand?" on p. 163, and **S15** observing how the participants made sense of the agent, respectively. On the other

hand, it was only subjective CEMs of *usefulness* and *aesthetic*. Examples are **S17** utilizing two external expert judges to assess the problem-solving ability of an idea on a five-point Likert scale and **S74** using external expert judges from four different backgrounds to evaluate the final designs, respectively.

Novelty

Oxford Dictionary defines *novelty* as "the quality of being new, different and interesting" [31]. Six different terms were utilized to evaluate *novelty* in the included studies. The word *novelty* was used in fourteen studies, but only half of them did not define it. The studies defining it did it differently: **S3** focused on *surprising*; **S11** on *uniqueness*; **S8** on *uniqueness*, *unusual*, and *surprising*; **S59** on *surprising*, *originality*, and *novel*; **S39** on *rarity*; **S46** on *rarity* and *infrequency*.

On the other hand, 24 studies utilized the term *originality* to evaluate *novelty*. Less than half of them define it, and the studies had a different focus as with *novelty*; **S2**, **S17**, and **S52** focused on *novelty*; **S6**, **S21**, and **S61** on *uniqueness*; **S55** on *unusual*; **S20** on *unusual*, *relevance*, and *rarity*; **S36** on *unusual*, *uniqueness*, *rarity*, and *surprising*; **S47** on *unusual*, *novelty*, and *infrequent*; **S50** on *unusual*, *innovative*, and *fresh*. Most of these words defining *originality* correspond with the words defining *novelty* and match the definition provided by Oxford Dictionary [31].

The included studies evaluated *originality* both objectively and subjectively. For example, **S61** objectively and psychophysically observed the participants' number of unique responses, while **S73** utilized external expert judges to subjective evaluate CSTs with a five-point Likert scale. *Novelty* was mainly evaluated using external judges. An example is **S8** used external expert judges with backgrounds in creativity studies to rate the outcomes of the CST on a five-point scale. On the other hand, **S32** observed the chords in the composers' creative process.

Fluency

Fluency is defined as "the quality of doing something in a way that is smooth and shows skill" by Oxford Dictionary [31]. The included studies utilized three different terms to evaluate this. None of the studies used multiple of these terms, contrasted to *valuable* and *novelty*. Two studies used the term *frequency*, 12 *idea generation*, and 15 *fluency*. Most of the studies utilizing *fluency* defined it homogeneously concurring with the definition provided by Oxford Dictionary [31]. Indeed, **S21**, **S27**, and **S28** used the same definition: "the number of ideas generated" on p. 3, 170, and 143, respectively. Still, the definition had some variations. For example, **S36** stated on

p. 14 “the number of interpretable, meaningful and relevant responses” and **S52** on p. 206 “the frequency of ideas generated”. *Fluency* is mostly assessed objectively and psychophysically by identifying the sum of relevant ideas.

Most of the studies lack theoretical grounding for the choice of *idea generation* as an evaluation factor. Nevertheless, **S2** explains on p. 424: “According to literature, idea generation (or fluency) is one fundamental creative ability that tends to lead to originality and novelty [43]. With YOLO, we aim to stimulate fluency during the creative process of storytelling”.

Enjoyment

Oxford Dictionary defines *enjoyment* as “the pleasure that you get from something” [31]. Five different terms were utilized to evaluate *enjoyment* in the included studies. The word *enjoyment* was used in twelve studies, *duration* in nine, and *fun* and *playfulness* in five each. *Enjoyment* was only defined by **S24** as “the degree to which the activity is perceived to be personally enjoyable” on p. 35. Only **S67** defined *playfulness* by stating: “Playfulness is the property of a tool to encourage unfettered trialability in design, helping the user to push intermediate solutions to final results iteratively. It includes the support of iterative development of (intermediate) creative products, the provision of simulation functionality, comparison as simultaneous representation of diverse sets of alternative data, and modification for altering of the problem of the creative task” on p. 8. Both these definition corresponds with Oxford Dictionary’s definition [31]. None of the studies defined *duration* or *fun*.

The majority of the studies evaluated *enjoyment* subjectively. **S64** performed on p. 163 self-reporting interviews asking the question “Do you enjoy using the system?”, while **S4** utilized on p. 6 a self-reporting questionnaire including the question “How much did you enjoy the creativity project? (Circle a number on the scale below.)”. On the other hand, **S7** psychophysically observed the *enjoyment* of interactions around the CST. *Duration* was mostly evaluated objectively by psychophysical observation. An example is **S33** observing the *duration* of the study session with the CST by video recording. On the other hand, **S60** evaluated *duration* by conducting a qualitative self-reporting questionnaire including the statement “Time appeared to go by quickly when I was interacting with the activity” on p. 109. *Fun* and *playfulness* were mostly evaluated subjectively by self-reporting. For example, **S57** asked the participant to, based on their experience of CST, rate their emotional level of boredom on a 10 point scale, and **S75** interviewed the participants regarding their playful experience of the CST.

User feeling

According to Oxford Dictionary [31], feeling is defined as "an attitude or opinion about something". Six different terms were used to evaluate *user feeling* in the included studies. Nine of the studies utilized the term *result satisfaction*, seven *user feeling*, and six *preference*. Few of the studies defined *result satisfaction*, but **S27**, **S34**, and **S66** did it similarly as "[the users'] perceptions of [...] their avatar" on p. 170, "requires the users to evaluate the final result of the creative process" on p. 64, and "happy with the final design" on p. 165, respectively. All the seven studies defined *user feeling* correspondingly. For example, **S24** defined it as "the degree to which a person feels positive about the activity" on p. 35, and **S32** as "how [the users'] felt the tool had impacted their creativity process" on p. 246. Most of the studies defined *preference*, and they did it similarly. An example is **S55** defining it on p. 29 as "the extent to which [the user] liked the [CSTs]". All these definitions concur with the definition of feeling provided by Oxford Dictionary [31].

It was utilized subjective self-reporting CEMs to evaluate *result satisfaction*, *user feeling*, and *preference*. For example, **S58** used the agreement statements of the CSI and **S70** a questionnaire including the statements "Take appropriate time to finish" and "Satisfied with generated path" on p. 1103. Examples of *user feeling* are **S36** interviewing the participants regarding their experience of using the CST and **S24** utilizing on p. 35 questionnaires including the Likert statements "My decision to participate in the activity was a wise one", "I was pleased with the activity", and "I was satisfied with the activity". On the other hand, **S22** utilized teachers as external judges to evaluate their *preference* for the student's outcomes.

Collaboration

Oxford Dictionary defines *collaboration* as "the act of working with another person or group of people to create or produce something" [31]. This factor was evaluated in twenty-two of the included studies, each of them utilizing the word *collaboration*. A small amount of the studies defined *collaboration*. **S8** defined it on p. 148 as "co-operation time, which is the time that both participants in a group were effectively co-manipulating the platform during the time needed to complete" and **S34** on p. 65 as "mutual influence, sharing and feedback with different agents". Both of these definitions correspond with Oxford Dictionary's definition [31].

To evaluate the *collaboration*, different CEMs were used. **S56** assessed it subjectively on p. 71, asking the questions "How much did you collaborate with the Graphical/Tangible interface?" and "With what system did you collaborate the most?" in the self-reporting Likert scale and the Fun Sorter, respectively. **S41** observed it objectively and psychophysically by pinpoint

sessions where both participants contributed to an idea. Besides, **S57** calculated collaboration psychophysically as $|\frac{1}{N} - \frac{O_i}{\sum_{i=1}^{N} O_i}| \times 100$, where N is the group size and O_i is each participant's observed collaborative behavior.

Expressiveness

Expressiveness is defined as "the quality of expressing somebody's thoughts and feelings" by Oxford Dictionary [31]. This factor was evaluated by twelve of the included studies, all using the word *expressiveness*. A small number of the studies defined it, and the definitions varied. **S34** matches on p. 64 Oxford Dictionary [31] with the definition "how well the users are able to be creative and express themselves in the creative process". Nevertheless, **S35**, **S54**, and **S79** define it more like the term *variance* with the definitions "the success in producing a variety of drawings for a specified task", "the wide range of constructions that the kit makes possible for both adults and kids", and "encourage users to explore topics through a new form of storytelling medium" on p. 181, 346, and 13, respectively.

Both subjective and objective CEMs were used in the studies. **S14** used on p. 6 a self-reporting questionnaire with the agreement statements "The system or tool allowed me to be very expressive" and "I was able to be very creative while doing the activity inside this system or tool". Likewise, **S54** psychophysically observed the range of outcomes.

Immersion

Oxford Dictionary defines *immersion* as "the state of being completely involved in something" [31]. The included studies utilized three different terms to evaluate *immersion*. Eight studies used the term *engagement*, while only three used *immersion*. A small amount of the studies defined *immersion*, and they understood it differently. **S79** saw it as *interest*, **S63** as *curiosity*, and **S66** as *absorption of the activity*, which is closest to Oxford Dictionary's definition [31].

Most of the studies evaluated *immersion* subjectively by self-reporting. Examples are **S79**'s interview of the participants' engagement and **S60**'s qualitative questionnaire, including the statements "As I carried out the activity I was absorbed in it" and "My interest in the subject matter grew as I did the task" on p. 109. Nevertheless, **S66** psychophysically calculated the variance between perceived and completion time. Some studies also utilized objective psychophysical observation; for example, **S75** observing the participant's engagement behavior by asking them to think aloud throughout the trial.

Flexibility

Flexibility is defined as "the ability to change to suit new conditions or situations" by Oxford Dictionary [31]. This factor was evaluated in eleven of the included studies, all utilizing the word *flexibility*. A majority of the studies specified how they understood the term, a similar understanding that the studies explained differently. **S36** defined on p. 14 *flexibility* as a "variety of categories of relevant responses", while **S39** on p. 11 elaborated this to be more similar to Oxford Dictionary's definition [31]: "consideration of alternative interpretations, which means ways of thinking and viewpoints. Flexibility in thinking describes the cognitive process of trying out a various ways of looking at a problem. Flexibility measures the span of the solution space explored during ideation".

All of the studies evaluated *flexibility* by identifying the number of different categories the ideas belonged to. A small amount of the studies specified the CEM in detail. Examples are **S8** rating the *flexibility* by using external expert judges and **S61** psychophysically calculating the number of principles, procedures, and tools utilized to explain ideas.

Interaction

As claimed by Oxford Dictionary [31], "if one thing has an interaction with another, or if there is an interaction between two things, the two things have an effect on each other". Eleven of the included studies evaluated *interaction*, eight utilizing the word *interaction* and three *behavior*. Nevertheless, none of them specified how they defined *interaction*. The factor was evaluated exclusively objectively by psychophysically observing the interaction between the digital CST and the participants.

Other

The twelve different terms presented in table 3.10 did not fit in the 10 mentioned factors and were categorized in the *Other* category. Six of the terms were only evaluated in one study each, while *exploration* was the most utilized term with a frequency of seven. Oxford Dictionary defines *exploration* as "an examination of something in order to find out about it" [31]. Only two of the included studies defined the term; **S14** as "to explore many different ideas, options, designs, or outcomes" on p. 6 and **S61** as "how sensitive the students were in exploring a new domain in response to issues surrounding them" on p. 181. These definitions concur with the definition provided by Oxford Dictionary [31].

Table 3.10: Distribution of the terms used to evaluate other factors in the included studies.

	Exploration	Knowledge	Imagination	Motivation	Inspiration	Emergence	Persistence	Risk taking	Self-determination	Start help	Success guaranty	Accessibility
S1		■										
S6										■	■	
S8				■								
S14	■											
S15						■						
S18	■											
S27					■							
S35		■		■			■					
S39						■						
S41			■					■	■			
S42	■	■			■							
S43	■											
S55			■									
S57				■								
S58	■											
S60		■	■									
S61	■											
S71					■							
S76	■											
S79												■
	7	4	3	3	3	2	1	1	1	1	1	1

The CEMs of *exploration* were both objective and subjective. **S42** observed psychophysically the users' processes of creation and performance. On the other hand, **S14** utilized on p. 6 a self-reporting questionnaire including the agreement statements "It was easy for me to explore many different ideas, options, designs, or outcomes, using this system or tool" and "The system or tool was helpful in allowing me to track different ideas, outcomes, or possibilities".

3.2.3 Context

This section addresses the context of the CEMs in the included studies by specifying software products, approaches, and methods.

Products

It was identified nine different creativity topics for the evaluated products in the included studies. Table 3.8 presents the distribution of these topics, while table 3.11 specifies the different software.

Table 3.11: The distribution of the creativity topics and software for the evaluated products.

Topic	Software	Number of studies
Idea generation	• Self-developed digital idea-generation tools (**S3**, **S31**, **S33**, **S35**, **S39**, **S44**, **S52**, **S59**, **S67**, **S80**) • Self-developed digital brainstorming tools (**S17**, **S60**, **S69**) and brainwriting tool (**S57**) • Brainstorming in a video conference system (**S51**) and Adobe Connect (**S47**) • The idea generation apps for children AppLab and Common Sense Media (**S55**) • Virtual reality (VR) idea generation tool (**S21**, **S77**) • Self-developed online multiplayer ideation game (**S73**)	20
Model making	• Self-developed digital model making kits (**S9**, **S54**, **S79**) and prototype tools (**S8**, **S66**, **S70**, **S76**) • Computer aided design (CAD) (**S5**), the CAD Rhino (**S74**), VR CAD (**S19**), and 3D-CAD (**S10**, **S11**) • Self-developed 3D-model tools (**S65**, **S75**) and 3D-sketch app (**S72**) • Self-developed digital multimedia learning tool for mechanism design (**S36**)	16
Storytelling	• Self-developed digital storytelling tools (**S1**, **S2**, **S7**, **S37**, **S38**, **S42**, **S62**, **S63**) • Short story writing in Google Docs (**S14**) • Storytelling with the software Tuxpaint combined with the Our Story app (**S41**) • The digital storytelling software Frames combined with Powerpoint (**S78**)	11
Game	• Self-developed game creation apps (**S13**, **S22**) • Game development platforms (ARIS (**S4**), Construct (**S48**), Powerpoint (**S48**), Powtoom (**S48**), and Roar (**S48**)) • Self-developed digital gamified learning management system (**S46**) • The digital problem-solving game Crayon Physics Deluxe (**S20**) • The digital constructing game Minecraft (**S68**) • The VR idea generation game Second Life (**S27**, **S28**) • The educational computer game I Spy Treasure Hunt (**S50**)	10
Music	• Self-developed digital music creation tools (**S18**, **S23**, **S32**, **S45**) • Self-developed app for music creation and manipulation through movement (**S30**) • Visual music interactive art system with Processing and Leap Motion (**S12**) • Self-developed adaptive digital musical instrument (**S25**) • Self-developed educational digital music game (**S53**)	8
Drawing	• Self-developed digital co-creative drawing agent (**S15**) • Self-developed sketch combination system integrating Mechanical Turk and Google Docs (**S81**), OdoScetch (**S14**), and AutoDesk Scetchbook (**S14**) • Self-developed digital painting tool (**S6**) and the Java app Pixelitor (**S49**) • Self-developed digital color picker (**S58**) and Mac OS's color picker BiCEP (**S14**)	6
Programming	• App Inventor (**S26**) • Scratch (**S24**, **S26**, **S48**, **S61**) • Unity (**S48**) • Unspecified programming tools (**S40**, **S56**)	6
Visualization	• Self-developed digital photography tool (**S29**) • Adobe Photoshop (**S14**) • Self-developed digital data visualization tool (**S64**) and font visualization tool (**S71**) • MySpace's Flash-based AdBuilder Tool (**S16**)	5
Writing	• Self-developed digital poetry tool (**S34**) • Self-developed digital journalism tool (**S43**)	2

Approaches

The distribution of the three basic approaches by Caroll and Latulipe [41] in the CEMs of the included studies is presented in table 3.8 and figure 3.2. The majority of the studies utilized a combination of these approaches.

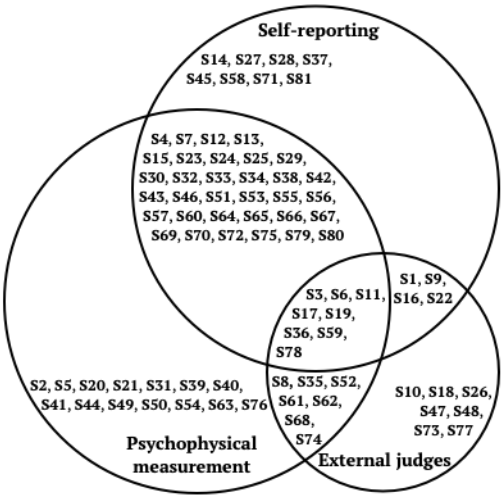


Figure 3.2: The distribution of the approaches in the included studies.

Methods

The included studies used different methods and a combination of different methods, as indicated in table 3.8. Most of the studies utilized several data collection methods, and approximately one-third a mixed method. Some of the studies also used existing CEMs, which are presented in table 3.12. A small amount of the included studies lacked information regarding the methods. Twenty-seven of the studies lacked information regarding the gender distribution of the participants, **S48** regarding the sample size, **S69** regarding the age, and **S25** regarding the participants.

Table 3.12: The distributions of existing Creativity Evaluation Methods in the included studies.

Creativity Evaluation Methods	Studies
Alternative Uses Task	**S21**, **S25**, **S31**
Torrance Tests of Creative Thinking	**S36**, **S50**, **S61**
Abbreviated Torrance test for adults	**S12**
Multidimensional Stimulus Fluency Measure	**S20**
Remote Associates Test	**S21**
Creativity Support Index	**S14**, **S58**
Ideation metrics of curation	**S39**
Likert Scale	**S1**, **S16**, **S17**, **S24**, **S25**, **S28**, **S31**, **S32**, **S34**, **S36**, **S37**, **S45**, **S55**, **S56**, **S57**, **S58**, **S59**, **S66**, **S73**, **S77**, **S79**, **S80**, **S81**
Creativity Achievement Questionnaire	**S16**
Kaufman Domains of Creativity Scale	**S77**
AttrakDiff	**S22**
Fun-Sorter	**S56**
MemoLine	**S53**
Theory of inventive problem solving	**S48**
Consensual Assessment Technique	**S18**, **S78**
Creative Product Semantic Scale	**S36**

3.2.4 Age of the Target Group

Table 3.8 indicates that the target group of the included studies is mainly adults, but 15% used both adults and children as participants. The distribution for each of the different evaluation factors, products, approaches, and methods for the two different target groups in the included studies is presented in table 3.13. The table indicates that the distribution depends on the age of the target group.

Table 3.13: Distribution of factors, products, approaches (appr.), and methods in the included studies, divided into children and adults as target age.

	Factors											Products									Appr.			Methods				
	Valuable	Novelty	Fluency	Enjoyment	User feeling	Collaboration	Expressiveness	Immersion	Flexibility	Interaction	Other	Idea generation	Model making	Storytelling	Game	Music	Drawing	Programming	Visualization	Writing	Psychophysical	Self-reporting	External judges	Observation	Questionnaire	Interview	Qualitative	Quantitative
Children	18	20	9	13	10	11	4	5	7	4	9	3	7	8	5	3	1	5	0	1	26	18	13	26	21	12	24	22
Adults	35	31	22	16	19	14	10	8	7	8	14	18	14	4	7	5	6	1	5	1	43	39	19	43	42	15	42	31

3.2.5 Effectiveness

This section addresses the effectiveness of the CEMs in the included studies, focusing on the definitions of creativity, the validity of the methods, and the outcomes.

Creativity

Forty-two of the included studies did not define creativity. Figure 3.3 indicates no unambiguous opinion of creativity in the 39 studies defining it since there are few overlaps among the different types of definitions. Only **S28** and **S57** utilized the same definition by Sternberg: "the ability to produce work that is both novel and appropriate" [34, p. 3]. Unclear definitions were used in eight of the studies. For example, **S69** states on p. 103: "Identifying new ideas can be difficult due to individuals' limited vision, knowledge, experience, motivation and time. Collaborative teamwork that pools and integrates efforts from multiple individuals is thus considered a useful way to approach creativity". Both the definitions focusing on the characteristics *novel* and *valuable* and the *process of creating* correspond with Oxford Dictionary's definition of creativity [31].

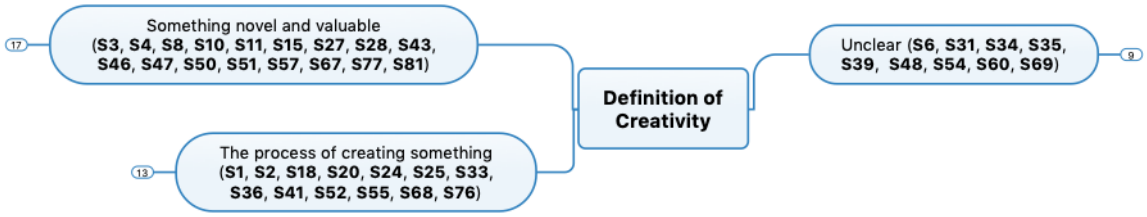


Figure 3.3: Definitions of creativity in the 39 studies defining the term.

Validity of the methods

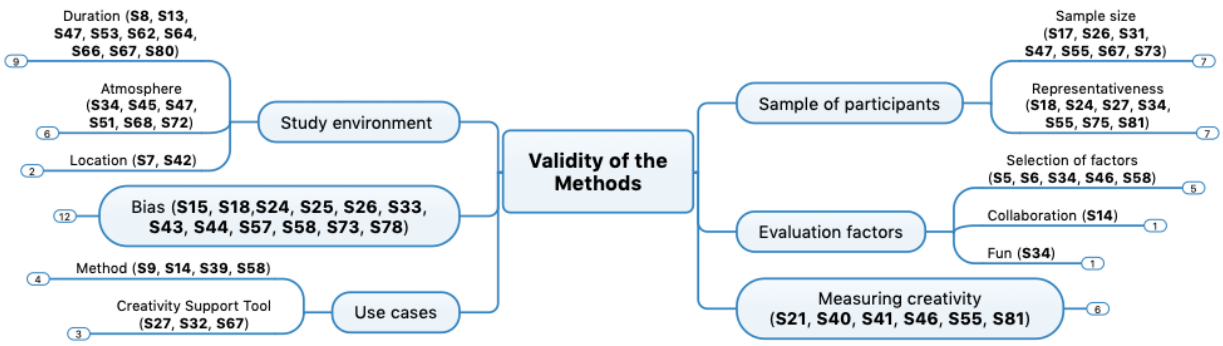


Figure 3.4: The validity of methods in the 44 studies including details of the validity.

The validity of the methods was not specified in 37 of the included studies. Figure 3.4 illustrates the six different categories discussed in the 44 studies that included details of the validity. A small amount of the studies included multiple of these categories.
S34 discovered the evaluation factor *fun* to be less important than expected, and **S14** highlighted the difficulty of including *collaboration* as an evaluation factor since many creative activities are not collaborative. The included studies used various approaches in the selection of evaluation factors. **S58** utilized the well-tested CEM CSI, while **S6** and **S34** chose the evaluation factors by identifying partially overlapping factors in three and nine articles, respectively. Nevertheless, **S46** presents how the factors *novelty*, *affect*, and *importance* can evaluate creativity in web-based information design systems, and **S5** recommends including *novelty* and *quality*.

The studies emphasized different CEMs. **S40** highlighted the calculation of divergence from a standard norm. On the other hand, **S81** utilized a binary measure that only qualified the design as creative if both the scales of the factors *practicality* and *originality* exceed 4.0 on a seven-point Likert scale. The studies also had different limitations regarding measuring creativity. **S21** indicated the main limitation of the study was the interpretations of AUT and RAT scores. **S41** emphasized the claim by Glăveanu [70] regarding how the community could impact a person assessing creativity. One limitation of **S46** was the blurry boundaries between the evaluation of creative design and creative pedagogy in the learning CST. A significant challenge in **S55** was the assessment of *engagement* for young children. Three different CEMs were utilized, but two were excluded due to the validity of the results.

Outcome

The majority of the included studies' outcomes was a description of how an existing or self-developed CST improved creativity. The products in these two CSTs categories have a different distribution of creativity topics, as presented in figure 3.5. Besides, a small amount of the studies investigated gender differences of creativity or developed a new CEM for CSTs.

The developed CEMs of **S14** and **S39** are already mentioned in section 2.1.2 as Cherry and Latulipe's CSI [39] and Kerne et al.'s IMC [71], respectively. The divergence test in **S40** is described in the previous section. Besides, **S9** developed a design model of user creativity platforms and **S46** a rubric for creativity evaluation in learning management systems using the evaluation factors *importance*, *affect*, and *novelty*.

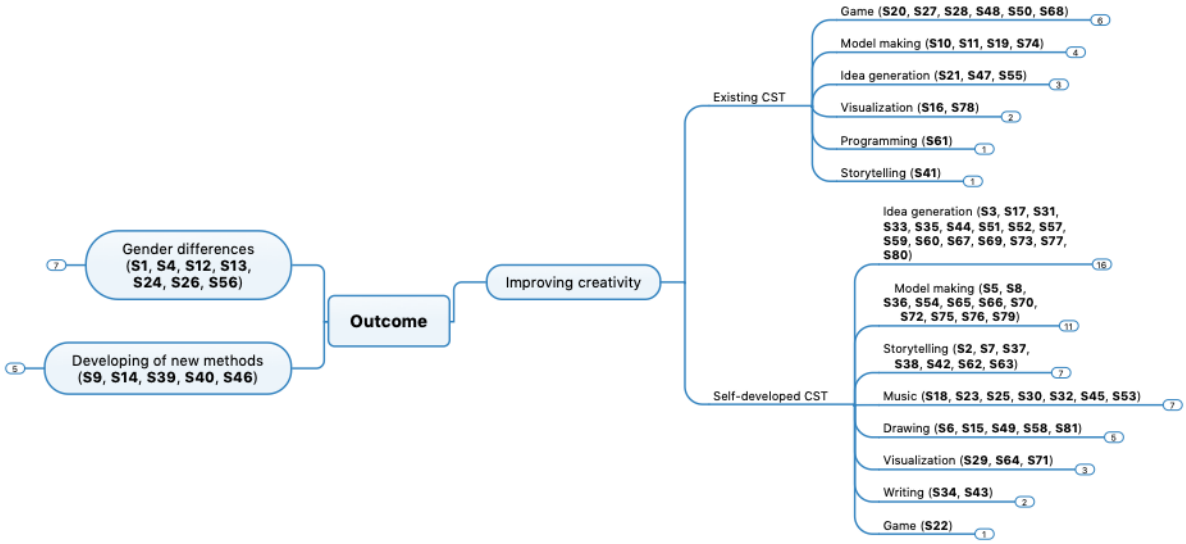


Figure 3.5: The outcomes of the included studies.

3.3 Discussion

3.3.1 Implication for Research

By identifying trends and opportunities for future works, this SLR has significance for further research. The wide specter of CEMs indicates a need for more research on this area. The SLR also pinpointed the knowledge gap of digital visualization and drawing CSTs for children. Moreover, the SLR identified that the two promising CEMs for children — Smileyometer and This or That — had not been explored for digital CSTs. Smileyometer and This or That is often used in the interaction design and children community, e.g., Sim and Horton [82], Zhang et al. [83], and Yarosh et al. [84]. Because of the wide specter of current CEMs of digital CSTs for children, there is also a need for more structured CEMs. Hence, we support in section 3.4 Remy et al.'s [72] recommendation **R6** (section 2.1.2) of developing a toolbox. However, we want the recommendation to focus on a tailored approach instead of a toolbox since the CEM of digital CSTs must be tailored to the specific case.

3.3.2 Implication for Practice

The identified trends of CEMs in this SLR could help practitioners in companies developing digital CSTs for children to make better solutions. The challenge is that a standardized CEM does not exist. The CEM could be more structured by presenting the companies the definition and specific examples of each of the 10 proposed factors in section 3.2.2. Moreover, the company

could select which factors to focus on based on target age and product category. This is presented as a recommendation in section 3.4. Besides, the psychophysical measurement approach with observation is promising as it occurred in 79% of the studies with children as participants. This is easiest to use on the youngest children because the participants are only observed and do not need to answer questions. The studies of older children often combined quantitative and qualitative methods. Therefore, we recommend in section 3.4 to start with a quantitative psychophysical measurement approach but considering expanding to qualitative methods.

3.3.3 Limitations of the Systematic Literature Review

Potential Biases

The outcomes of the included studies varied, and comparison between the studies is difficult. For example, a large amount of the included studies presented how a CST improved creativity, but many of them did not specify the way creativity was evaluated. As pinpointed by Remy et al. [72], there are three ways to evaluate the creativity of a CST; the creativity of the outcome, the usability of the CST, or the productivity of the process supported by the CST. Thus, we support in section 3.4 Remy et al.'s [72] recommendation **R1** of defining a clear goal of the CEM of the CST.

A great quantity of the included studies focused on evaluating usability instead of creativity. This bias also occurred in the literature review by Remy et al., where it was a clear trend that most of the recent studies evaluated CSTs using usability testing [72]. The important product evaluation aspect usability is one of the main characteristics in the software product quality model ISO/IEC 25010:2011 [18]. On the other hand, children's beneficial quality of creativity must not be forgotten. Thus, we merge Remy et al.'s [72] recommendation **R3** of deciding whether to evaluate creativity or usability with the recommendation of the 10 factors in section 3.4 since the evaluation factor *valuable* is closely related to usability.

A large amount of the studies did not ground their choice of CEM sufficiently in theory. This was also found in the literature review by Remy et al. [72], where less than half of the CEMs of the studies were built on identifiable theoretical foundations like *little-c creativity* in **S33** and **S43**. Therefore, we support in section 3.4 the recommendation **R2** of using theoretical grounding in the CEM of the CST. **S6** and **S34** are good examples of this, grounding their selection of evaluation factors on three and nine studies, respectively. In section 3.4, this recommendation of theoretical grounding is merged with the recommendation of the 10 factors that rely on theoretical groundings of CEMs.

Data Synthesis

To identify themes answering the RQs, the data were categorized. Howbeit, the categories had some overlap. An example is the self-developed educational digital music game product in **S53**. The CST was placed in the music category but also belonging to the game category. Since a few studies did not adequately describe the data to be extracted, wrong interpretation was also possible in the data extraction process. On the other hand, the interpretation was made consistently as only I did this. Further, this was discussed with the supervisors to reduce bias.

3.4 Take-Home Message

This SLR of the state-of-the-art procedures of CEMs of digital CSTs for children contributes to both practitioners and researchers by identifying trends in CEMs and areas needing further investigation, respectively. Future research could focus on the knowledge gap in the CEMs of digital drawing and visualization CSTs for children and exploring the promising CEMs Smiley-ometer and This or That for digital CSTs for children.

The same issues that led to Remy et al.'s [72] recommendations **R1**, **R2**, **R3**, and **R6** in section 2.1.2 occurred in the SLR and could be applied to CEMs of digital CSTs for children as the recommendations (RECs) **REC1**, **REC2**, and **REC4**. **REC1** is identical with **R1**, while **REC2** included **R2** and **R3** by changing the focus from theoretical groundings to the 10 factors identified by the findings. **REC4** was centered around a tailored approach instead of a toolbox as **R6** since the findings indicated that the CEM has to be tailored to the software product. We also added **REC3** based on the utilized methods in the included studies. Thus, we recommend the following four RECs for CEMs of digital CST for children:

- **REC1)** Clearly define the goal;
- **REC2)** Review and choose amongst the 10 evaluation factors *valuable*, *novelty*, *fluency*, *enjoyment*, *user feeling*, *collaboration*, *expressiveness*, *immersion*, *flexibility*, and *interaction*;
- **REC3)** Start with a quantitative psychophysical measurement approach but consider expansion to qualitative methods;
- **REC4)** Develop a tailored approach.

Chapter 4

Developing of the tailored Creativity Evaluation Method

To develop a tailored CEM of Ablemagic to investigate the evaluation of digital CSTs for children in practice, it was a collaboration between the ChildrenByDesign project and Ablemagic from August 2020 to February 2021. The development of the tailored CEM is the focus of this chapter.

4.1 Method

This section presents the method used to develop the tailored CEM of Ablemagic, the second step in figure 1.3. The overview of the method of the thesis is presented in figure 1.2.

4.1.1 Data Analysis

The research used primarily qualitative data analysis to answer the RQs by exploring Ablemagic's CEM in dept by summarizing, interpreting, and categorizing the collected data. In contrast to quantitative data, qualitative data is all non-numeric data [2]. The documents, the observation notes, and the transcribed interviews were analyzed inductively through Cruzes and Dybå's five recommended steps for thematic synthesis [81] in the same way as described in section 3.1.7. Details regarding the data analysis of the different data collection methods are presented in the next section.

4.1.2 Data Collection

As indicated in figure 1.2, the study used method triangulation with documents, observations, and interviews. A questionnaire with pre-defined questions in a standardized order was not a suitable data collection method for the low number of employees in Ablemagic [2].

Documents

Three existing internal documents of Ablemagic were used in the data collection. These documents included Ablemagic's design principles for digital CSTs for children, one observation guide for a user test of a digital drawing CST, and one observation note of another user test of a different digital drawing CST. Some of the documents also included pictures. Information from these documents gave insight into the current practice of Ablemagic's CEM of digital CSTs for children. Collecting documents was a time-saving way to understand the practice in Ablemagic compared to collecting all these data by own research. It was in each document extracted data of factors, products, and methods in the same way as the SLR, described in section 3.1.6. Moreover, this data was coded in the same themes used in the SLR, presented in table 3.8.

Observations

Ablemagic's SE process of two digital CSTs for children was observed to understand the current CEM. Oates defines observations as "watching and paying attention to what people actually do" [2, p. 36]. The observations included three meetings in the SE team, each lasting around one hour. This data collection method supplemented the documents by investigating if the reality corresponded with the presentation of Ablemagic's CEM in the documents. Observations also reveal what people actually do instead of what they say they do when questioned in interviews [2].

It was used *participant observation* where the observer took the role as a participant-observer and participated when it was possible [2]. This was chosen as it was beneficial to observe information of Ablemagic through the experience as a participant in the SE process. Besides, it was included aspects from systematic observation with an observation template regarding the categories creativity, the software engineering team, communication, and other relevant things. This gave a more systematic collection of data, so the observer knew what was important to focus on during the observation. The filled-out templates became documents used in the data analysis.

The first meeting happened in October. It was a physical concept development meeting of a digital drawing CST between one developer and two concept developers. A Ph.D. student and I participated as observers. In November, it was a digital creativity discussion between two developers and the tester regarding evaluating creativity in the digital drawing CST. A supervisor and I observed. The discussion was based on the SLR findings and the results from the first user test of the digital drawing CST. At last, I participated in a planning meeting with the tester in February, discussing suggestions for the tailored CEM of Ablemagic before an upcoming user test of the digital drawing CST and a digital visualization CST. The discussion was based on the practice of

Ablemagic, the tester's document with proposals, and my presentation of findings from the SLR.

Each observation was coded in an exploratory way with a total of four (the environment, the software development process, the communication, and the evaluation factors), two (the evaluation factors and the evaluation methods), and three (the evaluation factors, the practical implementation of the user test, and the tester's experience of the meeting), higher-order themes emerging from the observations, respectively.

Interview

The accumulated insight from the documents and observations were triangulated with an interview of the tester in Ablemagic regarding the software development process of digital CSTs for children. The interview was also based on the findings from the SLR. According to Oates, an interview is "a particular kind of conversation between people where, at least at the beginning of the interview if not all the way through, the researcher controls both the agenda and the proceedings and will ask most of the questions" [2, p. 36]. The interview lasted around half an hour and were semi-structured, so the subject had the possibility to turn the conversation into themes the interviewee had not thought about [2]. The interview guide is attached in Appendix C.1. The video communications software Zoom 5.4.9[1] was used to audio record the interviews with the built-in recorder. Further, the interviews were transcribed and became documents used in the data analysis. The interview was coded exploratory with three higher-order themes (the concept development phase, the insight phase, the design thinking process, and the CEM) and 47 sub-themes emerging.

4.2 Results

4.2.1 The Documents of Ablemagic

Table 4.1 present the distribution of the evaluation factors, products, approaches, and methods in the three different documents of Ablemagic. Each of the documents is presented in detail in the following subsections.

[1] zoom.us

Table 4.1: Distribution of the Creativity Evaluation Methods in the different documents of Ablemagic. (Appr. = Approaches)

	Factors											Products									Appr.			Methods						
	Valuable	Novelty	Fluency	Enjoyment	User feeling	Collaboration	Expressiveness	Immersion	Flexibility	Interaction	Other	Idea generation	Model making	Storytelling	Game	Music	Drawing	Programming	Visualization	Writing	Psychophysical	Self-reporting	External judges	Observation	Questionnaire	Interview	Qualitative	Quantitative	Adults	Children
Design principles	■			■	■	■	■		■	■	■								■		■			■				■		
Observation guide	■			■	■	■	■			■							■				■			■				■	■	
Observation notes	■	■		■	■	■	■			■							■				■	■		■		■	■	■		■
	3	1	0	3	3	3	3	0	1	3	1	0	1	1	1	1	3	0	1	0	3	1	0	3	0	1	1	3	1	3

The Design Principles

The development of the product Munti Box and the subsidiary company Munti Magic formed the basis of the nine design principles of Ablemagic. The principles indicated their focus on experience economy, which spotlights the value of the user's experience of digital CSTs. The CST must not disturb the spatial environment. Inspired by Japanese projects, Ablemagic is concerned with using the whole room by actively involving the physical environment in their digital CSTs for children.

The digital CSTs should have an intuitive and entertaining experience that begins on the first touch, and one can do as long as the time permits. In contrast to regular games, there is nothing to finish. Furthermore, their CSTs are for the whole family. Even though children are the target group, it is important that the rest of the family could watch it. Several children could also use the CST at once to create collaboration. The children should also be interactive co-creators with the CST to get indirect learning. To determine if this is working as intended, Ablemagic needs to do many tests on the target group.

The design principles used the term *intuitiveness* to evaluate the factor *valuable*. Further, the term *playfulness* was in addition to *enjoyment* utilized to evaluate *enjoyment*. Besides, the design principle included the factor *exploration* in the *Other* category.

The Observation Guide

The observation guide of the user test of a digital drawing CST for children comprised the observation questions and the CST features Ablemagic wanted to investigate. In this case, the features were drawing sheets, equipment, a scanner, and a screen. To evaluate the factor *valuable*, the guide utilized the terms *intuitiveness*, *aesthetics*, and *logical*. Besides, it was used the term *duration* in addition to *enjoyment* to evaluate *enjoyment*.

The Observation Notes

The observation notes encompassed the description of the user test of the digital drawing CST, including the observation questions, the results of the test, and a conclusion based on the results. The goal of the evaluation was to test different drawing sheets in the CST.

The participants were children between five and eight years old in an after-school program nearby. They were divided into two different groups with 10 and seven children, respectively. The last group used a prototype of the CST in contrast to the first group working as a control group. The data collection method was mainly observations of the children during the activity to compare the group using the prototype with the control group. Besides, the user test facilitator conducted a short group interview of the children's experience of the digital CST after the activity.

The observation notes used the term *ascetics* to evaluate the factor *valuable*. To evaluate *enjoyment*, the notes utilized the terms *duration* and *fun* in addition to *enjoyment*. It was also used the terms *result satisfaction* and *preference* in addition to *user feeling* to evaluate *user feeling*.

4.2.2 Observations of the Formal Meetings

A creativity meeting, a creativity discussion, and a planning meeting of an evaluation were observed. Each of these meetings is presented in the following subsections.

The Concept Development Meeting

The software development team in Ablemagic shares an old warehouse with several other companies. The employee sits in two different rooms, one room for the technical roles and one for the administrative roles. The door between the rooms is open. In each room, everyone sits around a table with an own computer. Besides, Ablemagic has an own meeting room with a big table, several blackboards, papers, post-it notes, and windows available to draw on. This room was the location for the concept development meeting. The subsidiary company Munti Magic is located in the room next door. The equipment is placed in this room.

The process of Ablemaic starts with clarifying the mission in the concept development phase. Further, there is the insight work, the idea work, and the evaluation investigating if the proposed solution is a possible solution. This is done by evaluating the following seven evaluation criteria: the distribution of the physical and the digital, the collaboration, the use of time, the immediate

understanding of the task, the need for follow-up, the effort, and the wow effect. These criteria utilize the evaluation factors *interaction*, *collaboration*, *enjoyment* (*duration*), *valuable* (*intuitiveness* and *logical*), and *user feeling*, respectively. The evaluation must be conducted with children in their environment. Ablemagic does not use the service design method with testing on the walls but tests in the target environment instead. Ablemagic has many parallel projects, and the employee has to find a balance between the work in the company and the work in the subsidiary company.

In the concept developing meeting of Ablemagic, it was important with a spatial understanding of the context. They utilized something they call *verbing*, which is to use active verbs in the discussion. Their inspiration is the book *The art of game design* by Schell [85]. The communication in the team in the concept developing meeting was honest, open, and direct. At meetings in Ablemagic, everyone can talk about everything independent of their role. The flow between the different team members was excellent. One of the members drew out the ideas, while the two others discussed them. They quickly pointed to the difficulty in another's solution. Still, the communication was always positive.

One example was the discussion regarding the challenge of finding the balance between control and freedom in the digital drawing CST. One of the members pointed out: "*The concept should not control the children*". Moreover, another member commented: "*It is important the children draw something that fits with the concept. The drawings must not be ugly in the digital CST*". The third member then highlighted: "*It is possible with some background figures on the drawing sheet to control the theme. Furthermore, these figures could be animated in the digital CST to make it more alive*". The second member then continued: "*This could be evaluated on children using sheets with different types of figures and blank sheets to observe how this impacts the motives of the outcome*".

Another example was the challenge of inspiration from other's work. A team member mentioned: "*Others could impact the children if they see their results before drawing*". Another member then reflected: "*This can lead to rush the drawing to see the result*". The third one commented: "*An alternative is to scan all drawings simultaneously, so no one sees the drawings' transformation to the result before they are finished with their drawing*".

The team members thought the most important evolution factors of creativity in their CSTs are *collaboration* and *user feeling*. Ablemagic wants to have concrete evaluation factors, but if this is difficult to do automatically with General Data Protection Regulation, it is possible to observe it manually with random sampling.

The Creativity Discussion

Creativity is important to evaluate in Ablemagic as it is the common denominator in all of their products. Table 4.2 illustrates how the team members disagreed on which of the ten evaluation factors were most important to Ablemagic. All picked *flow*, but they did not think about the flow of ideas like it is defined in the creativity context in section 3.2.2. Instead, they thought of *experience flow*. In Ablemagic, the experience flow is an important part of *intuitiveness*. The experience flow is how the users understand the different parts and the flow of the CST. Two of the employees picked *enjoyment*, *collaboration*, *user feeling*, and *interaction*. A developer meant *interaction* was most important. One picked *novelty*. The tester picked *expressiveness* because of the importance of showing the participant's result to others.

Table 4.2: The evaluation factors the employees of Ablemagic thought were most important amongst the 10 factors.

Factor	Number of employees
Fluency	3
Enjoyment	2
Collaboration	2
User feeling	2
Interaction	2
Novelty	1
Expressiveness	1

The employees were positive to use these ten factors in future CEMs in Ablemagic. On the other hand, they agreed that ten factors would be too much and found a maximum of five factors is more suitable for evaluating a digital CST on a given evaluation. Which of these factors is chosen for a specific CST depends on the product. Ablemagic wants to primary manually observe the factors. It is also desirable to have something digitally afterward. Examples launched by the employees were to track the *interaction* and measure the *duration* in front of a motion detector.

The employee totally agreed that observations are Ablemagic's most important data collection method today. Ablemagic could supplement with questionnaires and interviews if necessary. They also agreed the data analysis is mostly qualitative. They conduct user tests of different age groups in different environments. To avoid start problems for the children in the user test, they have found out it is best to inform the participant about the purpose of the task at the start of the test. They do not investigate gender differences in the user tests.

The Planning Meeting

The tester found *expressiveness* and *valuable* in the form of *intuitiveness* and *aesthetics* the most important evaluation factors of creativity for Ablemagic. To evaluate *expressiveness*, possible observation questions are *Do the users express themself as they want?* and *Do the animations in the display destroy the drawings?*. Accordig to the tester, *interaction* is also an important factor. It is interesting to observe the interaction between the participants and the digital CST and the interaction between the participants and the environment. Further, the CST must be *fun*, a sub-term of *enjoyment*. The tester was very positive regarding using Smileyometer to evaluate this. Besides, the tester was optimistic regarding using *This or That* to let the children tell which one they prefer of the two different CSTs. To evaluate the factor *collaboration*, one could observe the inspiration by the other drawings, communication between the participants when they draw and scan, and discussions after the drawings are displayed. *Novelty* is another factor that could be evaluated in the drawing CST. A possible observation question is: *Does the drawing paper restrict the motive?*.

The practical implementation of the test will use two rooms: one room with the digital drawing CST and another room with the digital visualization CST. The children could go back and forth between these two CSTs. It is important not to inform the participants too much about how the CSTs work. The maximum number of participants is 15. If it is many participants, Ablemagic will consider dividing them into two groups. The tester will facilitate the test. The developers observe the opposite CST than they have worked with. I will observe the same as the tester and a developer using the observing guide and compare afterward.

It was a beneficial experience for the tester to discuss the user test. The tester elaborated: "*It is not time to understand things properly during individual work, and it helps discuss it with others*".

4.2.3 Interview about the Software Development Process

The Design-thinking Process

According to the tester, the software development process of Ablemagic utilizes a design-thinking process that comprises iterative cycles. The tester reflected on the iterative process: "*Sometimes it can be very frustrating that we make so many iterations, but we have to do it this way to create the best concept*". This process circulating between developing and evaluating is experienced very flexibly. It is made a lot of prototypes in the process that are consecutively evaluated, and it is continuous problem-solving of occurred issues. This process is experienced

faster by the tester than expected from studies and other work.

The Concept Development Phase

In Ablemagic, the software development process begins with the concept development phase, where the software engineers collaborate across disciplines. The tester explained the company's openness: "*We also have a very open mode where there do not exist bad ideas. We are completely open at first, and we rather scale down afterward to redo the things we do not achieve*". As stated by the tester, they work based on the thought that anything is possible technically.

The Insight Phase

Further, it is the insight phase comprising interviews, workshops, and observation of relevant software. The concept developers also examining relevant theory. The phase is affected by a lack of time, and the tester wishes that there was more time to read method theory. The tester commented: "*In a recent large project, the insight phase lasted for almost a year. [...] Of course, we do like that on a smaller scale on the other projects, but we do not always have time for the large insight phase*".

The Creativity Evaluation Method

Ablemagic evaluates as early as possible. The tester elaborated: "*As soon as we have something — some function, some factor — that we want to test, we try to test it on the target group. Or internally*". According to the tester, they could evaluate internally when there are many mistakes. Still, they prioritize doing the evaluations with the target audience because they could miss significant aspects without including them. The target audience of Ablemagic is relatively wide, mostly between five and 10 years old. Ablemagic also wants to approve the prototype before proceeding to the next phase. As claimed by the tester, it is approximately a monthly evaluation in the company.

The recruitment method of Ablemagic is snowball sampling, where the employees ask acquaintances or an after-school program nearby. The tester reflected on the challenge of recruiting children: "*You can somehow not only post on Facebook and hear if any children want to join*". As stated by the tester, the sample size depends on the CST. Sometimes they could test to investigate if the experience works for a certain number of participants, but there are approximately 10 children on smaller user tests. They work for differentiation regarding gender, age, and ethnicity,

but the latter is challenging to achieve. Differentiation is easier when they use pilots, according to the tester.

As claimed by the tester, the most important data collection method for Ablemagic is observations. They want to test in the correct environment due to the importance of spatial experience. Thus, they use pilots of the CST in the correct environment when it is almost finished. Ablemagic often divides the participants into two groups — to observe one group and observe and ask questions to the other group. They also utilize interviews, both individual and focus groups of families. The tester experience these interviews with children as challenging: "*Children are more unpredictable [than adults], I think. Very unpredictable. [...] But it is so fun to test with children because they are as honest as you can be*". The least important data collection method for Ablemagic is questionnaires, but they sometimes use it to start the children's thought process.

The evaluation factors in Ablemagic are based on their own design principles. These principles form the goal for the CST and are a part of their culture. The design principles are based on several years of research on accessible installations for children in public spaces, according to the tester. *Intuitiveness*, *collaboration*, *user feeling*, and *expressiveness* are the factors that appear in these principles that Ablemagic sees as most important for the company. They also find *novelty* important, even though it is not included in the principles. The tester also highlighted to limit the control of the evaluation: "*Another thing that I forgot to say that is essential to us, and repeated in all our tests, is to really keep the mind as open as possible. Of course, some framework on how we should evaluate, but also be very open so we can be surprised by some findings*".

4.3 Discussion

4.3.1 Developing of the tailored Creativity Evaluation Method

The findings reveal Ablemagic as a high-functioning software development team heavily focusing on creativity, holding regular creativity meetings during the whole software development process to target integrating the quality creativity in their products. The design-thinking process of Ablemagic centers on iterations of prototype development and evaluations. The focus is not only on implementing software features; the solution is also consecutively discussed from the perspective of children.

The development of the CEM for Ablemagic was based on their existing CEM and the recommendations from the SLR in section 3.4. The findings indicate that Ablemagic has a good starting point, already following some of the recommendations. Based on the design principles,

Ablemagic has a clear goal of the CEM (**REC1**). The observation signaled Ablemagic was positive regarding utilizing the 10 evaluation factors as a starting point (**REC2**). The findings indicate that they already used some of these factors in their CEM. Only one of the factors in the findings, *exploration*, was not encompassed among these 10 factors. However, Ablemagic used various combinations of factors for different CSTs, and the employee disagreed in the formal meetings on which of these factors were most important. On the other hand, they agreed using all these factors to evaluate a specific CST is too much.

The documents, the observations, and the interview revealed Ablemagic utilizes a psychophysical measurement approach with observations that is more qualitative than quantitative (**REC3**). When needed, they use interviews and questionnaires and welcomed utilizing the child-friendly self-reporting questionnaires Smileyometer and This or That. Moreover, the CEM of Ablemagic is not optimized. In the interview, the tester expressed a wish for more time to study the theory of the method. The CEM is not standardized as the documents revealed that the factors in the evaluation criteria did not completely match the factors in the design principles and earlier evaluations. Thus, Ablemagic needs to develop a tailored approach for their CEM of digital CSTs (**REC4**).

4.3.2 Limitations

It is difficult to declare the existing CEM of Ablemagic as the documents, the observations, and the interview revealed that their earlier CEMs utilized different, factors, products, approaches, and methods. Still, the tailored CEM of Ablemagic could be more standardized regarding the methods and tailored to the digital CST that should be evaluated.

More data on Ablemagic may have given a better understanding of the company. On the other hand, data collection is time-consuming, and Ablemagic had no more documents to send. The method triangulation gave a broad data collection and emphasized the findings as the information from the different data collection methods concurred. The development of the tailored CEM was also based on the well-grounded findings in the SLR. Besides, the respondent validation with Ablemagic validated the findings from the documents, the observations, and the interview.

4.4 Take-Home Message

Based on the insight in the existing CEM of Ablemagic and the findings in the SLR, the focus areas to tailor the CEM of Ablemagic were to use the 10 evaluation factors more systematically

and utilize more qualitative methods like the self-reporting questionnaires Smileyometer and This or That.

Chapter 5

Evaluation of the tailored Creativity Evaluation Method

The tailored CEM of Ablemagic was evaluated between Mars and April 2021. This chapter describes the evaluation. The paper summarizing the evaluation that was submitted to the *14th International Conference on the Quality of Information and Communications Technology* could be found in Appendix E. The last user test and the documents of evaluation routines were not included in the paper because it happened after the submitting deadline.

5.1 Method

This section presents the method used to evaluate the tailored CEM of Ablemagic, the last step in figure 1.3. The overview of the method of the thesis is presented in figure 1.2.

5.1.1 Goal Question Metric

The evaluation of the tailored CEM of Ablemagic was based on the Goal Question Metric (GQM), an effective data collection method for evaluating software development methodologies [86]. The GQM comprises three levels:

- The conceptual level **Goal**: Defines the specific goal for the object;
- The operational level **Question**: A set of questions utilized to characterize how the goal will be measured;
- The quantitative level **Metric**: A set of data associated with each of the questions to answer them quantitatively.

Figure 5.1 presents the GQM for the evaluation. The questions and metrics were adapted from an examination of evaluating the individual acceptance of methodologies by Riemenschneider, Hardgrave, and Davis [87]. This examination investigated existing models of individual acceptance of IT tools, including Technology Acceptance Model (TAM), TAM2, Perceived

Characteristics of Innovating (PCI), Theory of Planned Behavior (TPB), and Model of Personal Computer Utilization (MPCU).

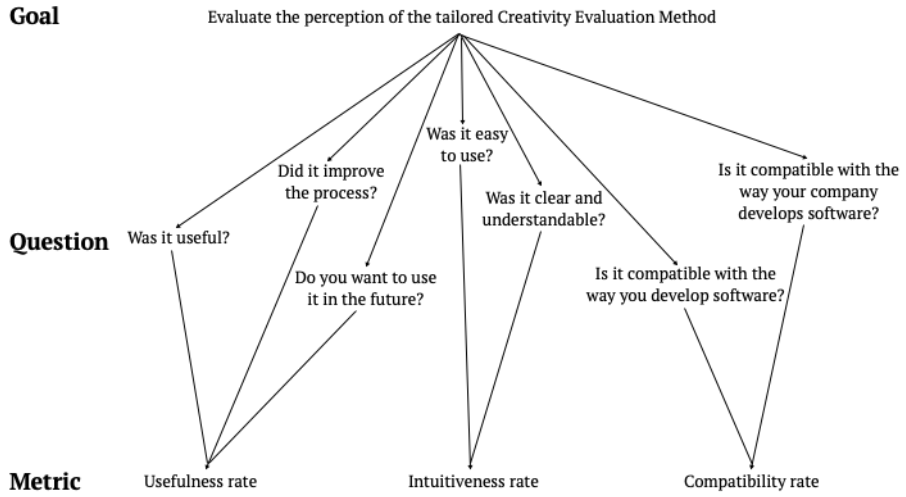


Figure 5.1: Goal Question Metric of the evaluation of the tailored Creativity Evaluation Method of Ablemagic.

The TAM explains the organizational workplace's adaption of IT apps [88]. This model was later extended to include mandatory and voluntary usage situations in the TAM2 [89]. Moreover, the PCI is the personal workstations' adaptation of IT innovations [90]. In contrast, the TPB links beliefs to behavior utilizing three primary determinants of intention [91]. The MPCU redefined the TPB to estimate personal computer utilization behavior [92].

These models utilized different metrics. To evaluate the tailored CEM, it was most relevant to assess *usefulness*. All these models included *usefulness*, and Riemenschneider et al. [87] claimed that the developers must find a methodology useful (i.e., the users' perception of improved job performance by using the methodology) if the methodology should succeed. Furthermore, *intuitiveness* (ease of use) was included in the evaluation because all the models except TPB assessed ease of use (i.e., the users' perception of free of effort by using the methodology). Lastly, the evaluation included *compatibility* as Riemenschneider et al. [87] found *compatibility* of the methodology driving the methodology adoption intentions. *Compatibility* (i.e., the degree to which the users perceived using the methodology as consistent with the existing experience, needs, and values of potential adopters) was only included in PCI.

5.1.2 Data Collection

The evaluation of the tailored CEM of Ablemagic comprised two observations of user test utilizing the CEM, an interview evaluating the CEM, and documents of the evaluation routines of Ablemagic. The data analysis described in section 4.1.1 was utilized for all these data collection methods. Details regarding the data analysis of the different data collection methods are presented in the following subsections.

Observations

Two user tests utilizing the tailored CEM of Ablemagic were observed with participant observation as described in section 4.1.2. To investigate how it was to use the CEM in practice, I observed the software engineers conducting the user test through the experience as an observer in the user test following the tailored CEM. The duration of the user tests was one hour and 15 minutes and two hours, respectively. It was used the observation template presented in section 4.1.2. The data sources also included the artifacts of evaluation templates and results of the user test.

The first user test was conducted in Mars. This test was the planned user test of the digital drawing and visualization CSTs. The participants were 10 children from the after-school program nearby between seven and eight years old. After a while, the leader of the after-school program was also allowed to try the drawing CST.

In April, it was a new user test of a modified version of the digital visualization CST. The participants were family members of the employees in Ablemagic. A total of five children between three and nine years old and one of the parents participated. The children were divided into different groups with two, one, and two participant(s), respectively. It was tested two different versions, one with animations and one without animations. The groups of two children tested the version without animations before the version with animations, and the group of one child tested only the version with animations.

The observations were coded in an exploratory way. Still, they ended up with a total of the same four higher-order themes emerging from the observation notes: the test setup, the team, the CEM, and evaluation of the user test.

Interview

The insight from the observations was combined with a personal interview of the tester in Ablemagic regarding the first user test and the tailored CEM. The interview was semi-structured and recorded in Zoom as described in section 4.1.2. The GQM formed the basis of the questions regarding the tailored CEM. The interview guide is attached in Appendix C.2. The interview was categorized with a top-down approach based on this, resulting in four higher-order themes (the user test, usefulness rate, intuitiveness rate, and compatibility rate) and 31 sub-themes.

Documents

The accumulated insight from the observations and interview were triangulated with two documents of the evaluation routines of Ablemagic. The tester made these documents after the user tests. It was extracted and coded data in the same way as the documents presented in section 4.1.2.

5.2 Results

5.2.1 Observation of the first User Test

Figure 5.2 presents the test setup of the digital drawing CST. In the middle of the main room of the location of Ablemagic, it was a table with colorful markers in cups. On another table, it lay four different drawing sheets with different motives. Besides, it was on a stool with a suitcase covered with silver glitter paper. The prototype of the digital drawing CST was placed on a third table. The prototype comprises four screens connected to a computer to show the scanned drawings. Two of the screens showed curtains and the two other blinds. The digital visualization CST was in the location of Munti Magic. Other CSTs were placed in the hallway between the location of Munti Magic and Ablemagic.

Before the user test started, the atmosphere in Ablmagic was a bit stressed. Approximately half of the employees in Ablemagic participated in the user test. The test was facilitated and observed by the tester. The drawings in the suitcase were consecutively collected and scanned in the office by one of the designers. One developer was responsible for the revealing of the scanned drawings on the screens. A bachelor student and I observed the digital drawing CST. The digital visualization CST was facilitated by another developer and observed by the person concerned, in addition to the last bachelor student and developer.

Figure 5.2: The user test setup of the digital drawing Creativity Support Tool.

The tester had made an evaluation template based on the user test planning meeting. This template laid the foundation for the CEM. Table 5.1 presents the distribution of evaluation factors, products, approaches, and methods in the two different CSTs. To evaluate the factor *enjoyment*, an analog Smileyometer comprised of Ablemagic's own figures was utilized. The Smileyometer is depicted in figure 5.3. The poster with the Smileyometer was laid on the table by the facilitator. All the children should at the same time point at the smiley representing their perception of the CST. Moreover, the facilitator asked the children why they chose specific smileys to get the children to elaborate their perception. As the remainder of the factors, *enjoyment* was also evaluated in the observation through the different observation questions. Both of the CSTs evaluated *enjoyment* with the term *enjoyment*. Besides, the drawing CST utilized the terms *duration* and the visualization CST *duration* and *fun*. To evaluate *valuable*, the drawing CST utilized the terms *usefulness* and *logical*, while the visualization CST used *intuitiveness* and *logical*.

Table 5.1: Distribution of the Creativity Evaluation Methods of the two different digital Creativity Support Tools (CSTs) for children in the user test. (Appr. = Approaches)

	Factors											Products									Appr.			Methods						
	Valuable	Novelty	Fluency	Enjoyment	User feeling	Collaboration	Expressiveness	Immersion	Flexibility	Interaction	Other	Idea generation	Model making	Storyteling	Game	Music	Drawing	Programming	Visualization	Writing	Psychophysical	Self-reporting	External judges	Observation	Questionnaire	Interview	Qualitative	Quantitative	Adults	Children
Drawing CST																														
Visualization CST																														
	2	0	2	2	2	1	1	0	0	0	0	0	0	0	0	0	1	0	1	0	2	2	0	2	2	2	2	2	1	2

The employees discussed how they should change the CSTs based on the result of the evaluation on their Slack workspace and in the office immediately after the test. Further, the user test was in plenary discussed after the children had left. The employees agreed the Smileyometer worked in their CEM of digital CSTs for children and that the evaluation was successful. Besides, I asked the two bachelor students how they experienced utilizing the observation guide. They found it beneficial to start focusing on something concrete not to waste time on irrelevant aspects.

5.2.2 Interview evaluating the tailored Creativity Evaluation Method

The User Test

The tester told the employees in Ablemagic were satisfied with the first user test of the digital drawing CST. This evaluation was mainly about being surprised by the childrens' reactions and testing whether the animations made the CST magical. The evaluation quickly revealed what worked in the CST. This user test was not located in the correct environment, but the other digital CSTs that the children could try after the drawing made it a bit realistic. The tester highlighted that this demonstrates the balance between development and evaluation. Further, the tester reflected on this balance: "*But it just has to be that way. I think it's best in this order because then you kind of avoid developing something that does not work at all*". According to the tester, the focus of the next evaluations will be to polish the digital drawing CST. Equipment and animations could be evaluated internally in Ablemagic, but the experience flow of the CST needs to be evaluated with the target group. Furthermore, the prototype should be more accomplished. They also want a family workshop with a wider age group, including children from four to 12 years old. The next test will also be conducted in the same test location, as they still do not have access to the correct test environment.

Usefulness Rate

The evaluation template made by the tester comprises the goal for the digital CST, the goal for the test, and the ten evaluation factors. For each evaluation, the most important factors should

be determined and further used to make observation questions. As claimed by the tester, it is time-saving and important to have a standardized test frame. Ablemagic could later expand the CEM to conduct interviews based on the selected factors. Further, the tester explained that with the use of this evaluation template, "*we knew for sure that we have somehow elaborated everything, all factors, and this is what is important to focus on in this test*". The observation questions reflected Ablemagic's preliminary thoughts. The tester emphasized the assurance of the fact that the template was academically grounded and not just experience-based. It helps to find out how to test, evaluate and make progress. The tester concluded that the Smileyometer worked excellent for children to evaluate the *enjoyment* of a digital CST. Still, the test exposed that the children did not always understand which of the CSTs they should evaluate with the Smileyometer, so the employees did not find the user test of the visualization CST completely valid. The tester underlined one must be very concrete and visual with seven to eight years old children and remarked: "*So I do not want to say the evaluation [e.g., the questionnaire and the interview] is weighted so highly with this age group. It is more the observation and such - what they are talking about, what they answer*".

Intuitiveness Rate

The CEM was experienced easy to use by the tester. In the process of understanding the CEM, the tester transferred it easily to Ablemagic's own design principles. The tester underlined that some of the factors were not suitable for the CSTs evaluated in the user test but remarked: "*Some factors may not have been so immediate or fit so well, but they may fit later*".

Compatibility Rate

The CEM was compatible with the way the tester evaluated digital CSTs. The tester pinpointed that the user must understand the method and continued: "*It is a method that I would not be able to use without working with it myself*". The tester experienced it good to have something concrete as a basis. The CEM was also compatible with the way Ablemagic evaluated digital CSTs. The tester will present the CEM at the user test evaluation meeting and make a part of what they call *the method of Ablemagic*. The tester emphasized the importance of stopping and evaluating the method, and the whole team must be involved in this process. Moreover, the tester suggested: "*So that, maybe I should set it up as a point, that I should bring it up, after all this. We'll see if time allows*".

5.2.3 Observation of the second User Test

The test setup of the digital visualization CST was the same as the first user test described in section 5.2.1. Three employees in Ablemagic participated in this user test in addition to me. One developer managed the CST and observed the test together with the tester facilitating the test, a designer, and me.

The evaluation used the same observation template as the first user test, including observation questions and evaluation factors. Thus, the CEM is presented in table 5.1. The only difference in the CEM was the fact that one adult participated in this test.

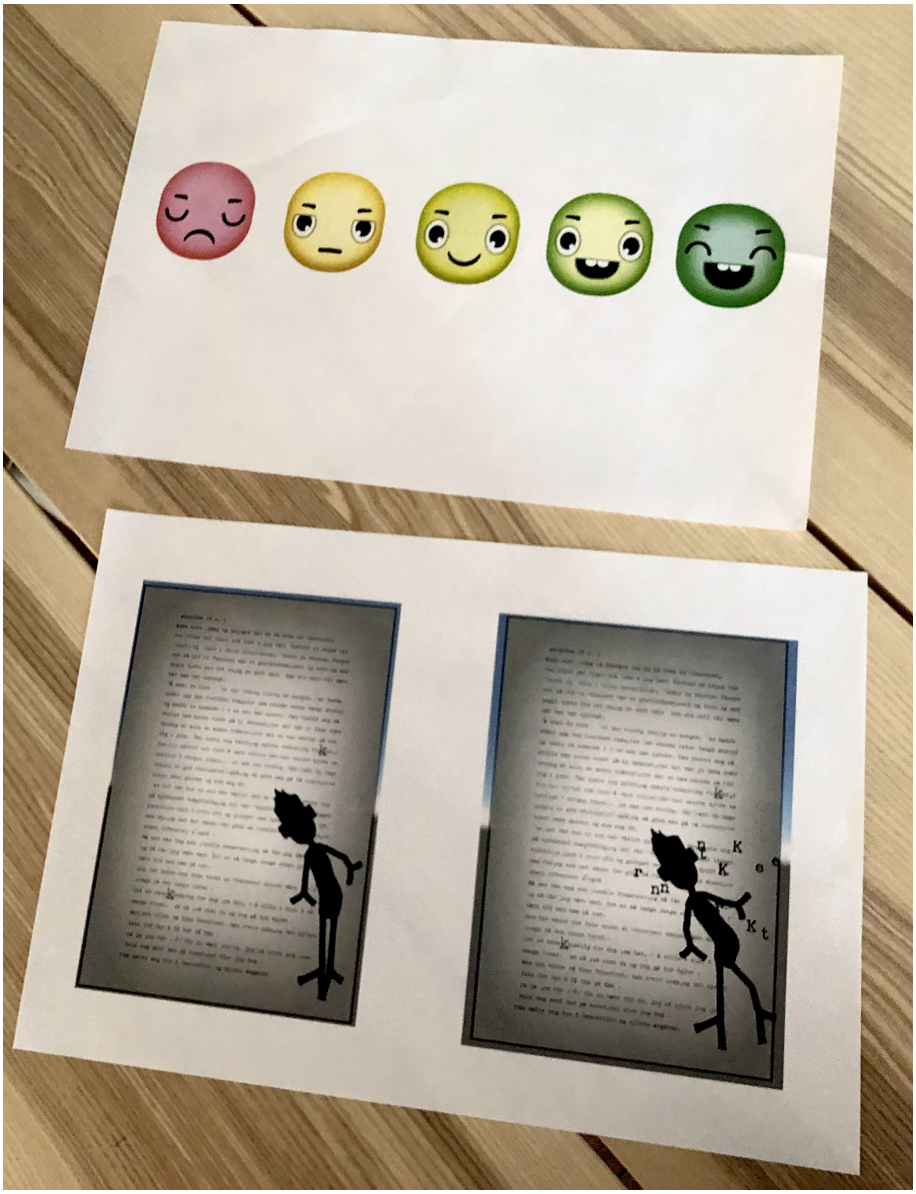


Figure 5.3: The Smileyometer and This or That.

After each test, the children evaluated the CST with the same analog version of a Smileyometer as described in section 5.2.1. The groups who tested both versions also evaluated the CST with an analog version of This or That by pointing at the most preferred version of the CST. The Smileyometer and This or That are depicted in figure 5.3. The youngest child did not use This or That. Smileyometer and This or That worked well with the children and got them to talk about the CST when the facilitator asked why they chose a specific smiley or option. It seems that the children understood what they should answer in these different self-reporting questionnaires. In contrast, the youngest one picked a smiley that not represented the child's outward *enjoyment*. The employees concluded that this was since the child liked that specific smiley the most.

The employee started to discuss a test immediately after a group finished. The focus of the discussion was opportunities for change and improvement based on the findings from the test, the concept, and the budget.

5.2.4 Documents of Evaluation Routines

Table 5.2 presents the distribution of evaluation factors, products, approaches, and methods in the updated evaluation routines of Ablemagic. These routines comprise information describing Ablemagic and Munti Magic, the process, and the design principles (described in section 4.2.1). In addition to the evaluation factors included in the design principles, the routines have an own page with evaluation factors, including *valuable* (and its sub-theme *intuitiveness*), *novelty*, *enjoyment* (and its sub-theme *duration*), *collaboration*, *flexibility*, *expressiveness*, and *interaction*. Lastly, it is a page with an example of an observation guide of a drawing CST. The tester informed that these documents were sent to Ablemagic's latest customer.

Table 5.2: The Creativity Evaluation Method of the updated evaluation routines of Ablemagic. (Appr. = Approaches)

	Factors											Products									Appr.			Methods						
	Valuable	Novelty	Fluency	Enjoyment	User feeling	Collaboration	Expressiveness	Immersion	Flexibility	Interaction	Other	Idea generation	Model making	Storytelling	Game	Music	Drawing	Programming	Visualization	Writing	Psychophysical	Self-reporting	External judges	Observation	Questionnaire	Interview	Qualitative	Quantitative	Adults	Children
Evaluation routines																														

5.3 Discussion

5.3.1 Evaluation of the tailored Creativity Evaluation Method

The findings indicated that the tailored CEM of Ablemagic had good usability, intuitiveness, and compatibility rate. The observations of the discussions in the software development team after the user tests finished signified that the CEM helped to identify the prototypes' limitations, something the tester also emphasized in the interview. It was also very promising that the tester made an evaluation template based on the findings from the SLR. The bachelor students preferred to utilize the observation guide as a starting point to observe the user test. Besides, the CEM became a part of the method of Ablemagic in the updated documents of evaluation routines. This document has also Ablemagic begin to send to costumers.

The observations of the user test indicated the analog Smileyometer and This or That worked well to get children to evaluate *enjoyment*. It also helped the children talk about their experience of the digital CST when the facilitator asked why they chose as they did. The observations revealed the children understood how to use these different questionnaires, but for children under the age of four, the validity of their answers should be carefully considered. In the second user test, the youngest child picked a smiley not representing the outward *enjoyment*, probably because the child preferred that specific smiley. The tester pinpointed in the interview that the observation is the most valid data collection method for children. To evaluate several CSTs consequently, the observation of the first user test indicated Smileyometer did not work so well. Instead, This or That worked to identify children's preferred option in the second user test. In the future, it is a possibility that the Smileyometer and This or That could be implemented in the digital CST.

5.3.2 Limitations

The interdisciplinary research of creativity, CCI, and SE made it difficult to determine the research methods. There was no research investigating how to evaluate an evaluation method, but the SE approach GQM seems to fit such cases [86]. This approach has never been used in the creativity area before but is utilized in CCI (e.g., Kusumo, Sabariah, and Wiharj [93]). The question and metrics in the GQM were adapted from Riemenschneider et al.'s [87] examination of evaluating the individual acceptance of methodologies. Moreover, this examination was based on five existing SE models found transferable to evaluate a method, but they modeled individual acceptance of IT tools. The models utilized 12 different metrics, where the three metrics that were found most relevant were adapted. Nevertheless, there is a possibility that some of the remaining metrics also could be helpful for the evaluation. An example is *subjective norm* (i.e., the degree to which a person's motivation to perform a behavior is based on what people who

are important to them think), included in TAM2, TPB, and MPCU [87].

It was only observed two different user tests utilizing the tailored CEM of Ablemagic. Multiple user tests evaluating more than two different CSTs and more participants had given a broader evaluation of the CEM. Unfortunately, it was no more user test in Ablemagic before the deadline of the master thesis due to the COVID-19 restrictions. Still, the two user tests indicated if the tailored CEM of Ablemagic worked or not.

It had also been advantageous with more than one interviewee to get other perspectives than the tester. However, only this employee had at that time enough knowledge of the tailored CEM of Ablemagic. Besides, the interview was delayed two times because of the large amount of work in Ablemagic, so I did not want to take more time from the company. Nonetheless, multiple employees were observed during the user tests.

5.4 Take-Home Message

The results indicated that the tailored CEM of Ablemagic is promising with good usability, intuitiveness, and compatibility rate. The tester in Ablemagic adopted the CEM and discussed it with the rest of the team to integrate it into their evaluation routines.

Chapter 6

Discussion

This chapter discusses the RQs, the implications for research and practice, the limitations and lessons learned of this master thesis.

6.1 Research Questions

6.1.1 RQ1: How to Evaluate digital Creativity Support Tools for Children?

The SLR of 81 articles identified a great diversity of CEMs of digital CSTs for children caused by the lack of a standardized CEM. The CEMs utilized multiple evaluation factors, but *valuable*, *novelty*, *fluency*, *enjoyment*, *user feeling*, *collaboration*, *expressiveness*, *immersion*, *flexibility*, and *interaction* were the 10 most common factors. The CEMs were performed in different contexts using various products, approaches, data collection methods, and data analysis.

The age of the target group impacted the CEMs. For children and adults differed the distribution of the methods, products, and factors. For example, the factors *novelty*, *enjoyment*, and *collaboration* were more utilized in the studies with children as participants, while *valuable*, *fluency*, and *user feeling* were used more for adults. Storytelling and programming products were more used in the studies targeting children, in contrast to idea generation and drawing products for adults. The CEMs of adults included more qualitative questionnaires with self-reporting, while children were evaluated more objectively utilizing a larger amount of quantitative psychophysical measurements with observation. Besides, the age of the children impacted the CEMs, as it for younger children is more important with quantitative observations than qualitative questionnaires and interviews.

The effectiveness of CEMs is difficult to determine as the majority of the included studies did not specify the threats to the validity, and the outcomes vary. The studies did not either have a uniform understanding of creativity using different definitions. Besides, multiple studies did not define creativity. This was also the case for the majority of the evaluation factors. Consequently,

a comparison between the included studies is difficult. However, the most uniform evaluation factor was *flexibility*, where most studies provided a similar definition.

Based on the findings from the evaluation, the recommendations for CEMs of digital CST for children in section 3.4 grounded in the SLR was upgraded. **REC2** was specified to use the goal to select the evaluation factors because the evaluation of the tailored CEM of Ablemagic indicated the CEM depended on the CST that should be evaluated. **REC3** was elaborated and divided into two parts: one part focusing on quantitative methods and the other on qualitative methods. This was due to the evaluation of the tailored CEM of Ablemagic emphasized the value of utilizing qualitative questionnaires and interviews in the CEM. These changes gave the following five recommendations:

- **REC1)** Clearly define the goal;
- **REC2)** Review and choose based on the goal amongst the 10 evaluation factors *valuable*, *novelty*, *fluency*, *enjoyment*, *user feeling*, *collaboration*, *expressiveness*, *immersion*, *flexibility*, and *interaction*;
- **REC3)** Start with a quantitative psychophysical measurement approach to make observation questions based on the selected factors;
- **REC4)** Expand to qualitative methods utilizing Smileyometer (and eventually This or That if multiple products/versions) to start a discussion regarding the digital CST by asking the children to explain their choice;
- **REC5)** Develop a tailored approach.

The development of the tailored CEM of Ablemagic indicated that to improve a specific CEM of a company, the CEM should be analyzed centered on these recommendations to find the needed changes in the CEM to better evaluate digital CSTs for children.

6.1.2 RQ2: How to Integrate evaluation of Creativity in the Software Development Process targeting Children?

The tailored CEM of Ablemagic seemed compatible, intuitive, and useful. The fact that Ablemagic was already aware of creativity as a quality factor may have made the process of including the tailored CEM easier. Ablemagic hold regular creativity meetings during the whole software development process, and the employees were aware of its value. Thus, making software companies aware of creativity as a quality factor was added as a recommendation regarding integrating evaluation of creativity in the software development process.

The tailored CEM of Ablemagic utilized a mix of experience from earlier work in Ablemagic and theory from the SLR. In the interview regarding the CEM, the tester highlighted the importance of the evaluation factors originating from the theory, ensuring a thoughtful evaluation. It was easy for the tester to transform the tailored CEM from the design principles of Ablemagic that was built on experience. Thus, we recommend to use a mix of theory- and experience-based CEM.

The tester emphasized the value of working with the CEM. This is supported by the importance of employee involvement in change management [94]. Consequently, employee involvement is added as a recommendation. This also indicates it is important to build on the company's software development process to integrate creativity as a quality factor. Therefore, it was added a recommendation of analyzing the development method of the specific company's process to find points to improve.

The evaluation of the tailored CEM of Ablemagic also indicated this process of using a tailored CEM is time-consuming. For example, This or That was not added before the second user test despite the tester was aware of the benefits of both of these self-reporting questionnaires in the planning meeting described in section 4.2.2. Thus, it is wise not to do all the changes simultaneously and to use the time this integration need was added as a recommendation.

To summarize, we have the following five recommendations (RECOs) for integrating evaluation of creativity in the software development process targeting children:

- **RECO1)** Make software companies aware of creativity as a quality factor;
- **RECO2)** Analyze the specific company's process to find points to improve;
- **RECO3)** Use a mix of theory- and experience-based CEM;
- **RECO4)** Involve employees in changing the software development process;
- **RECO5)** Use the time this integration needs.

6.2 Implication for Research

The case study investigated some of the knowledge gaps discovered by the SLR regarding the digital visualization and drawing CSTs for children and the CEMs This or That and Smileyometer evaluating digital CSTs for children. This research corresponded with Jaccheri and Morasca's [1] recommendation of systematically collect and analyze empirical data of software engineers developing software for children.

The thesis contributes to researchers by starting the investigation of creativity as a quality factor,

following Jaccheri and Morasca's [1] recommendation of including investigating other aspects than just technical aspects. The research was submitted as a journal article to *Computing Surveys* and will be submitted as a conference article to the *44'th International Conference on Software Engineering: Software Engineering in Society*. Further research could evaluate the CEM in other companies and other target groups than children to continue this investigation of evaluation of creativity in the software development process.

6.3 Implication for Practice

The research has influenced Ablemagic by improving their software development process of digital CST for children towards a more systematic CEM. The tester made an evaluation template and updated their evaluation routines based on the findings in this thesis.

The findings could also help practitioners in other software companies improve their software developing process by including creativity as a quality factor. This will follow Jaccheri and Morasca's [1] recommendation of software developers designing software for children to consider each of the four qualities in figure 1.1, including creativity.

Improving the software development process for companies targeting children can lead to better digital CSTs as well as other software for children, which further can enhance children's creative skills.

6.4 Limitations

6.4.1 Completeness

The SLR answered **RQ1.1** regarding the current CEMs well, except **RQ1.1.4** since various findings made it difficult to conclude regarding the effectiveness. Following the guidelines by Kitchenham [21], a structured SLR was performed in relevant journals and databases, resulting in 81 peer-review articles with 77 different main authors. Due to continuously new research, it could not be guaranteed that all literature in this area was captured. Table 3.6 illustrated relatively few publications in 2020, but this may be due to the COVID-19 pandemic restriction in research and the automatic search conducted before the year was over. The search strategy was another limitation in the SLR. Relevant studies could be missed since the term *"assessment"* was not included in the search string in the automatic search. Nevertheless, this was compensated

to a certain degree by the manual search opening for more relevant studies using other search strings than the automatic search.

Studies targeting adults were included in the SLR because the research on CEMs of digital CSTs for children was limited. This could be caused by the ethical issues for research utilizing children as participants related to protection, consent, and confidentiality [22]. The inclusion of studies examination adults allowed investigating if the CEMs for adults could be transferable to children's CEMs and comparing the CEMs to the age of the target group in **RQ1.1.3**.

Combined with the case study, the SLR also contributed to a certain degree to answer **RQ1.2** regarding changes needed to improve the current CEMs and **RQ1** regarding evaluating digital CSTs for children. Since the thesis is a case study of the company Ablemagic, the findings could not be generalized to software companies in general. Ablemagic was already aware of creativity as a quality factor in digital CSTs for children, something that could be very different in other companies. To evaluate the CEM in multiple software companies had been beneficial, but it is time-consuming to get into a company. Further research could test the CEM in other companies and other target groups than children to continue this investigation of creativity as a software quality factor. Nevertheless, some aspects of the findings could be transferable to other software companies. Thus, the case study gave only an indication of the answer of **RQ2** regarding integrating evaluation of creativity in the software development process, and further research needs to investigate **RQ2** in other software companies.

6.4.2 Potential biases

The focus on children in the research was challenging from an ethical perspective. Even though the user tests of the CSTs included children, it was chosen to investigate the software engineers during the test, as described in section 1.4.3. Information regarding the children was only introduced in this thesis if it was relevant for the CEM. Moreover, this information was anonymized and not sensitive.

As I was the only observer in most observations, there is no guarantee that others would observe the same as humans notice, ignore, and forget different aspects [2]. An observation guide was utilized to standardize the observations, and notes were taken during the sessions to reduce this potential bias. The two first observations were also observed by a Ph.D. student and a supervisor, respectively, to see if we noticed the same things. Besides, the triangulation of data collection methods and the respondent validation with Ablemagic validated the findings.

The COVID-19 pandemic caused difficultly in the cooperation with Ablemagic. The user tests were delayed because it was problematic for Ablemagic to recruit children to the tests. The strict and constant changing restrictions also made some of the physical meetings digital, including the interviews. The small picture of the interviewee in the digital video interviews made it more difficult to notice the body language than in a physical interview. Body language is an important part of communication that sometimes replaces verbal communication [95]. A digital video interview could also give the interviewee a feeling of confrontation from the interviewer since they sit 180 degrees to one another in an eye-to-eye position. According to Oates [2], the ideal comfortable interaction position between the interviewer and interviewee is to sit 90 degrees to one another because it allows eye contact without any feeling of confrontation.

6.5 Lessons Learned

During this master thesis period, I have learned a lot about research. I had no previous experience with the research of this scale; my earlier research project had lasted for a maximum of three months. It was beneficial to get experience from the whole research process, including both an SLR and empirical research. To start in September and determine the aim of the research, the RQs and the methods for the next ten months were difficult. I experienced that it is important to have flexibility in such a big task because things never go as planned, and you have to be able to change the plan continuously. For example, the RQs made in the research proposal were redefined several times during the process and were not finally defined before Mars.

It was also an educational process of writing articles. I experienced writing both a manuscript and a conference paper, finding a relevant journal or conference to submit them to, and the process of submitting and review. The feedback from experts in the area improved the quality of the thesis. At first, I submitted the conference paper to the *14th International Conference on the Quality of Information and Communications Technology*, but it was not accepted. The feedback from the reviewers helped me improve the thesis and make the message of the research clearer.

The most important lesson of this thesis was how to use the acquired knowledge from an SLR to improve the current SE practice in a company. It really showed me the value of cooperation between researchers and practitioners. Software companies are hectic and mostly experience-based and need help from researchers to investigate the theory to include useful elements from it in their practice. The tester pinpointed the value of this in the interview about the CEM.

The user test with children was very overwhelming and different from earlier user tests I had

experience with adults as participants. The noise level was higher, and it was harder to get them to focus on the task. Some of the children also experienced the user test setting scary. This gave me an important perspective on this investigating of designing software for children.

Chapter 7

Conclusion

This chapter concludes this master thesis by answering the RQs, addresses limitations, states contributions, and propose further work.

7.1 Research Questions

7.1.1 RQ1: How to Evaluate digital Creativity Support Tools for Children?

The SLR found a broad spectrum of CEMs for children performed in different contexts impacted by the target group's age. It is difficult to conclude regarding the effectiveness of the CEMs because of the lack of a standardized CEM. The 10 most common evaluation factors in the CEMs are presented in table 3.7.

Based on the findings from the SLR and the evaluation of the tailored CEM of Ablemagic, we made five recommendations for CEMs of digital CSTs for children:

- **REC1)** Clearly define the goal;
- **REC2)** Review and choose based on the goal amongst the 10 evaluation factors *valuable*, *novelty*, *fluency*, *enjoyment*, *user feeling*, *collaboration*, *expressiveness*, *immersion*, *flexibility*, and *interaction*;
- **REC3)** Start with a quantitative psychophysical measurement approach to make observation questions based on the selected factors;
- **REC4)** Expand to qualitative methods utilizing Smileyometer (and eventually This or That if multiple products/versions) to start a discussion regarding the digital CST by asking the children to explain their choice;
- **REC5)** Develop a tailored approach.

7.1.2 RQ2: How to Integrate evaluation of Creativity in the Software Development Process targeting Children?

Based on promising findings from the evaluation of the tailored CEM of Ablemagic indicating good usability, intuitiveness, and compatibility rate of the CEM, we made five recommendations for integrating evaluation of creativity in the software development process targeting children:

- **RECO1)** Make software companies aware of creativity as a quality factor;
- **RECO2)** Analyze the specific company's process to find points to improve;
- **RECO3)** Use a mix of theory- and experience-based CEM;
- **RECO4)** Involve employees in changing the software development process;
- **RECO5)** Use the time this integration needs.

7.2 Limitations

This master thesis has limitations regarding the choice of research methods and the COVID-19 restrictions during the research period. However, the possibility of generalizing the findings is the main limitation, especially regarding **RQ2**. The case study of Ablemagic gave only an indication of the answer of this RQ.

7.3 Contribution

The master thesis contributes to both researchers and practitioners. This research started investigating creativity as a quality factor in SE, an integration that can give children better software and improve their creative skills. Moreover, the research helps practitioners in software companies towards a more systematic evaluation of creativity as a quality factor to improve the software developing process.

7.4 Further Work

Future research could build on our recommendations to create more knowledge regarding integrating creativity in the software development process by testing the CEM in other companies and other target groups than children.

Bibliography

[1] Jaccheri L, Morasca S. Toward inclusion of children as software engineering stakeholders. 2021 Jan;Available from: https://arxiv.org/pdf/2101.02704.pdf.

[2] Oates BJ. Researching information systems and computing. London, UK: SAGE Publications Ltd.; 2006.

[3] Amabile TM. Creativity in Context: Update to the Social Psychology of Creativity. Boulder, USA: Westview Press; 1996.

[4] Anderson LW, Krathwohl DR, Airasian PW, Cruikshank KA, Mayer RE, Pintrich PR, et al. A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. New York, USA: Longman; 2001.

[5] Csikszentmihalyi M. Creativity: Flow and the Psychology of Discovery and Invention. New York, USA: Harper Perennial; 1997.

[6] Giannakos MN, Jaccheri L, Proto R. Teaching Computer Science to Young Children Through Creativity: Lessons Learned from the Case of Norway. In: van Eekelen M, Barendsen E, Sloep P, van der Veer G, editors. Proceedings of the 3rd Computer Science Education Research Conference on Computer Science Education Research. CSERC'13. Heerlen, Netherland: Open Universiteit; 2013. p. 103–111. Available from: https://dl.acm.org/doi/10.5555/2541917.2541927.

[7] Holmes RM, Gardner B, Kohm K, Bant C, Ciminello A, Moedt K, et al. The relationship between young children's language abilities, creativity, play, and storytelling. Early Child Development and Care. 2019;189(2):244–254. Available from: https://doi.org/10.1080/03004430.2017.1314274.

[8] Hsiao HS, Wong KH, Wang MJ, Yu KC, Chang KE, Sung YT. Using Cognitive Affective Interaction Model to Construct On-Line Game for Creativity. In: Pan Z, Aylett R, Diener H, Jin X, Gobel S, Li L, editors. First International Conference on Technologies for E-Learning and Digital Entertainment. Edutainment'06. Hangzhou, China: Technologies for E-Learning and Digital Entertainment; 2006. p. 409–418. Available from: https://doi.org/10.1007/11736639_52.

[9] United Nations Human Rights. Convention on the Rights of the Child [Internet]. United Nations; 1990 [cited 2020 Sep 23]. Available from: https://www.ohchr.org/en/professionalinterest/pages/crc.aspx.

[10] van Laar E, van Deursen AJAM, van Dijk JAGM, de Haan J. The relation between 21st-century skills and digital skills: A systematic literature review. Computers in Human Behavior. 2017 Jul;72(2017):577–588. Available from: https://doi.org/10.1016/j.chb.2017.03.010.

[11] Papavlasopoulou S, Giannakos MN, Jaccheri L. Empirical studies on the maker movement, a promising approach to learning: A literature review. Entertainment Computing. 2017 Jan;18(1):57–78. Available from: https://doi.org/10.1016/j.entcom.2016.09.002.

[12] Shneiderman B, Fischer G, Czerwinski M, Resnick M, Myers B, Candy L, et al. Creativity support tools: Report from a US National Science Foundation sponsored workshop. International Journal of Human-Computer Interaction. 2006 May;20(2):61–77. Available from: https://doi.org/10.1207/s15327590ijhc2002_1.

[13] Resnick M. Seven issues for creativity support tool researchers. In: Shneiderman B, Fischer G, Czerwinski M, Myers B, Resnick M, editors. National science foundation Workshop Report on Creativity Support Tools; 2005. p. 67–70. Available from: http://www.cs.umd.edu/hcil/CST.

[14] Cassell J, Ryokai K. Making space for voice: Technologies to support childrens fantasy and storytelling. Personal and Ubiquitous Computing. 2001 Aug;5(3):169–190. Available from: https://doi.org/10.1007/PL00000018.

[15] Yarosh S, Radu I, Hunter S, Rosenbaum E. Examining values: an analysis of nine years of IDC research. In: Moher T, Quintana C, editors. Proceedings of the 10th International Conference on Interaction Design and Children. IDC '11. New York, USA: ACM; 2011. p. 136–144. Available from: https://doi.org/10.1145/1999030.1999046.

[16] Read JC, Fitton D, Horton M. Giving ideas an equal chance: inclusion and representation in participatory design with children. In: Proceedings of the 2014 Conference on interaction Design and Children. IDC '14. New York, USA: ACM; 2014. p. 105–114. Available from: https://doi.org/10.1145/2593968.2593986.

[17] Iivari N, Kinnula M, Kuure L. With best intentions: A Foucauldian examination on children's genuine participation in ICT design. Information Technology & People. 2014 Jun;28(2):246–280. Available from: https://doi.org/10.1108/ITP-12-2013-0223.

[18] International Organization for Standardization. Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – System and software quality models [Internet]. Geneva, Switzerland: International Organization for Standardization; 2011 [cited 2021 Mar 17]. ISO/IEC 25010: 2011. Available from: https://webstore.iec.ch/publication/11245.

[19] de la Barra CL, Crawford B, Soto R, Misra S, Eric. Agile Software Development: It Is about Knowledge Management and Creativity. In: Murgante B, Misra S, Carlini M, Torre CM, Nguyen HQ, Taniar D, et al., editors. Proceedings of the 13th International Conference on Computational Science and Its Applications. ICCSA '13. Berlin, Germany: Springer; 2013. p. 98–113. Available from: https://doi.org/10.1007/978-3-642-39646-5_8.

[20] Neverland JM. Adferd i offentlig område med digital lek [Behavior in public areas with digital play] [Master Thesis];. Trondheim, Norway: Norwegian University of Science and Technology; 2018. Available from: https://ntnuopen.ntnu.no/ntnu-xmlui/bitstream/handle/11250/2575863/19280_COVER.pdf?sequence=2.

[21] Kitchenham BA. Procedures for Performing Systematic Reviews [Internet]. Keele University, UK: Department of Computer Science; 2004 [cited 2020 Oct 4]. Technical Report TR/SE-0401. Available from: https://www.bibsonomy.org/bibtex/2e48137ec01b6308876e05ab1ecdf4bc4/wiljami74.

[22] Ferdousi N. Children as Research Subjects: The Ethical Issues. Bangladesh Journal of Bioethics. 2015 Aug;6(1):6–10. Available from: https://doi.org/10.3329/bioethics.v6i1.24398.

[23] NTNU. Collection of personal data for research projects [Internet]. NTNU; 2021 [cited 2021 Apr 17]. Available from: https://innsida.ntnu.no/wiki/-/wiki/English/Collection+of+personal+data+for+research+projects.

[24] NSD. Notification Form for projects processing personal data [Internet]. NSD; 2020 [cited 2020 Sep 28]. Available from: https://nsd.no/personvernombud/en/notify/meldeskjema?eng.

[25] Lovdata [The Lovdata Foundation]. Personopplysningsloven [The Personal Data Act] [Internet]. Lovdata; 2016 [cited 2020 Oct 29]. Available from: https://lovdata.no/dokument/NL/lov/2018-06-15-38/KAPITTEL_gdpr#KAPITTEL_gdpr.

[26] NTNU. Tools for interaction and collaboration [Internet]. NTNU; 2021 [cited 2021 May 22]. Available from: https://innsida.ntnu.no/wiki/-/wiki/English/Tools+for+interaction+and+collaboration.

[27] Resurchify. ACM Computing Surveys [Internet]. Resurchify; 2021 [cited 2021 May 14]. Available from: https://www.resurchify.com/all_ranking_details_2.php?id=2237.

[28] Resurchify. ICSE Ranking and Rating Metrics [Internet]. Resurchify; 2021 [cited 2021 May 24]. Available from: https://www.resurchify.com/conference_ranking_details.php?id=276.

[29] Ritter SM, Rietzschel EF. Lay Theories of Creativity. In: Zedelius CM, Muller BCN, Schooler JW, editors. The Science of Lay Theories. Cham, Germany: Springer; 2004. p. 95–126. Available from: https://doi.org/10.1007/978-3-319-57306-9_5.

[30] Kaufman JC, Beghetto RA. Beyond big and little: The four c model of creativity. Review of General Psychology. 2009 Mar;13(1):1–12. Available from: https://doi.org/10.1037/a0013688.

[31] Oxford Dictionary. Oxford Learner's Dictionary [Internet]. Oxford University Press; 2021 [cited 2021 May 25]. Available from: https://www.oxfordlearnersdictionaries.com/definition/english.

[32] Amabile TM. The social psychology of creativity: A componential conceptualization. Journal of Personality and Social Psychology. 1983 Nov;45(2):357–376. Available from: https://doi.org/10.1037/0022-3514.45.2.357.

[33] Runco MA, Jaeger GJ. The Standard Definition of Creativity. Creativity Research Journal. 2012 Jan;24(1):92–96. Available from: https://doi.org/10.1080/10400419.2012.650092.

[34] Sternberg RJ, Lubart TI. The concept of creativity: Prospects and paradigms. In: Sternberg RJ, editor. Handbook of creativity. Cambridge, UK: Cambridge University Press; 1999. p. 3–15. Available from: https://doi.org/10.1017/CBO9780511807916.003.

[35] Sarkar P, Chakrabarti A. Studying Engineering Design Creativity - Developing a Common Definition and Associated Measures. In: Gero J, editor. Invited Paper in the Proceedings of the NSF Workshop on Studying Design Creativity. Alexandria, USA: National Science Foundation; 2008. Available from: https://cpdm.iisc.ac.in/cpdm/ideaslab/paper_scans/UID_17.pdf.

[36] Jordanous A, Keller B. Modelling creativity: Identifying key components through a corpus-based approach. PLOS ONE. 2016 Oct;11(10):e0162959. Available from: https://doi.org/10.1371/journal.pone.0162959.

[37] Weisberg RW. On the Usefulness of "Value" in the Definition of Creativity. Creativity Research Journal. 2015 Jun;27(2):111–124. Available from: https://doi.org/10.1080/10400419.2015.1030320.

[38] Vygotsky LS. Imagination and Creativity in Childhood. Journal of Russian and East European Psychology. 2004 Jan;42(1):7–97. Available from: https://doi.org/10.1080/10610405.2004.11059210.

[39] Cherry E, Latulipe C. Quantifying the Creativity Support of Digital Tools through the Creativity Support Index. ACM Transactions on Computer-Human Interaction. 2014 Jun;21(4). Available from: https://doi.org/10.1145/2617588.

[40] Tack T, Jasperneite J. Application specific Automation Devices-the App-based approach. In: Jumar U, Jasperneite J, editors. 4th Annual Colloquium Communication in Automation. KommA '13; 2013. p. 1–10. Available from: https://www.researchgate.net/publication/257403049_Application_specific_Automation_Devices_-_the_App-based_approach.

[41] Carroll EA, Latulipe C. Triangulating the Personal Creative Experience: Self-report, External Judgments, and Physiology. In: Brooks S, Hawkey K, editors. Proceedings of the Graphics Interface Conference 2012. GI '12. Mississauga, Canada: Canadian Information Processing Society; 2012. p. 53–60. Available from: https://dl.acm.org/doi/10.5555/2305276.2305286.

[42] Guilford JP. Creativity. American Psychologist. 1950;5(9):444–454. Available from: https://doi.org/10.1037/h0063487.

[43] Guilford JP. The Nature of Human Intelligence. New York, USA: McGraw-Hill; 1967.

[44] Torrance EP. The nature of creativity as manifest in its testing. In: Sternberg RJ, editor. The nature of creativity: Contemporary psychological perspectives. New York, USA: Cambridge University Press; 1988. p. 43–75.

[45] Torrance EP. The Torrance Tests of Creative Thinking: Directions manual Figural (Streamlined) forms A & B. Bensenville, USA: Scholastic Testing Service; 1998.

[46] Torrance EP. Torrance tests of creative thinking: Norms technical manual (Research Ed.). Princeton, USA: Personnel Press; 1966.

[47] Goff K, Torrance EP. Abbreviated Torrance Test for Adults: Manual. Bensenville, USA: Scholastic Testing Service; 2002.

[48] Almeida LS, Prieto LP, Ferrando M, Oliveira E, Ferrandiz C. Torrance Test of Creative Thinking: The question of its construct validity. Thinking Skills and Creativity. 2008 Apr;3(1):53–58. Available from: https://doi.org/10.1016/j.tsc.2008.03.003.

[49] Moran III JD, Milgram RM, Sawyers JK, Fu VR. Original thinking in preschool children. Child Development. 1983 Aug;54(4):921–926. Available from: https://doi.org/10.2307/1129896.

[50] Mednick MT, Mednick SA, Mednick EV. Incubation of creative performance and specific associative priming. Journal of Abnormal and Social Psychology. 1964 Jul;69(1):84–88. Available from: https://doi.org/10.1037/h0045994.

[51] Carson SH, Peterson JB, Higgens DM. Reliability, validity, and factor structure of the creative achievement questionnaire. Creativity Research Journal. 2005 Dec;17(1):37–50. Available from: https://doi.org/10.1207/s15326934crj1701_4.

[52] Kaufman JC. Counting the muses: Development of the Kaufman Domains of Creativity Scale (K-DOCS). Psychology of Aesthetics, Creativity, and the Arts. 2012 Nov;6(4):298–308. Available from: https://doi.org/10.1037/a0029751.

[53] Singh YK. Fundamental of Research Methodology and Statistics. New Delhi, India: Newage International (P) Ltd. Publisher; 2006.

[54] McKay KM Alexander S, Kaufman JC. Measuring the muses: Validating the kaufman domains of creativity scale (K-DOCS). Psychology of Aesthetics, Creativity, the Arts. 2017 May;11(2):216–230. Available from: https://doi.org/10.1037/aca0000074.

[55] Hassenzahl M, Burmester M, Koller F. AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualitat. In: Szwillus G, Ziegler J, editors. Mensch & Computer 2003: Interaktion in Bewegung. Wiesbaden, Germany: Vieweg+Teubner Verlag; 2003. p. 187–196. Available from: https://doi.org/10.1007/978-3-322-80058-9_19.

[56] Zaman B. Introducing a Pairwise Comparison Scale for UX Evaluations with Preschoolers. In: Gross T, Gulliksen J, Kotzé P, Oestreicher L, Palanque P, Prates RO, et al., editors. Human-Computer Interaction – INTERACT 2009. INTERACT 2009. Lecture Notes in Computer Science, vol 5727. Berlin, Germany: Springer; 2009. p. 634–637. Available from: https://doi.org/10.1007/978-3-642-03658-3_68.

[57] Zaman B, Vanden Abeele V, De Grooff D. Measuring product liking in preschool children: An evaluation of the Smileyometer and This or That methods. International Journal of Child-Computer Interaction. 2013 May;1(2):61–70. Available from: https://doi.org/10.1016/j.ijcci.2012.12.001.

[58] Read JC. Validating the Fun Toolkit: an instrument for measuring children's opinions of technology. Cognition Technology and Work. 2008 Apr;10(2):119–128. Available from: https://doi.org/10.1007/s10111-007-0069-9.

[59] Read JC, MacFarlane S. Using the Fun Toolkit and Other Survey Methods to Gather Opinions in Child Computer Interaction. In: Raiha KJ, Hoysniemi J, editors. Proceedings of the 2006 Conference on Interaction Design and Children. IDC '06. New York, USA: ACM Press; 2006. p. 81–88. Available from: https://doi.org/10.1145/1139073.1139096.

[60] Read JC, MacFarlane S, Casey C. Endurability, engagement and expectations: Measuring children's fun. In: Bekker MM, Markopoulos P, Kersten-Tsikalkina M, editors. Proceedings of Interaction Design and Children 2002. IDC'02. New York, USA: ACM; 2002. p. 1–23. Available from: https://www.researchgate.net/publication/228870976_Endurability_Engagement_and_Expectations_Measuring_Childrenaposs_Fun.

[61] Vissers J, De Bot L, Zaman B. MemoLine: evaluating long-term UX with children. In: Hourcade JP, editor. Proceedings of the 12th International Conference on Interaction Design and Children. IDC '13. New York, USA: ACM; 2013. p. 285–288. Available from: https://doi.org/10.1145/2485760.2485836.

[62] Savransky S. Engineering of Creativity: Introduction to TRIZ Methodology of Inventive Problem Solving. Boca Raton, USA: CRC Press; 2000.

[63] Besemer SP, Treffinger DJ. Analysis of creative products: Review and synthesis. The Journal of Creative Behavior. 1981 Sep;15(3):158–178. Available from: https://doi.org/10.1002/j.2162-6057.1981.tb00287.x.

[64] Chang Y. 3D-CAD effects on creative design performance of different spatial abilities students. Journal of Computer Assisted Learning. 2014 Oct;30(5):397–407. Available from: https://doi.org/10.1111/jcal.12051.

[65] Gaggioli A, Mazzoni E, Milani L, Riva G. The creative link: Investigating the relationship between social network indices, creative performance and flow in blended teams. Computers in Human Behavior. 2015 Jan;42(2015):157–166. Available from: https://doi.org/10.1016/j.chb.2013.12.003.

[66] Horn D, Salvendy G. Consumer-based assessment of product creativity: A review and reappraisal. Human Factors and Ergonomics in Manufacturing & Service Industries. 2006 Mar;16(2):155–175. Available from: https://doi.org/10.1002/hfm.20047.

[67] Sarkar P, Chakrabarti A. Assessing design creativity. Design Studies. 2011 Jul;32(4):348–383. Available from: https://doi.org/10.1016/j.destud.2011.01.002.

[68] Sarooghi H, Libaers D, Burkemper A. Examining the relationship between creativity and innovation: A meta-analysis of organizational, cultural, and environmental factors. Journal of Business Venturing. 2015 Sep;30(5):714–731. Available from: https://doi.org/10.1016/j.jbusvent.2014.12.003.

[69] Hocevar D. Measurement of Creativity: Review and Critique. Journal of Personality Assessment. 1981 Oct;45(5):450–464. Available from: https://doi.org/10.1207/s15327752jpa4505_1.

[70] Glăveanu VP. Paradigms in the study of creativity: Introducing the perspective of cultural psychology. New Ideas in Psychology. 2010 Apr;28(1):79–93. Available from: https://doi.org/10.1016/j.newideapsych.2009.07.007.

[71] Kerne A, Webb AM, Smith SM, Linder R, Lupfer N, Qu Y, et al. Using Metrics of Curation to Evaluate Information-Based Ideation. ACM Transactions Computer-Human Interactions. 2014 Jun;21(3). Available from: https://doi.org/10.1145/2591677.

[72] Remy C, MacDonald Vermeulen L, Frich J, Biskjaer MM, Dalsgaard P. Evaluating Creativity Support Tools in HCI Research. In: Wakkary R, Andersen K, editors. Proceedings of the 2020 ACM Designing Interactive Systems Conference. DIS '20. New York, USA: ACM; 2020. p. 457–476. Available from: https://doi.org/10.1145/3357236.3395474.

[73] Frich J, MacDonald Vermeulen L, Remy C, Biskjaer MM, Dalsgaard P. Mapping the Landscape of Creativity Support Tools in HCI. In: Brewster S, Fitzpatrick G, editors. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. CHI '19. New York, USA: ACM; 2019. p. 1–18 (Paper 389). Available from: https://doi.org/10.1145/3290605.3300619.

[74] Lamb C, Brown DG, Clarke CLA. Evaluating Computational Creativity: An Interdisciplinary Tutorial. ACM Computing Surveys. 2018 Feb;51(2). Available from: https://doi.org/10.1145/3167476.

[75] Giannakos MN, Jaccheri L. What motivates children to become creators of digital enriched artifacts? In: Do EYL, Dow S, Ox J, Smith SS, editors. Proceedings of the 9th ACM Conference on Creativity & Cognition. C&C '13. New York, USA: ACM; 2013. p. 104–113. Available from: https://doi.org/10.1145/2466627.2466634.

[76] Guinard JX. Sensory and consumer testing with children. Trends in Food Science & Technology. 2000 Aug;11(8):273–283. Available from: https://doi.org/10.1016/S0924-2244(01)00015-2.

[77] Khaddaj S, Horgan G. The Evaluation of Software Quality Factors in Very Large Information Systems. Electronic Journal of Information Systems Evaluation. 2004 Jan;7(1):43–48. Available from: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.64.1026.

[78] Fenton NE, Bieman JM. Software Metrics: A Rigorous and Practical Approach. 3rd ed. Chapman Hall/CRC Innovations in Software Engineering and Software Development Series. Abingdon, USA: Taylor Francis; 2014.

[79] Goggin G. Ubiquitous apps: politics of openness in global mobile cultures. Digital Creativity. 2011 Sep;22(3):148–159. Available from: https://doi.org/10.1080/14626268.2011.603733.

[80] Dybå T, Dingsøyr T. Empirical studies of agile software development: A systematic review. Information and Software Technology. 2008 Aug;50(9–10):833–859. Available from: https://doi.org/10.1016/j.infsof.2008.01.006.

[81] Cruzes DS, Dybå T. Recommended Steps for Thematic Synthesis in Software Engineering. In: Werner B, editor. Proceedings of the 2011 International Symposium on Empirical Software Engineering and Measurement. ESEM '11. Piscataway, USA: IEEE; 2011. p. 275–284. Available from: https://doi.org/10.1109/ESEM.2011.36.

[82] Sim G, Horton M. Investigating Children's Opinions of Games: Fun Toolkit vs. This or That. In: Proceedings of the 11th International Conference on Interaction Design and Children. IDC '12. New York, USA: ACM; 2012. p. 70–77. Available from: https://doi.org/10.1145/2307096.2307105.

[83] Zhang C, Yao C, Liu J, Zhou Z, Zhang W, Liu L, et al. StoryDrawer: A Co-Creative Agent Supporting Children's Storytelling through Collaborative Drawing. In: Kitamura Y, Quigley A, Isbister K, Igarashi T, editors. Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems. New York, USA: ACM; 2021. Available from: https://doi.org/10.1145/3411763.3451785.

[84] Yarosh S, Thompson S, Watson K, Chase A, Senthilkumar A, Yuan Y, et al. Children Asking Questions: Speech Interface Reformulations and Personification Preferences. In: Proceedings of the 17th ACM Conference on Interaction Design and Children. IDC '18. New York, USA: ACM; 2018. p. 300–312. Available from: https://doi.org/10.1145/3202185.3202207.

[85] Schell J. The Art of Game Design: A Book of Lenses. Boca Raton, USA: CRC Press; 2014.

[86] Basili VR, Weiss DM. A Methodology for Collecting Valid Software Engineering Data. IEEE Transactions on Software Engineering. 1984 Nov;SE-10(6):728–738. Available from: https://doi.org/10.1109/TSE.1984.5010301.

[87] Riemenschneider CK, Hardgrave BC, Davis FD. Explaining Software Developer Acceptance of Methodologies: A Comparison of Five Theoretical Models. IEEE Transactions on Software Engineering. 2002 Jan;28(12):1135–1145. Available from: https://doi.org/10.1109/TSE.2002.1158287.

[88] Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly. 1989 Sep;13(3):319–340. Available from: https://pdfs.semanticscholar.org/ea34/9162d97873d4493502e205968ffccb23fcf2.pdf?_ga=2.13065591.1119970440.1614277031-209301614.1583827202.

[89] Venkatesh V, Davis F. A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. Management Science. 2000 Feb;46(2):186–204. Available from: https://doi.org/10.1287/mnsc.46.2.186.11926.

[90] Moore GC, Benbasat I. Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation. Information Systems Research. 1991 Sep;2(3):192–222. Available from: https://doi.org/10.1287/isre.2.3.192.

[91] Ajzen I. From Intentions to Actions: A Theory of Planned Behavior. In: Kuhl J, Beckmann J, editors. Action Control. Berlin, Germany: Springer; 1985. p. 11–39. Available from: https://doi.org/10.1007/978-3-642-69746-3_2.

[92] Thompson RL, Higgins CA, Howell JM. Personal Computing: Toward a Conceptual Model of Utilization. MIS Quarterly. 1991 Mar;15(1):125–143. Available from: https://doi.org/10.2307/249443.

[93] Kusumo DS, Sabariah MK, Wiharja KRS. A Goal Question Metric (GQM) Approach for evaluating interaction design patterns in drawing games for preschool children. Computers in Human Behavior. 2017 Jun;10(2):96–101. Available from: https://doi.org/10.21609/jiki.v10i2.459.

[94] Hussain ST, Lei S, Akram T, Haider MJ, Hussain SH, Ali M. Kurt Lewin's change model: A critical review of the role of leadership and employee involvement in organizational change. Journal of Innovation Knowledge. 2018 Dec;3(3):123–127. Available from: https://www.sciencedirect.com/science/article/pii/S2444569X16300087.

[95] Phutela D. The Importance of Non-Verbal Communication. IUP Journal of Soft Skills. 2015 Dec;9(4):43–49. Available from: https://ssrn.com/abstract=2715432.

Appendices

Appendix A

NSD Application

NSD NORSK SENTER FOR FORSKNINGSDATA

Meldeskjema 731873

Sist oppdatert

28.10.2020

Hvilke personopplysninger skal du behandle?

- Navn (også ved signatur/samtykke)
- Lydopptak av personer
- Bakgrunnsopplysninger som vil kunne identifisere en person

Type opplysninger

Du har svart ja til at du skal behandle bakgrunnsopplysninger, beskriv hvilke

Yrke, utdanning og ansettelsesvarighet.

Skal du behandle særlige kategorier personopplysninger eller personopplysninger om straffedommer eller lovovertredelser?

Nei

Prosjektinformasjon

Prosjekttittel

ChildrenByDesign-Creativity-MasterThesis

Prosjektbeskrivelse

Prosjektet er en masteroppgave på studiet i informatikk ved Norges teknisk-naturvitenskapelige universitet (NTNU). Det er en del av NTNU-prosjektet ChildrenByDesign som handler om hvordan man skal designe programvare for barn. Masteroppgaven fokuserer på kreative programvarer og samarbeider med et IT-selskap som lager programvarer for barn. Målet med prosjektet er å finne en metode som kan brukes i designprosessen av en kreativ programvare for barn for å evaluere produktet underveis i prosessen på en mest mulig optimal måte.

Begrunn behovet for å behandle personopplysningene

For å finne ut hvordan man skal designe kreative programvarer for barn er det et behov for å behandle personopplysninger. Det skal gjennomføres et intervju med ansatte i IT-selskapet for å finne ut hvordan deres utviklingsprosess er i dag. Enkelte bakgrunnsopplysninger som yrke, utdanning og ansattelsesvarighet er relevant informasjon å innhente for å se om disse faktorene påvirker svarene om utviklingsprosessen. Det skal tas lydopptak av samtalen for å sørge for at intervjuene blir transkibert og analysert riktig i ettertid.

Ekstern finansiering

Type prosjekt

Studentprosjekt, masterstudium

Kontaktinformasjon, student

Marte Hoff Hagen, marte.h.hagen@ntnu.no, tlf: 4741633507

Behandlingsansvar

Behandlingsansvarlig institusjon

Norges teknisk-naturvitenskapelige universitet / Fakultet for informasjonsteknologi og elektroteknikk (IE) / Institutt for datateknologi og informatikk

Prosjektansvarlig (vitenskapelig ansatt/veileder eller stipendiat)

Letizia Jaccheri, letizia.jaccheri@ntnu.no, tlf: 91897028

Skal behandlingsansvaret deles med andre institusjoner (felles behandlingsansvarlige)?

Nei

Utvalg 1

Beskriv utvalget

Ansatte i IT-selskapet.

Rekruttering eller trekking av utvalget

Alle ansatte (10 stk) i IT-selskapet vil bli spurt om de ønsker å delta.

Alder

21 - 67

Inngår det voksne (18 år +) i utvalget som ikke kan samtykke selv?

Nei

Personopplysninger for utvalg 1

- Navn (også ved signatur/samtykke)
- Lydopptak av personer
- Bakgrunnsopplysninger som vil kunne identifisere en person

Hvordan samler du inn data fra utvalg 1?

Personlig intervju

Grunnlag for å behandle alminnelige kategorier av personopplysninger

Samtykke (art. 6 nr. 1 bokstav a)

Informasjon for utvalg 1

Informerer du utvalget om behandlingen av opplysningene?

Ja

Hvordan?

Skriftlig informasjon (papir eller elektronisk)

Tredjepersoner

Skal du behandle personopplysninger om tredjepersoner?

Nei

Dokumentasjon

Hvordan dokumenteres samtykkene?

- Elektronisk (e-post, e-skjema, digital signatur)

Hvordan kan samtykket trekkes tilbake?

Kontakte marte.h.hagen@ntnu.no.

Hvordan kan de registrerte få innsyn, rettet eller slettet opplysninger om seg selv?

Kontakte marte.h.hagen@ntnu.no.

Totalt antall registrerte i prosjektet

1-99

Tillatelser

Skal du innhente følgende godkjenninger eller tillatelser for prosjektet?

Behandling

Hvor behandles opplysningene?

- Maskinvare tilhørende behandlingsansvarlig institusjon
- Ekstern tjeneste eller nettverk (databehandler)

Hvem behandler/har tilgang til opplysningene?

- Student (studentprosjekt)
- Interne medarbeidere
- Databehandler

Hvilken databehandler har tilgang til opplysningene?

Microsoft OneDrive brukes til lagring av den innsamlede dataen. NTNU har en databehandlingsavtale med Microsoft, og alle tjenestene er beskyttet med passord. Studenten som gjennomfører og transkriber intervjuene blir eneste med tilgang til lydopptakene.

Tilgjengeliggjøres opplysningene utenfor EU/EØS til en tredjestat eller internasjonal organisasjon?

Nei

Sikkerhet

Oppbevares personopplysningene atskilt fra øvrige data (koblingsnøkkel)?

Ja

Hvilke tekniske og fysiske tiltak sikrer personopplysningene?

- Opplysningene anonymiseres fortløpende
- Adgangsbegrensning
- Endringslogg

Varighet

Prosjektperiode

15.11.2020 - 30.06.2021

Skal data med personopplysninger oppbevares utover prosjektperioden?

Nei, data vil bli oppbevart uten personopplysninger (anonymisering)

Hvilke anonymiseringstiltak vil bli foretatt?

- Personidentifiserbare opplysninger fjernes, omskrives eller grovkategoriseres
- Lyd- eller bildeopptak slettes
- Koblingsnøkkelen slettes

Vil de registrerte kunne identifiseres (direkte eller indirekte) i oppgave/avhandling/øvrige publikasjoner fra prosjektet?

Nei

Tilleggsopplysninger

Appendix B

Consent Form

Vil du delta i forskningsprosjektet ”ChildrenByDesign-Creativity-MasterThesis”?

Dette er et spørsmål til deg om å delta i et forskningsprosjekt hvor formålet er å finne ut hvordan man bør designe kreative programvareprodukter for barn, i forhold til å evaluere kreativitetsstøtten til produktet underveis i utviklingsprosessen. I dette skrivet får du informasjon om målene for prosjektet og hva deltakelse vil innebære for deg.

Formål

Prosjektet er en masteroppgave på studiet i informatikk ved Norges teknisk-naturvitenskapelige universitet (NTNU). Oppgaven fokuserer på programvare for barn som oppmuntrer til kreativitet, ettersom dette er en fordelaktig egenskap å utvikle for barn da det blant annet forbedrer deres ferdigheter innen problemløsing, programmering og språk. Formålet med prosjektet er å finne ut hvordan man bør designe slike programvareprodukter. Dette får praktisk betydning for IT-firmaer som lager slike produkter og i dag mangler effektive metoder til å evaluere kreativitetstøtten underveis i utviklingsprosessen.

Hvem er ansvarlig for forskningsprosjektet?

Prosjektet er en del av en masteroppgave ved Institutt for datateknologi og informatikk (IDI) ved NTNU. Prosjektet utføres av masterstudent Marte Hoff Hagen og veiledes av Letizia Jaccheri og Daniela Soares Cruzes.

Hvorfor får du spørsmål om å delta?

Du får spørsmål om å delta på grunn av din rolle som ansatt i IT-firmaet Ablemagic. Vi har fått kontaktinformasjonen din fra Ablemagic. Din deltakelse vil bidra til å gi innsikt i hvordan Ablemagic sin utviklingsprosess av kreative programvareprodukter for barn er i dag, som videre skal brukes til å finne ut hva som skal endres i denne prosessen for at det skal bli bedre evaluering av kreativitetsstøtten til produktene.

Hva innebærer det for deg å delta?

Om du velger å delta, vil du bli med på et semistrukturert intervju for å innhente informasjon om Ablemagic sin utviklingsprosess av kreative programvareprodukter for barn. Dette intervjuet vil ta deg ca. 30 minutter å gjennomføre. Det blir tatt lydopptak av intervjuet som skal brukes til å transkribere og analysere den innhentede dataen i ettertid.

Det er frivillig å delta

Det er frivillig å delta i prosjektet. Hvis du velger å delta, kan du når som helst trekke samtykket tilbake uten å oppgi noen grunn. Alle dine personopplysninger vil da bli slettet. Det vil ikke ha noen negative konsekvenser for deg hvis du ikke vil delta eller senere velger å trekke deg.

Ditt personvern – hvordan oppbevares og brukes dine opplysninger

Det vil bare brukes opplysningene om deg til formålene det er fortalt om i dette skrivet. Opplysningene behandles konfidensielt og i samsvar med personvernregelverket. Det vil kun være masterstudenten og veilederne, som nevnt tidligere, som har tilgang til opplysningene.

Navn og kontaktopplysninger om deg vil bli erstattet med en kode som lagres på en egen navneliste adskilt fra øvrige data. Datamaterialet blir lagret på Microsoft OneDrive. NTNU har en databehandlingsavtale med Microsoft, og alle tjenestene er beskyttet med passord. Studenten som gjennomfører og transkriber intervjuene blir eneste med tilgang til lydopptakene fra intervjuene.

Deltakere vil ikke kunne gjenkjennes i publikasjon, med mindre samtykke for dette gis i samtykkeerklæringen (nederst i dette dokumentet).

Hva skjer med opplysningene dine når forskningsprosjektet avsluttes?

Opplysningene anonymiseres når prosjektet avsluttes/masteroppgaven er godkjent, noe som etter planen er ca. 30.06.21, og etter dette vil datamaterialet bli anonymisert. Alle lydopptak vil da bli slettet.

Dine rettigheter

Så lenge du kan identifiseres i datamaterialet, har du rett til:

- innsyn i hvilke personopplysninger som er registrert om deg, og å få utlevert en kopi av opplysningene,
- å få rettet personopplysninger om deg,
- å få slettet personopplysninger om deg, og
- å sende klage til Datatilsynet om behandlingen av dine personopplysninger.

Hva gir oss rett til å behandle personopplysninger om deg?

Vi behandler opplysninger om deg basert på ditt samtykke.

På oppdrag fra NTNU har NSD – Norsk senter for forskningsdata AS vurdert at behandlingen av personopplysninger i dette prosjektet er i samsvar med personvernregelverket.

Hvor kan jeg finne ut mer?

Hvis du har spørsmål til studien, eller ønsker å benytte deg av dine rettigheter, ta kontakt med:

- NTNU - Norges teknisk-naturvitenskapelige universitet ved Letizia Jaccheri på epost (letizia.jaccheri@ntnu.no) eller telefon: 918 97 028
- NTNU - Norges teknisk-naturvitenskapelige universitet ved masterstudent Marte Hoff Hagen på epost (marte.h.hagen@ntnu.no)
- Vårt personvernombud: Thomas Helgesen på thomas.helgesen@ntnu.no

Hvis du har spørsmål knyttet til NSD sin vurdering av prosjektet, kan du ta kontakt med:

- NSD – Norsk senter for forskningsdata AS på epost (personverntjenester@nsd.no) eller på telefon: 55 58 21 17.

Med vennlig hilsen

Letizia Jaccheri	Marte Hoff Hagen
Prosjektansvarlig	Prosjektdeltaker
(Forsker/veileder) .	(Masterstudent)

Samtykkeerklæring

Jeg har mottatt og forstått informasjon om prosjektet "ChildrenByDesign-Creativity-MasterThesis", og har fått anledning til å stille spørsmål. Jeg samtykker til å delta i semistrukturert intervju der det blir tatt lydopptak, og at mine opplysninger behandles frem til prosjektet er avsluttet og masteroppgaven er godkjent, ca. 30.06.21.

(Signert av prosjektdeltaker, dato)

Appendix C

Interview Guides

C.1 Software Development Process

C.1.1 The Software Development Process

- How would you describe the software development process of the CSTs in Ablemagic?
- Is the software development process of Ablemagic different from what you are used to from other jobs and/or education? In that case, how?

C.1.2 The Creativity Evaluation Methods for Children

- How is your CEM for children?
- Do you have a clear goal for the CEM?
- Do you use any theoretical grounding?
- What is the frequency of the CEMs?
- When in the software development process is the CEM?
- Are there any elements in CSTs you can evaluate without involving children?
- How many children are included in a CEM?
- How do you recruit children?
- Do you think about the gender and the age of the children?
- How do you communicate with the children in conjunction with the CEM?
- What data collection method are you using?
- Which of the ten evaluation factors — *valuable*, *novelty*, *enjoyment*, *collaboration*, *flow*, *user feeling*, *flexibility*, *immersion*, *expressiveness*, and *interaction* — do you think is most important for Ablemagic?
- Do you evaluate usability?

C.1.3 Other Comments

- Do you have more comments, or is it something else you would like to talk about?

C.2 Evaluation of the tailored Creativity Evaluation Method

C.2.1 The User Test

- How do you feel regarding the user test?
- Did you notice something special before the test?
- Did you notice something special after the test?
- How was it to use the Smileyometer?
- What would you do differently next time?
- What is the further plan based on the results from the user test?
- Do you consider using more interviews or questionnaires next time?

C.2.2 Evaluation of the tailored Creativity Evaluation Method

Usefulness rate

- Was it useful? Why/why not?
- Did it improve the process? Can you explain?
- Do you want to use it in the future? Do you have any comments on your answers?

Intuitiveness rate

- Was it easy to use? What do you think makes it easy/not easy to use? How was your process of trying to understand it?
- Was it clear and understandable? If no: What was not clear and/or understandable?
- Was the use of it frustrating at some point? If yes: Which parts?

Compatibility rate

- Is it compatible with the way you develop software? Why/why not?
- Is it compatible with the way your company develops software? Can you explain?

Other Comments

- Do you have any other comments? Would you like to talk about other advantages or challenges of using this method? Or what you think should be improved?

Appendix D

Journal Article

Evaluating Digital Creativity Support Tools for Children: A Systematic Literature Review

MARTE HOFF HAGEN, Norwegian University of Science and Technology (NTNU), Norway
DANIELA SOARES CRUZES, Norwegian University of Science and Technology (NTNU), Norway
LETIZIA JACCHERI, Norwegian University of Science and Technology (NTNU), Norway

Creativity is beneficial for children because it improves their skills and development. Digital Creativity Support Tools are software apps that are supposed to improve creativity. This systematic literature review investigated the current Creativity Evaluation Methods of children's digital Creativity Support Tools by identifying 81 peer-reviewed articles from the last 10 years. The review showed that the Creativity Evaluation Methods differ widely, and the area lacks a standard Creativity Evaluation Method. This research contributes to practitioners by providing recommendations regarding Creativity Evaluation Methods of digital Creativity Support Tools for children. It contributes to researchers by covering areas needing further investigation.

CCS Concepts: • **Human-centered computing → HCI design and evaluation methods**.

Additional Key Words and Phrases: creativity, creativity support tool, creativity evaluation method, children, systematic literature review

ACM Reference Format:
Marte Hoff Hagen, Daniela Soares Cruzes, and Letizia Jaccheri. 2021. Evaluating Digital Creativity Support Tools for Children: A Systematic Literature Review. 1, 1 (June 2021), 33 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

According to Article 1 in the United Nations Convention on the Rights of the Child (CRC), a child is defined as a "human being below the age of eighteen years unless under the law applicable to the child, majority is attained earlier" [67]. This international legal framework is ratified by 192 of 194 member countries and comprises 54 different articles that protect children's rights and asserts that all children are individual right-holders.

Child-computer interaction (CCI) aims to actively involve children in designing technologies that will contribute to their intellectual, social, and creative growth [73]. This is beneficial because children contribute other ideas than adult designers [48]. However, it can be challenging to collaborate with children in design processes [33]. During previous years, the interest in CCI has grown, which encompasses the entire age specter of children [51]. The research field is continuously growing and evolving by linking various approaches, techniques, methodologies, end-user groups, technologies,

Authors' addresses: Marte Hoff Hagen, Norwegian University of Science and Technology (NTNU), Trondheim, Norway, marte.h.hagen@ntnu.no; Daniela Soares Cruzes, Norwegian University of Science and Technology (NTNU), Trondheim, Norway, daniela.s.cruzes@ntnu.no; Letizia Jaccheri, Norwegian University of Science and Technology (NTNU), Trondheim, Norway, letizia.jaccheri@ntnu.no.

XXXX-XXXX/2021/6-ART $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

and scientific disciplines [20].

According to Ritter and Rietzschel [41], the English word *creativity* stems from the Latin word *creō*, which means "to create, to make" [41, p. 97]. The term has evolved from psychology theory and is defined in many different ways [37]. Oxford Dictionary defines it as "the use of skill and imagination to produce something new or to produce art" [45]. This is similar to the common definition of creativity as the ability to generate ideas or problem solutions that are both *novel* (i.e., perceived unique or original) and *valuable* (i.e., perceived useful or functional) [2, 54, 62]. Sarkar and Chakrabarti analyzed over 160 definitions to propose a common definition of creativity: "Creativity occurs through a process by which an agent uses its ability to generate ideas, solutions or products that are novel and valuable" [55, p. 11].

By examining the meaning of the word creativity, Jordanous and Keller [35] expanded the two components *novel* and *valuable* to 14 components. The 12 remaining components were *active involvement*, *immersion*, *generation of results*, *dealing with uncertainty*, *domain-specific competence*, *general intellect*, *freedom*, *progression*, *collaboration*, *spontaneity*, *decision-making*, and *variety*. According to Weisberg [72], *valuable* should not be a criterion for creativity because what society thinks is *valuable* changes over generations. Instead, he suggested defining creativity as "intentional novelty" [72, p. 119]. Another approach to creativity by Vygotsky [70] is the process of constructing these novel ideas or behaviors by revising old memories.

Creativity is considered one of the best cognitive abilities for humans [4], and it is linked to innovation and growth in society [3]. It is a rather beneficial quality for children because it improves their problem-solving, computer science, programming, storytelling, play, and language skills and contributes to individual personal development [13, 19, 30, 32]. It is one of the seven core dimensions of the digital skills for students of the 21st century and central in CRC's Article 31 regarding the right to play [67, 68].

Certain software is supposed to enhance children's creativity [46]. Such software is called Creativity Support Tools (CSTs), which Cherry and Latulipe define as "any tool that can be used by people in the open-ended creation of new artifacts" [11, p. 2]. In computer science, CSTs are often apps used in the process of completing an artifact or used to create digital artifacts [11]. According to Tack and Jasperneite [63], an app is "a small piece of software, that provides a specific functionality and is able to use device features. Furthermore, the app is totally integrated into a distribution system, that provides benefits for device suppliers, developers and customers" [63, p. 2]. The focus on CSTs in this paper is on apps. There are a large number of such apps; for example, the visual programming language Scratch (figure 1), the photo editor program Adobe Photoshop, and the collaborative writing program Google Docs.

It is difficult to evaluate CSTs because there are no single metrics to quantify them [60]. According to Resnick [53], creativity is strongly dependent on human factors, which gives the Creativity Evaluation Methods (CEMs) of CSTs unusual issues and challenges for the human-computer interaction (HCI) field. In addition, children's technology is often intended for entertaining rather than supporting creativity [9].

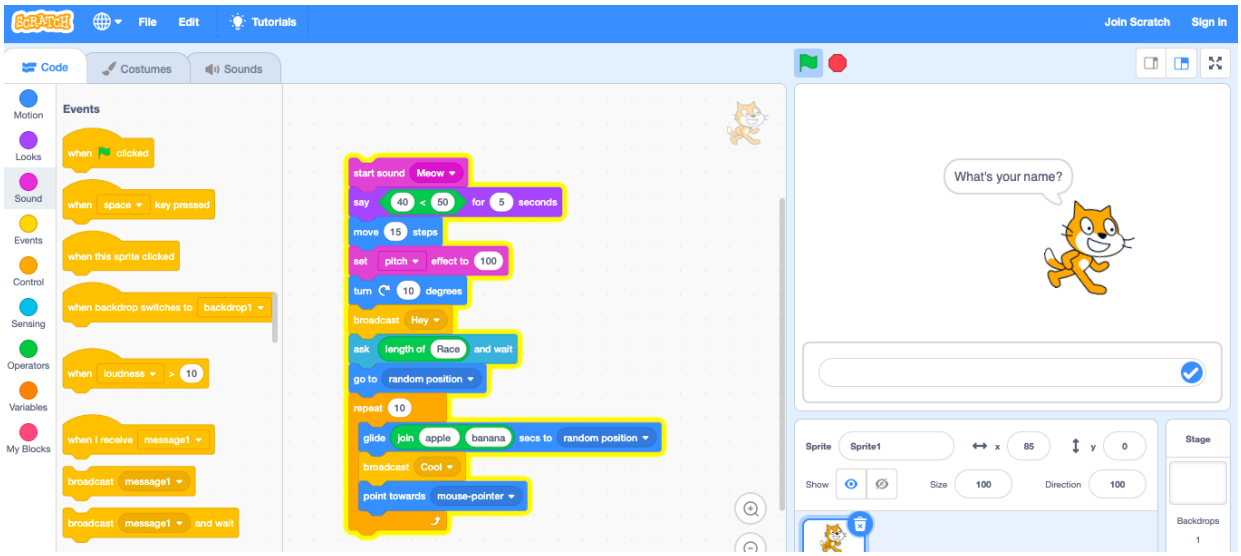


Fig. 1. Screenshot of the digital CST Scratch.

This research aims to explore the current procedures of CEMs of digital CSTs for children. A systematic literature review (SLR) was used since it is a trustworthy, rigorous, and auditable methodology to identify, evaluate, interpret, and summarize all available relevant research in the focus area [39]. With the SLR, this paper aims to answer the following research questions (RQs):

- **RQ1:** What are the current CEMs of digital CSTs for children?
  - **RQ1.1:** Which factors have been considered in CEMs?
  - **RQ1.2:** In which context are CEMs performed?
  - **RQ1.3:** How does age of the target group impact CEMs?
  - **RQ1.4:** How effective are CEMs?

The remainder of this paper is structured in the following manner. First, section 2 presents the background regarding CEMs and CSTs. Section 3 thoroughly describes the SLR, while section 4 presents the SLR findings. Section 5 discusses these findings, and section 6 concludes by answering the RQs and presents suggestions for future research.

2 BACKGROUND

This section begins with a presentation of different approaches to CEMs. Then, it discusses CEMs used for CSTs.

2.1 Creativity Evaluation Methods

According to Carroll and Latulipe [7], the three basic approaches to CEMs are *self-reporting* by users, *psychophysical measurement* using biometrics, and *external judges*. Self-reporting is a subjective CEM since the user itself does it. External judges are a more objective CEM because it utilizes externals. However, the most objective CEMs is done by observers in the form of the psychophysical measurement with biometrics like the user's behavior with the product.

2.1.1 Psychophysical measurement. In 1950, Guilford initiated modern creativity research by developing the psychophysical measurement Alternative Uses Task (AUT) based on cognitive psychology [24]. The AUT measures creativity in the form of the divergent thinking ability and reveals how many alternative uses for an object a participant managed to mention within a specific amount of time [25].

Torrance built on Guilford's thoughts that creativity depended on divergent thinking. He proposed an analysis including the following four factors of problem responses by humans: the number of relevant answers (*fluency*), the number of different categories in the answers (*flexibility*), the statistical rarity of the answers (*originality*), and the number of details in the answers (*elaboration*) [65]. These four factors could be measured by the psychophysical measurement Torrance Tests of Creative Thinking (TTCT), which is a standardized measure of children's creativity [66]. The TTCT consists of a verbal test with writing and a figural test with drawings, each with different subtests [64]. Based on the principle of TTCT, the abbreviated Torrance test for adults required a shorter test time [22]. However, TTCT has been criticized for being uncertain since the factor *elaboration* contributes substantially to the score despite being less important [1].

A similar psychophysical measurement is the Multidimensional Stimulus Fluency Measure (MSFM) creativity test, which evaluates *fluency* and *originality* [44]. It can be applied to young children as it requires verbal responses and uses tactile and visual stimuli [44]. On the other hand, the psychophysical measurement Remote Associates Test (RAT) is a convergent creativity test in which the participants are presented with three words and have to guess the fourth word that these three words are semantically related to [43].

2.1.2 Self-reporting. The CEM Creativity Achievement Questionnaire (CAQ) is a self-reporting questionnaire of creativity scores across 10 domains (writing, culinary arts, architecture, visual arts, dance, music, theater, scientific discovery, inventions, and humor) [8]. This is similar to Kaufman Domains of Creativity Scale (K-DOCS), a five-point Likert scale with 50 subdivisions in the following domains: everyday, academic, mechanical or science, artistic, and performance [36]. The Likert scale is a self-reporting questionnaire with a set of statements where the participants are asked to evaluate the level of agreement from strongly disagree to strongly agree, normally on a five-point scale [61]. K-DOCS has been proved to be a valid and reliable CEM for evaluating domain-specific creativity [42].

An acceptance self-reporting CEM for products is AttrakDiff, which evaluates both the pragmatic (*functionality*), the hedonic perception (*emotional experience*), and attractiveness (*aesthetic*) on a seven-point scale [28]. It differs from the Likert since the user does not evaluate a statement but how the product corresponds between two opposites (e.g., "complicated-simple").

There exist different self-reporting CEMs designed for children. One self-reporting preference CEM is This or That, where a child compares two different options and selects which product (s)he mostly prefers [74, 75]. According to Guinard [26], such paired-preference testing is the only valid method for two to three-year-old children. It can be considered preference ranking for older children when more than two products and acceptance scales are between three- to nine-point [26]. Another broadly used CEM for children is the self-reporting acceptance CEM Smileyometer [47]. It is a five-point Likert scale where the response items are replaced by smiley faces representing the user's perceived *enjoyment* level [49]. To evaluate the level of *fun*, Fun-Sorter is a CEM used

to rank statements for a series of connected activities to establish which is considered the most and the least *fun* [50]. The self-reporting instrument MemoLine is also child-friendly. It measures long-term user experience by asking the child to color a timeline according to their experience with the product changed in a predefined timespan [69].

2.1.3 External judges. The systematic human-oriented knowledge-based method of inventive problem-solving called theory of inventive problem solving (TIPS) uses external judges to compare the process with this method [58]. To manage and order the generated ideas in this method, it is possible to use Buzan's approach MindMap, which is a figure that illustrates the ideas to the main concept with different themes radiating from the central concept as labeled branches [6].

The Consensual Assessment Technique (CAT) uses external expert judges to evaluate products based on their subjective perception of creativity [2]. As reported by Hocevar [29], an issue with external judges is their inability to distinguish between creativity, aesthetics, and technical skills. According to Glăveanu [21], this judgment depends on the press since the same artifact could be assessed differently by members of different communities.

The evaluation rubrics in the two CEMs Creativity Product Analysis Matrix and Creative Product Semantic Scale (CPSS) use external judges to assess three aspects. These are *novelty* (i.e., originality and intuitiveness), *resolution* (i.e., usefulness, logic, and relevance), and *elaboration and synthesis* (i.e., aesthetic and beauty) [5, 10, 17]. Others have simplified these aspects to the two dimensions *novelty* and *valuable* [31, 56, 57]. One method to assess the *novelty* of a product is by comparing the characteristics of the product with that of another [59].

2.1.4 Creativity Evaluation Methods of Creativity Support Tools. There exist several different CEMs of CSTs. An example is Warr and O'Neil [71], who evaluated a CST using psychophysical observation from video footage and post-evaluative self-reporting questionnaires to quantify the types of interactions for boundary objects created by the participants and artifacts provided by the tools in the different steps of the design process. Kerne et al. [38] showed that measuring the levels of creativity components — based on the *fluency*, *flexibility*, *novelty*, *emergence*, *relevance*, *visual presentation*, and *exposition* — using observation of a person's development of new ideas combined with post-evaluative questionnaires can reveal how a CST supports creative engagement.

Cherry and Latulipe [11] developed the Creativity Support Index (CSI), a self-reporting standardized psychophysical measurement designed to evaluate a CST's ability to assist a user engagement in creative work. This is done by measuring six factors of creativity support: *collaboration*, *expressiveness*, *exploration*, *enjoyment*, *immersion*, and *results satisfaction*. The CSI enables comparable and quantifiable results by providing a rating scale section of two agreement statements for each factor and a paired-factor comparison. This was inspired by the standardized survey metric NASA Task Load Index, which was used to quantify the complex phenomenon workload associated with completing a specific task [27]. The psychometric survey in the CSI may be too complicated for children to use. To evaluate CSTs designed for children, it could be valuable to better understand several aspects from their perspective [18].

Remy, Vermeulen, Frich, Biskjaer, and Dalsgaard performed a recent and relevant in-depth literature review of CEMs of CSTs in HCI research [52]. The literature review aimed to improve the CEM strategies for developing CSTs in HCI research by providing an objective description of the

CEMs of the most leading CSTs in HCI research. This paper reused the corpus of their earlier in-depth literature review of CSTs in HCI research [16] by selecting all papers in this sample that evaluated a CST. Frich, Vermeulen, Remy, Biskjaer, and Dalsgaard [16] used the search string *"creativity" OR "creativity support tool"* in the *ACM Digital Library*. The sample size was reduced based on the above-average citations per year [16]. In addition, Remi et al. used the same method for papers after 2016 but reduced the sample size based on the average download count per year [52].

Based on the findings, Remy et al. [52] provided recommendations for future CEMs of CSTs. These recommendations were centered around six major aspects: 1) to clearly define the goal of the CEM since numerous studies are unclear about this goal, 2) to use theoretical grounding in the CEM because this is lacking in today's CEMs, 3) to decide whether to evaluate creativity or usability as many of the CEMs focus on usability instead of creativity, 4) to consider longitudinal in-situ studies since short-term controlled CEMs are prioritized, 5) to recruit domain experts if appropriate because the CEMs lack expert participants, and 6) to contribute to developing a toolbox for CEMs of CSTs as it is not existing [52]. Lamb, Brown, and Clarke [40] highlighted the need for such a toolbox as they referred to use the existing CEM standards of CSTs to evaluate co-creative systems in their interdisciplinary tutorial on evaluating computational creativity. Remy et al. [52] did not investigate the age of the target group in the studies. Thus, there is a need for an SLR focusing on the CEMs of digital CSTs for children.

To summarize, there exists a vast amount of research on creativity and CCI, but there is scarce research in the intersection of evaluating creativity in software for children. This survey of papers investigates this knowledge gap. Thus, the goal of producing this SLR is to produce valuable information regarding the intersection between creativity and CCI.

3 REVIEW METHOD

An SLR was conducted based on SLR's original guidelines for software engineering provided by Kitchenham [39].

3.1 Protocol development

In the initial phase of the SLR, a review protocol was developed to plan the SLR [39]. The protocol contained the rationale for the survey, the research questions, the search strategy, the study selection procedures and criteria, the study quality assessment procedures, the data extraction procedure, the synthesis strategy, and the project timetable. The SLR is divided into five steps, as illustrated in figure 2. The remainder of this section will describe each step in detail.

3.2 Identification of Research

The first step of conducting the SLR was the identification of research [39]. A pilot search was used to find the most optimal search string and databases to identify the current CEMs of digital CSTs for children. This search also found the mentioned literature review by Remy, Vermeulen, Frich, Biskjaer, and Dalsgaard [52]. The corpus of 113 studies from this literature review was added to the collection.

The literature search was conducted in October 2020. To include as many relevant studies as possible, the search strategy included the electronic bibliographic databases *Scopus*, *ACM Digital*

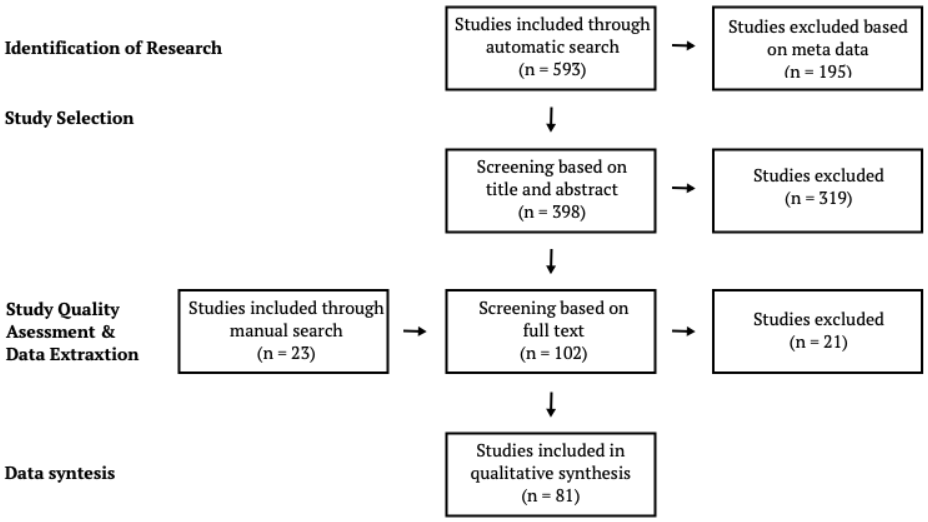


Fig. 2. Study selection process.

Library, and *IEEE Xplore* and journal hand search in the *International Journal of Child-Computer Interaction*. The documenting of the search is presented in table 1.

Table 1. Documenting of the automatic search.

Data Source	Search string	Search date	Hits
Recent and relevant literature review [52]	Based on *"creativity" OR "creativity support tool"* in the *ACM Digital Library* [16]	02.10.20 (pilot search)	113
Scopus	*(TITLE AND ABS)("creativity" AND ("measure" OR "measuring" OR "metrics" OR "evaluate" OR "evaluating") AND ("software" OR "app"))*	09.10.20	259
ACM Digital Library		10.10.20	89
IEEE Xplore		13.10.20	91
International Journal of Child-Computer Interaction	*"creativity" AND ("measure" OR "measuring" OR "metrics" OR "evaluate" OR "evaluating") AND ("software" OR "app")*	16.10.20	41
		Total:	**593**

It was intentionally decided not to include the term *"children"* in the search string, despite the focus on children in RQ1. This was done to explore if the CEMs of CST used on adults could be transferable to children.

This query resulted in 593 hits, and the bibliographic package Endnote X9.3.3[1] was used to manage this large number of references.

3.3 Study Selection

The potentially relevant primary studies from the search were then assessed for their actual relevance with exclusion criteria (C) [39]. After multiple iterations, these criteria were defined in the following manner:

- **C1:** Publishing date before 2010 (Excluded: 157)
- **C2:** Duplicated studies (Excluded: 37)
- **C3:** Non-English language (Excluded: 1)
- **C4:** No CEMs (Excluded: 319)

[1] endnote.com

Due to the focus on current CEMs in RQ1, C1 excluded non-recent studies. The year 2010 was selected as the lower time limit because this year represented the beginning of the big growth of apps since Apple released its Software Development Kit in 2009 [23]. Following Kitchenham [39], if a duplicate of a study was published more recently than the other duplicate(s), the most recent duplicate publication was kept in C2. Duplicate exclusion was done in Endnote. Even though Kitchenham [39] recommends avoiding exclusion based on the language, it was decided to include C3 as an exclusion criterion. The English abstracts in non-English papers did not give as much detail as required for the analysis. C4 was based on the RQs, and the exclusion in this criteria was based on reading titles and abstracts.

This exclusion resulted in a total of 79 primary studies.

3.4 Manual Search

To include additional papers that were not discovered by the automatic search, manual searches were conducted in relevant journals in CCI, HCI, creativity, and software engineering (SE) in November 2020. The documenting of the manual search is presented in table 2.

The manual search resulted in 23 studies, which yielded 102 primary studies in the SLR.

Table 2. Documenting of the manual search.

Data Source	Search string	Search date	Hits	Relevant
International Journal of Child-Computer Interaction	*"creativity" AND PUBYEAR ≥ 2010*	16.11.20	60	2
Interaction with Computers			142	2
ACM Transactions on Computer Human Interaction			74	2
AIS Transaction on Human-Computer Interaction			33	0
Computers in Human Behavior			408	8
Computer-Aided Design and Application			155	1
International Journal of Design Creativity and Innovation	*("software" OR "app") AND PUBYEAR ≥ 2010*	18.11.20	39	0
Journal of Creativity behavior			91	2
Thinking Skills and Creativity			155	6
Digital Creativity			183	0
IEEE Transactions on Software Engineering	*"creativity" AND PUBYEAR ≥ 2010*	20.11.20	2	0
IEEE Software			11	0
Transactions on Software Engineering and Methodology			15	0
Empirical Software Engineering			34	0
Information & Software Technology			61	0
Automated Software Engineering			4	0
Requirements Engineering			29	0
Software Quality Journal			13	0
Journal of Systems and Software			96	0
Software Testing, Verification & Reliability			6	0
Journal of Software: Evolution and Process			57	0
Software: Practice & Experience			29	0
International Journal of Software Engineering and Knowledge Engineering			59	0
			Total:	**23**

3.5 Assessment of Study Quality

Further, it was through multiple iterations made quality assessment that was used to assess each primary study's quality based on full-text reading [39]. Each primary study had to pass three quality assessment (QA) questions regarding relevance, rigor, and credibility, adapted from the checklist to Dybå and Dingsøyr [14]:

- **QA1:** Do the results of the study include a CEM of a digital CST? (Excluded: 6)
- **QA2:** Is the research method appropriate in terms of using evaluation factors? (Excluded: 15)

- **QA3:** Do the presented findings have a validity aspect? (Excluded: 0)

This step of quality assessment with full-text reading of the studies was combined with the next step on data extraction and monitoring to save time. Overall, the QAs excluded 21 studies, which resulted in an inclusion of 81 studies in the qualitative data synthesis.

3.6 Data Extraction

As proposed by Kitchenham [39], it was designed a systematic data extraction to collect all the information required from the primary studies to address the RQs and QAs. This predefined extraction form is presented in table 3.

The data was extracted in the specialist software package for qualitative analysis of textual data MAXQDA Analytics Pro 2020 (Release 20.2.1)[2]. This was done independently by the first authors but discussed with the second and third author to ensure it was done systematically.

Table 3. Data extraction form.
RQ = Research Question, QA = Quality Assessment

Focus area	Data	Mapping to QAs	Mapping to RQs
Demographics	• Publication channel • Publication year		
Factors	• Factors evaluating creativity	QA2	RQ1.1, RQ1.3
Products	• Main functionality • Characteristics that improve creativity • Creativity outcome • Software type • Creativity topic	QA1	RQ1.2
Methods	• Data gathering method • Data analysis • Participants – Sample size – Gender distribution – Age • Approach • Validity of method • Study environment	QA1, QA2	RQ1, RQ1.2, RQ1.3, RQ1.4
Research	• Objective • RQs • Findings • Outcome	QA3	RQ1.4
Creativity	• Definition of creativity • Unclear words describing creativity	QA3	RQ1.4

3.7 Data Synthesis

Data synthesis was used to collect and summarize the results of the included primary studies to answer the RQs [39]. The five recommended steps for thematic synthesis in SE proposed by Cruzes and Dybå [12] were utilized. The first two steps related to extracting data from primary studies and systematic coding of this data based on the RQs are described in section 3.6. The next two steps were to translate the codes into themes, higher-order themes, and sub-themes and create a model of the higher-order themes by exploring the relationships between the themes based on the RQs,

[2] maxqda.com

which was done in the mind mapping software MindManager 13.0.181[3]. The last step was to assess the trustworthiness of the interpretation leading to the synthesis. This was done by looking for counter-evidence.

4 RESULTS

4.1 Demographics

The distribution of the publication year for the studies is presented in table 4. Figure 3 illustrates the publication frequency between 2010 and 2020. It increased from 2013 until 2016, with 12 published studies, and has decreased thereafter. The distribution of the publication channels for the studies is illustrated in table 4. The number of conference papers were more than the number of journal articles. The most frequent conference was the *the SIGCHI Conference on Human Factors in Computing Systems* with 11 publications. Further, *Creativity and Cognition Conference* had eight and *ACM Conference on Computer-Supported Cooperative Work*, *ACM International Conference on Tangible, Embedded & Embodied Interaction*, and *ACM Conference on Designing Interactive Systems* had three each. With regard to the journals, eight studies were published in the journal *Computers in Human Behavior*, five in *Thinking Skills and Creativity* and *International Journal of Child-Computer Interaction* and three in *ACM Transaction on Computer-Human Interaction*. In addition, S20 was published as a chapter in the book entitled *Young Children and Families in the Information Age* and S73 was published as a chapter in *Gesellschaft für Informatik e.V's lecture notes in informatics*.

Table 4. Publication frequency.

Year	Journal	Conference	Book
2020	S21, S48, S71		
2019	S4, S11, S22, S34, S56, S77	S12, S13	
2018	S19, S24, S55, S63	S26, S31, S43, S70	
2017	S28, S46, S58, S68	S2, S3, S76	
2016	S5, S10, S27, S47, S53	S15, S32, S33, S45, S54, S66, S80	
2015	S18, S35, S41, S62	S37, S49, S59, S60, S65, S79	S20
2014	S14, S36, S39	S6, S23, S51	
2013	S61	S25, S67, S75	
2012	S9, S50, S57	S1, S8, S29	S73
2011	S64, S74	S30, S38, S40, S42, S78, S81	
2010	S16	S7, S17, S44, S52, S69, S72	
%	45	53	2

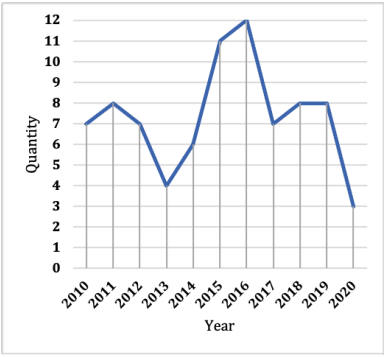


Fig. 3. Publication frequency.

4.2 Evaluation factors

To answer the main RQ1 regarding current CEMs of digital CSTs for children, this section addresses the RQ1.1 regarding the factors used evaluating digital CSTs for both adults and children. A total of 53 different factors have been considered in the CEMs of the included studies. These were categorized into higher-order themes. The categorizing yielded 11 different categories, including the 10 most frequently occurring categories that covered 90.3% of the factors and one *Other* category with the remaining factors. The distribution of these categories in the included studies is illustrated in table 5. The remainder of this section will investigate each of these categories in detail.

[3] mindmanager.com

Table 5. Overview of the distribution of factors, products, approaches (appr.), and methods.

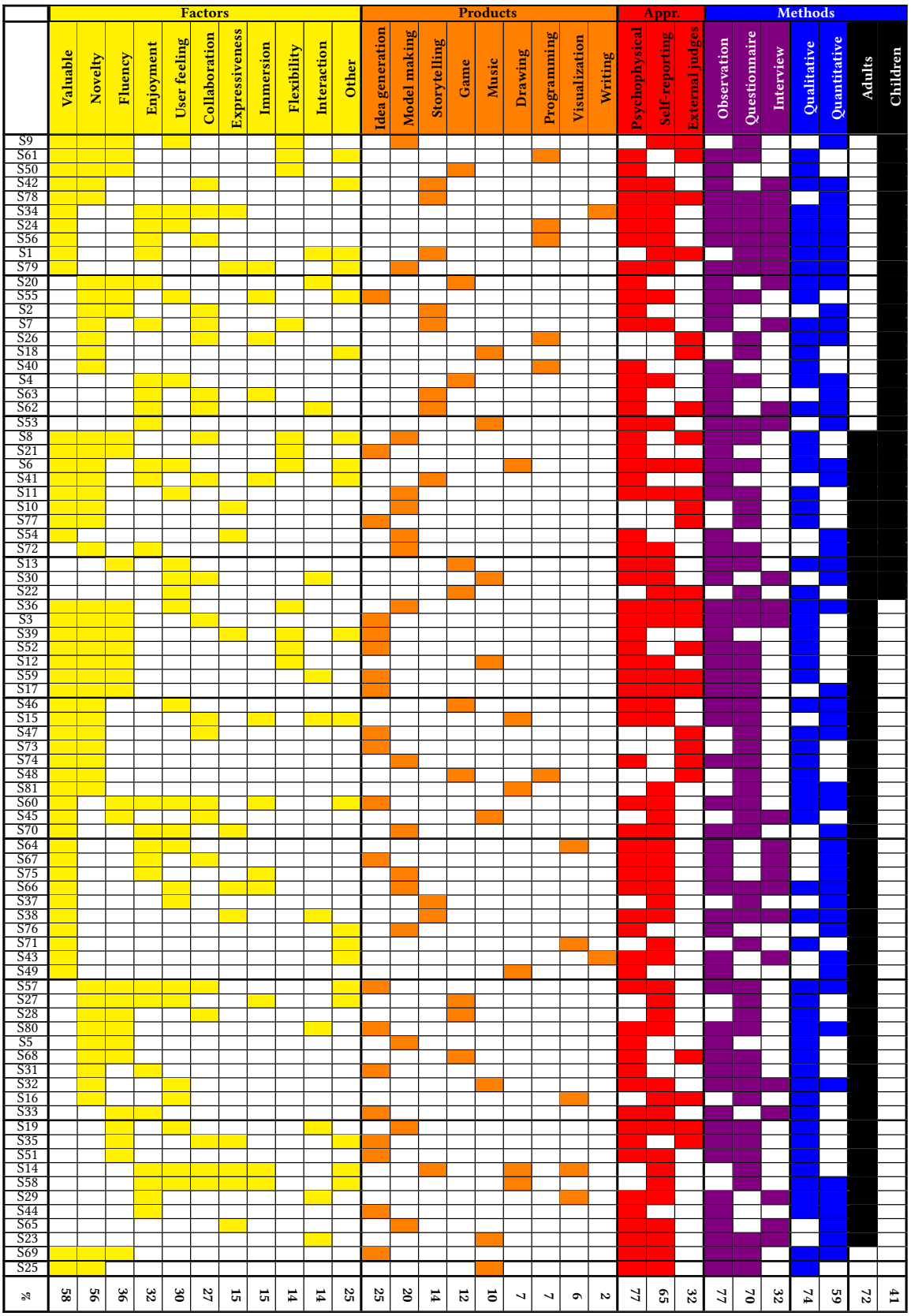
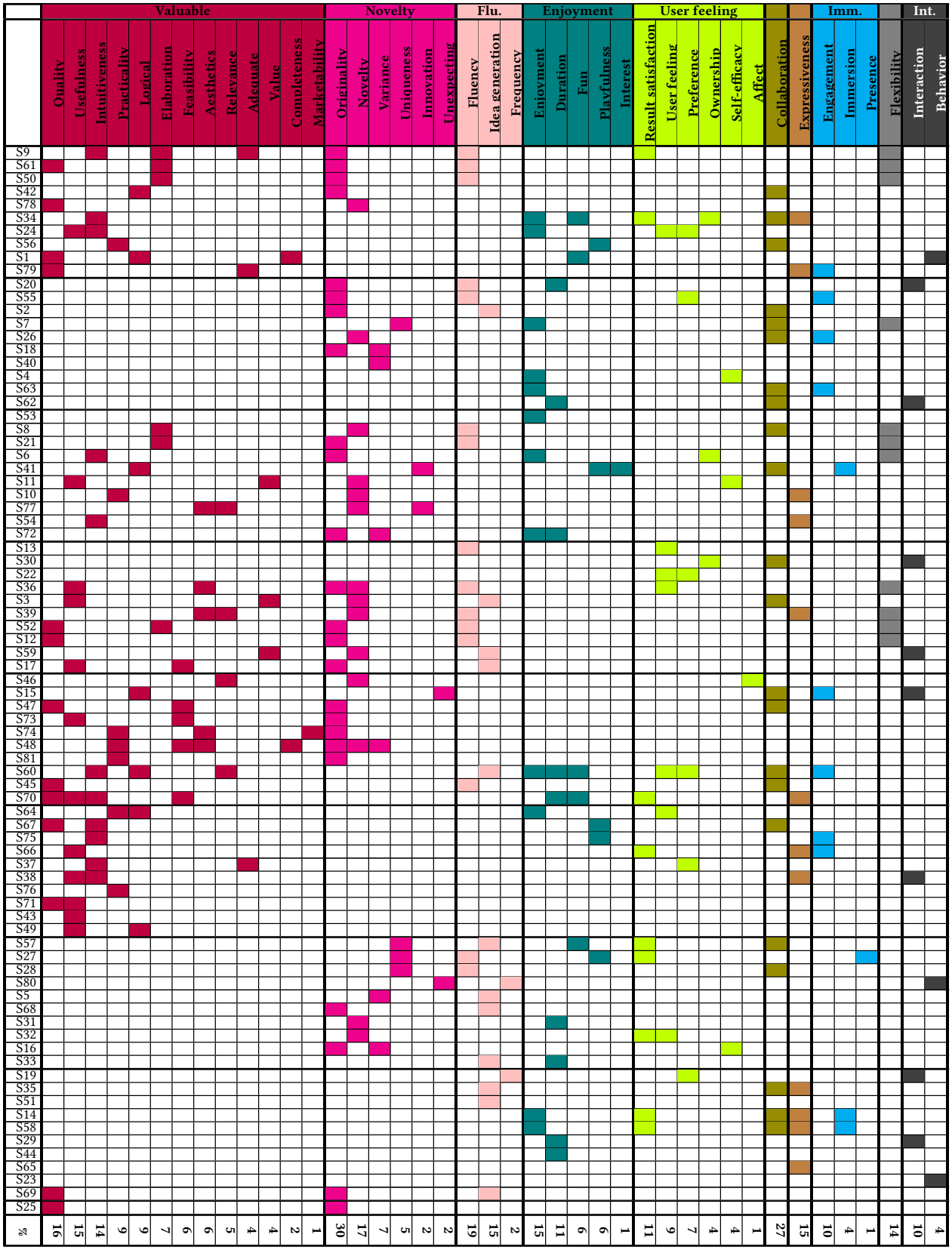


Table 6. Distribution of terms used to evaluate the different factors.
(Flu. = Fluency, Imm. = Immersion, Int. = Interaction)



4.2.1 Valuable. Oxford Dictionary defines *valuable* as "very useful or important" [45]. Table 6 indicates that the studies used 13 different terms to evaluate *valuable*, where *value* is the most related term. However, the most used term is *quality*, which is included in 13 studies. Four of these studies defined *quality*, but it was a variance in the definitions. S25 defined it as "the length of the participant response" on p. 150, S67 as "the property of a tool to provide the user with task specific support and to allow selecting and arranging this support for future re-use" on p. 8, S78 as "coherence, continuation and completion" on p. 7, and S79 as "the kit is to be used by people of different age groups, the device should be safe and should work reliably" on p. 13. Despite this, all these definitions agree with Oxford Dictionary's definition of *valuable* [45] since it is a term depending on the context.

Three of the 12 studies that used the term *usefulness* defined it. S11 used the definition "functions, practicality, and attractiveness" on p. 105, while S36 defined it as "how well the product does what it is supposed to do" on p. 14. In addition, four of the 11 studies focusing on *intuitiveness* defined it. They did it with a similarity; S6 as "easy to use" on p. 421, S9 as "user-friendliness of product" on p. 26, S24 as "the degree to which an individual believes that attending the activity is easy" on p. 35, and S67 elaborating these definitions to "the property of a tool to foster a rapid and clear understanding of the artifacts employed for idea development" on p. 8. These definitions of *usefulness* and *intuitiveness* correspond with each other and the definition of *valuable* to a larger degree than the definitions of *quality*.

An example of a CEM of *quality* was S25 specifying an objective psychophysical CEM based on AUT by examining the participant's response length. On the other hand, S45 used a subjective self-reporting questionnaire including questions like "How would you rate the overall quality of that performance?" on p. 2299. The CEMs of *intuitiveness* also included both objective and subjective CEMs. For example, S54 conducted objective psychophysical observations of the participants' prototyping and outcomes, and S75 did subjective self-reporting interviews where the participants self-reported their experience. However, the CEMs of *usefulness* was only subjective. An example was S17 using two external expert judges to evaluate how an idea helped solve the problem on a five-point Likert scale.

4.2.2 Novelty. According to the Oxford Dictionary [45], *novelty* is "the quality of being new, different and interesting". The studies used six different terms to evaluate *novelty*, as indicated in table 6. Fourteen of the studies used exactly the word *novelty*, but approximately half of them did not specify how they defined it. For example, S10 only stated that a scale including *novelty* was included. However, those who defined it did it differently. S3 focused on *surprising*; S11 on *uniqueness*; S39 on *rarity*; S46 on *infrequency* and *rarity*; S8 on *unusual*, *uniqueness*, and *surprising*; and S59 on *novel*, *originality*, and *surprising*.

Nevertheless, 24 of the studies used the term *originality* to evaluate *novelty*. More than half of them did not specify how they defined it. As with *novelty*, those who did had a different focus. S2, S17, and S52 focused on *novelty*; S6, S21, and S61 on *uniqueness*; S55 on *unusual*; S20 on *unusual*, *relevance* and *rarity*; S36 on *surprising*, *unusual*, *uniqueness*, and *rarity*; S47 on *novelty*, *unusual*, and *infrequent*; and S50 on *innovative*, *unusual*, and *fresh*. Most of these terms correspond with the mentioned terms used to define *novelty*, matching Oxford Dictionary's definition [45].

Originality was evaluated both subjectively and objectively in the included studies. S73 used a subjective CEM with external expert judges using a five-point Likert scale, while S61 objectively observed the number of unique responses of the participants psychophysically. *Variance* could also be measured objectively. An example is S40, which psychophysically compared the nine-element vector v of a student's programming solutions based on absorption, user control, diffusion, generation, transportation, collision, hill-climbing, push, and pull to the corresponding nine-element vector u of a tutorial norm calculated as the difference between these vectors with the formula $\frac{\sqrt{\sum_{i=1}^{n}(u_i - v_i)^2}}{\sqrt{n}}$.

4.2.3 Fluency. Oxford Dictionary defines *fluency* as "the quality of doing something in a way that is smooth and shows skill" [45]. The studies used three different terms to evaluate this, as illustrated in table 6. In contrast to *valuable* and *novelty*, none of the studies use several of these terms. Most of the studies used exactly the word *fluency*, while 12 used *idea generation* and two used *frequency*. Most of the studies that used the term *fluency* defined it. They did so in a uniform manner that corresponds with the definition provided in Oxford Dictionary [45]. In fact, S21, S27, and S28 used the same definition on respectively p. 3, 170, and 143: "the number of ideas generated". However, there were some variations, for example, S36 stating "the number of interpretable, meaningful and relevant responses" on p. 14 and S52 stating "the frequency of ideas generated" on p. 206.

Fluency is mostly measured objectively by psychophysically counting the number of relevant ideas. Most of the studies miss theoretical grounding for choosing *idea generation* as an evaluation factor. On the other hand, S2 said on p. 424: "According to literature, idea generation (or fluency) is one fundamental creative ability that tends to lead to originality and novelty [25]. With YOLO, we aim to stimulate fluency during the creative process of storytelling".

4.2.4 Enjoyment. According to Oxford Dictionary [45], *enjoyment* is "the pleasure that you get from something". Table 6 shows that the included studies used five different terms to evaluate *enjoyment*. Twelve studies used exactly the word *enjoyment*, but only S24 defined it. This definition was on p. 35: "the degree to which the activity is perceived to be personally enjoyable", which corresponds with the definition provided by Oxford Dictionary [45].

Most of the studies used subjective self-reporting CEMs of *enjoyment*. S4 used a questionnaire with the question "How much did you enjoy the creativity project? (Circle a number on the scale below.)" on p. 6, while S64 conducted interviews including the question "Do you enjoy using the system?" on p. 163. However, S7 used psychophysical observation in the form of a camera to recorded interactions around TellTable, including *enjoyment*. *Duration* could also be measured objectively by psychophysically observing users' interaction time like it was done in S62. On the other hand, S60 evaluated *duration* subjectively with a self-reporting questionnaire including the statement "Time appeared to go by quickly when I was interacting with the activity" on p. 109.

4.2.5 User feeling. Feeling is defined as "an attitude or opinion about something" by Oxford Dictionary [45]. The included studies used six different terms to evaluate *user feeling*, as illustrated in table 6. Seven of the studies used exactly the word *user feeling*, but nine used the term *result satisfaction*. Few defined *result satisfaction*, but S27, S34, and S66 did it equally and corresponding to Oxford Dictionary's definition of feeling [45]. The definition was respectively "[the users'] perceptions of [...] their avatar" on p. 170, "requires the users to evaluate the final result of the creative process"

on p. 64, and "happy with the final design" on p. 165.

User feeling was evaluated subjectively by self-reporting, both with questionnaires and interviews. For example, S24 used on p. 35 the Likert statements "I was satisfied with the activity", "I was pleased with the activity", and "My decision to participate in the activity was a wise one". In addition, S36 interviewed the subjects about their opinion of using the CST.

4.2.6 Collaboration. According to Oxford Dictionary [45], *collaboration* is "the act of working with another person or group of people to create or produce something". Twenty-two of the included studies evaluated this factor, as presented in table 6, each of them using the word *collaboration*. Very few defined *collaboration*, but S34 did it as "mutual influence, sharing and feedback with different agents" on p. 65, and S8 as "co-operation time, which is the time that both participants in a group were effectively co-manipulating the platform during the time needed to complete" on p. 148. Both of these definitions correspond with the definition provided by Oxford Dictionary [45].

Different CEMs were used to evaluate *collaboration*. S41 observed it objectively by psychophysically identifying episodes when both participants contributed to an idea. S56 subjectively measured it on p. 71 with the question "With what system did you collaborate the most?" in the self-reporting Fun Sorter and the Likert question "How much did you collaborate with the Graphical/Tangible interface?". Further, S57 calculated *collaboration* psychophysically through the inequity index $I = |\frac{1}{N} - \frac{O_i}{\sum_{i=1}^{N} O_i}| \times 100$, where N is the size of the group and O_i is the observed collaborative behavior for each participant.

4.2.7 Expressiveness. The Oxford Dictionary defines *expressiveness* as "the quality of expressing somebody's thoughts and feelings" [45]. Twelve studies evaluated this factor and all of them used the word *expressiveness*, as illustrated in table 6. Few defined it, and those who did used different definitions. S34's definition matches Oxford Dictionary's definition [45] stating "how well the users are able to be creative and express themselves in the creative process" on p. 64. On the other hand, S35's definition ("the success in producing a variety of drawings for a specified task" on p. 181), S54 ("the wide range of constructions that the kit makes possible for both adults and kids" on p. 346), and S79 ("encourage users to explore topics through a new form of storytelling medium" on p. 13) is more about *variance*.

The studies used both subjective and objective CEMs. For example, S14 had on p. 6 a self-reporting questionnaire including the agreement statements "I was able to be very creative while doing the activity inside this system or tool" and "The system or tool allowed me to be very expressive". Simultaneously, S54 observed psychophysically the range of constructions created with the kit by both children and adults.

4.2.8 Immersion. According to the Oxford Dictionary [45], *immersion* is "the state of being completely involved in something". As indicated in table 6, three different terms are used to evaluate *immersion*. Only three of the studies used the word *immersion*, while eight used the term *engagement*. Few of these studies specified how they understood the term, and those who did, understood it differently. S66 saw it as *absorption with the activity*, which is most similar to the definition of *immersion* provided by Oxford Dictionary [45]. On the other hand, S63 saw it as *curiosity* and S79

as *interest*.

A majority of the studies used self-reporting subjective CEMs of *immersion*. For example, S79 interviewed the participants regarding their engagement and S60 conducted a qualitative questionnaire on p. 109 including the statements "My interest in the subject matter grew as I did the task" and "As I carried out the activity I was absorbed in it". On the other hand, S66 calculated psychophysically the difference between perceived time and completion time. Other studies used objective psychophysical observation, like S75 asking the participants to think aloud throughout the trial to observe their engagement behavior.

4.2.9 Flexibility. Oxford Dictionary defines *flexibility* as "the ability to change to suit new conditions or situations" [45]. Eleven studies evaluated this factor, as illustrated in table 6, and all used the word *flexibility*. Most of the included studies specified how they understood the term. *Flexibility* was understood in the same manner but explained differently in the studies. In short, S36 defined *flexibility* as a "variety of categories of relevant responses" on p. 14. S39 elaborated this to be more similar to the definition to Oxford Dictionary [45], stating on p. 11: "consideration of alternative interpretations, which means ways of thinking and viewpoints. Flexibility in thinking describes the cognitive process of trying out a various ways of looking at a problem. Flexibility measures the span of the solution space explored during ideation".

All the studies evaluated *flexibility* as S36 and S39 by psychophysically observing the number of different categories that the ideas belonged to. A few studies specified the CEM in detail. For example, S61 calculated *flexibility* by the number of tools, principles, and procedures used to explain ideas. On the other hand, S8 used external expert judges to rate the user's *flexibility*.

4.2.10 Interaction. According to the Oxford Dictionary [45], "if one thing has an interaction with another, or if there is an interaction between two things, the two things have an effect on each other". Eleven of the included studies evaluated this factor, as shown in table 6, where eight used the word *interaction* and three used the word *behavior*. However, none explained how they understood these terms. The CEMs was exclusively objectively and psychophysically observing the participants' interaction with the CST.

4.2.11 Other. Twelve different terms did not fit in the 10 previous factors and were included in the *Other* category. It was variance between these terms. *Persistence*, *risk-taking*, *self-determination*, *start help*, *success guarantee*, and *accessibility* were only present in one study. On the other hand, *emergence* occurred in two studies; *inspiration*, *motivation*, and *imagination* in three; and *knowledge* in four. *Exploration* was the term included in most studies with a frequency of seven.

4.3 Context

This section addresses RQ1.2 regarding the context of the CEMs. In this case, the context is the evaluated products, the approach used to evaluate creativity, and the methods used for data collecting and data analysis.

4.3.1 Products. The distribution of the creativity topics for the evaluated products in the studies is presented in table 5 and further specified in table 7. A total of nine different creativity topics

appeared, and the ones that appeared the most frequently were idea generation, model making, and game.

Table 7. The distribution of the creativity topics and software for the evaluated products.

Topic	Software	Percentage of studies
Idea generation	• Self-developed digital idea-generation tools (S3, S31, S33, S35, S39, S44, S52, S59, S67, S80) • Self-developed digital brainstorming tools (S17, S60, S69) and brainwriting tool (S57) • Brainstorming in a video conference system (S51) and Adobe Connect (S47) • The idea generation apps for children AppLab and Common Sense Media (S55) • Virtual reality (VR) idea generation tool (S21, S77) • Self-developed online multiplayer ideation game (S73)	25
Model making	• Self-developed digital model making kits (S9, S54, S79) and prototype tools (S8, S66, S70, S76) • Computer aided design (CAD) (S5), the CAD Rhino (S74), VR CAD (S19), and 3D-CAD (S10, S11) • Self-developed 3D-model tools (S65, S75) and 3D-sketch app (S72) • Self-developed digital multimedia learning tool for mechanism design (S36)	20
Game	• Self-developed game creation apps (S13, S22) • Game development platforms (ARIS (S4), Construct (S48), Powerpoint (S48), Powtoom (S48), and Roar (S48)) • Self-developed digital gamified Learning Management System (S46) • The digital problem-solving game Crayon Physics Deluxe (S20) • The digital constructing game Minecraft (S68) • The VR idea generation game Second Life (S27, S28) • The educational computer game I Spy Treasure Hunt (S50)	14
Storytelling	• Self-developed digital storytelling tools (S1, S2, S7, S37, S38, S42, S62, S63) • Short story writing in Google Docs (S14) • Storytelling with the software Tuxpaint combined with the Our Story app (S41) • The digital storytelling software Frames combined with Powerpoint (S78)	12
Music	• Self-developed digital music creation tools (S18, S23, S32, S45) • Self-developed app for music creation and manipulation through movement (S30) • Visual music interactive art system with Processing and Leap Motion (S12) • Self-developed adaptive digital musical instrument (S25) • Self-developed educational digital music game (S53)	10
Drawing	• Self-developed digital co-creative drawing agent (S15) • Self-developed sketch combination system integrating Mechanical Turk and Google Docs (S81), OdoScetch (S14), and AutoDesk Scetchbook (S14) • Self-developed digital painting tool (S6) and the Java app Pixelitor (S49) • Self-developed digital color picker (S58) and Mac OS's color picker BiCEP (S14)	7
Programming	• App Inventor (S26) • Scratch (S24, S26, S48, S61) • Unity (S48) • Unspecified programming tools (S40, S56)	7
Visualization	• Self-developed digital photography tool (S29) • Adobe Photoshop (S14) • Self-developed digital data visualization tool (S64) and font visualization tool (S71) • MySpace's Flash-based AdBuilder Tool (S16)	6
Writing	• Self-developed digital poetry tool (S34) • Self-developed digital journalism tool (S43)	2

4.3.2 Approaches. Figure 4 illustrates the distribution of Caroll and Latulipe's [7] three basic approaches to CEMs in the included studies. This distribution is also illustrated in table 5. The most frequently used approach is psychophysical measurement, with a coverage of 77% of the included studies. Self-reporting is also a well-used approach with a coverage of 65%, while external judges only have a coverage of 32%. Figure 4 illustrates that most of the studies used a combination of these approaches, particulary psychophysical measurement and self-reporting, which was used by 33 of the studies. In addition, seven of the studies employed all these approaches.

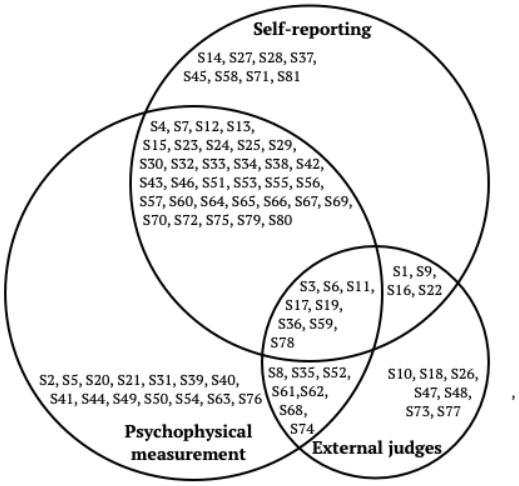


Fig. 4. The distribution of the approaches in the included studies.

4.3.3 Methods. Table 5 indicates that the studies used different methods and a combination of different methods. A majority of the studies used multiple data collection methods. Observation was the most frequently used data collection method with 77% coverage, while questionnaire and interview followed with 70 and 32% respectively. The data analysis was mostly qualitative, with 74% coverage. However, 59% of the studies utilized quantitative data analysis and 33% a mixed method.

A few studies also used existing CEMs. The most frequent was the Likert Scale, which appeared in 23 of the studies. It was used in both three- (S55), five- (S1, S17, S31, S32, S36, S55, S56, S58, S66, S73, S77, S79, S80), seven- (S16, S24, S25, S28, S37, S57, S59, S81), nine-(S45) and 10-point scales (S34). Other known CEMs that appeared in the studies were AUT (S21, S25, S31), TTCT (S36, S50, S61), (S12), MSFM (S20), RAT (S21), CAQ (S16), K-DOCS (S77), TIPS combined with Buzan (S48), CAT (S18, S78), AttrakDiff (S22), CPSS (S36), CSI (S14, S58), Fun-Sorter (S56), and MemoLine (S53).

A few studies missed information regarding the methods. S25 did not mention anything about participants. Further, 27 of the studies mention participants, but miss information regarding the gender distribution among the participants, while S48 and S69 miss information regarding sample size and age, respectively.

4.4 Age of the target group

This section addresses RQ1.3 regarding how the target group's age impacts the included studies. Table 5 indicates that the studies' target group is mostly adults. Further, 72% of the included studies used adults as participants and 41% used children, while 15% used both types of participants; 2% of the studies (S25, S69) did not specify the participant's age.

Table 8 indicates that the age of the target group impacts the evaluation factors, products, and methods. The most frequent factors are not similar for these different age groups. For example, *enjoyment* with 39% coverage and *collaboration* with 33% coverage for children have 28% and 24% coverage for adults, respectively. However, *fluency* has 38% coverage for adults but 27% coverage

for children.

Table 8. Coverage percentage in studies with children and adults. (Appr. = Approaches)

	Factors											Products									Appr.			Methods				
	Valuable	Novelty	Fluency	Enjoyment	User feeling	Collaboration	Expressiveness	Immersion	Flexibility	Interaction	Other	Idea generation	Model making	Storytelling	Game	Music	Drawing	Programming	Visualization	Writing	Psychophysical	Self-reporting	External judges	Observation	Questionnaire	Interview	Qualitative	Quantitative
Children	55	61	27	39	30	33	12	15	21	12	27	9	21	24	15	9	3	15	0	3	79	55	39	79	64	36	73	67
Adults	60	53	38	28	33	24	17	14	12	14	24	31	27	7	12	9	10	2	9	2	74	67	33	74	72	26	72	53

The products in the studies are also different for adults and children. The most common product for children is storytelling with 24% coverage, which only has 7% coverage for adults. In addition, programming is the third most common product for children with 15% coverage but has only 2% coverage for adults. For adults, the definitively most common product is idea generation with 31% coverage, which only has 9% coverage for children. Another interesting fact is that there are no visualization products for children and only 3% coverage of drawing product, which are the fifth and fourth most common product for adults with 9% and 10% coverage, respectively. Thus, there is a knowledge gap in evaluating the creativity support for children's drawing and visualization products.

The approaches in the included studies depend on the participant's age. For adults, there was just more psychophysical measurement (74%) than self-reporting (67%), while the studies with children definitely had the most psychophysical measurement (79% against 55%). This indicates that the CEMs in the studies with children have slightly more qualitative (73%) than quantitative (67%) data analysis, while adults clearly have the most qualitative data analysis (72% against 53%). With regard to the data collection method, there are an almost equal number of observations (74%) and questionnaires (72%) for adults, while the studies with children have clearly most observations of their interacting with the product (79%) instead of complicated questionnaires (64%).

4.5 Effectiveness

This section addresses RQ1.4 on the effectiveness of the CEMs in the included studies. It is divided into three parts: the first is regarding how the included studies understood the term creativity, the second is regarding the validity of the methods in the included studies, and the last is regarding the outcomes of the studies.

4.5.1 Creativity. Forty-two of the primary studies do not define creativity, and nine of the 39 studies that defined it used unclear definitions, as illustrated in figure 5. An example is S69 stating on p. 103, "Identifying new ideas can be difficult due to individuals' limited vision, knowledge, experience, motivation and time. Collaborative teamwork that pools and integrates efforts from multiple individuals is thus considered a useful way to approach creativity".

There is no unambiguous opinion of creativity. There are a few overlaps among the different definitions, but 17 of the studies include a definition of something with the two characteristics *novel* and *valuable*. Only two of the studies (S28, S57) uses the same definition of creativity by Sternberg as "the ability to produce work that is both novel and appropriate" [62, p. 3]. Thirteen of the studies

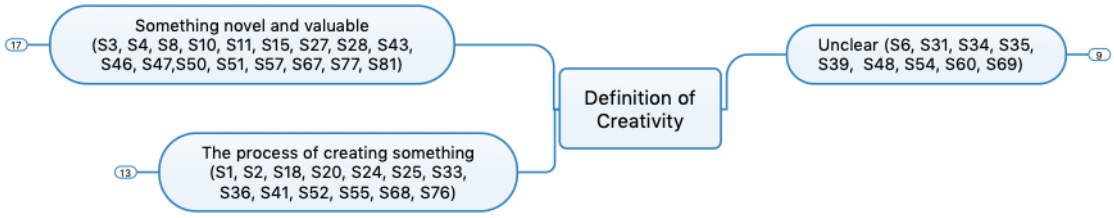


Fig. 5. Definitions of creativity in the 39 studies defining the term.

focus more on the process of creating something. For example, S20 defined creativity as "the process of producing a work, of any kind (e.g. artifact, device, idea), that is considered remarkable and original within the framework of a community" on p. 209. Both of these two different types of definitions correspond to the definition provided by Oxford Dictionary [45].

4.5.2 Validity of the methods. The validity of the methods was not mentioned in 37 of the included studies. Among the 44 studies including details of the validity, 17 discussed the study environment regarding duration, atmosphere, or location, 14 the sample of participants regarding sample size or representativeness, 12 potential biases, seven the evaluation factors, seven use cases of the method or the CST, and six measuring creativity, as depicted in figure 6. A few of the studies included several of these six categories. With regard to evaluation factors, S14 emphasized the difficulty of including *collaboration* in a standardized creativity index because many creative activities are not collaborative. In addition, S34 finds the factor *fun* to be less important than expected.

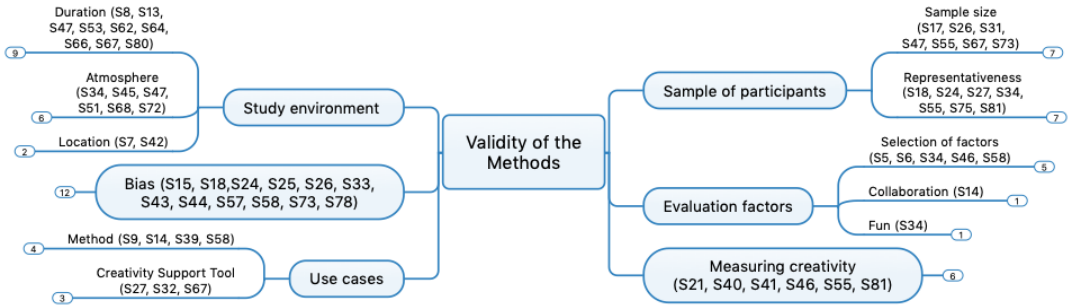


Fig. 6. The validity of methods in the 44 studies including details of the validity.

The studies have different approaches for selection of evaluation factors. S6 and S34 chose the evaluation metrics by the partially overlapping metrics in three and nine articles, respectively, while S58 uses the well-tested CSI. On the other hand, S5 recommends including *quality* and *novelty* as evaluation factors, and S46 indicates how *importance*, *novelty*, and *affect* can be used to assess creativity in a web-based information design system.

Further, the studies have different limitations in measuring creativity. S21 indicated their measuring of creativity with interpretations of AUT and RAT scores as the study's main limitation. On the other hand, S41 focused on the aspect mentioned by Glăveanu [21, p. 91] regarding the difference in how "members of different communities assess the creativity of one and the same artifact". One limitation of S46 is the difficulty of distinguishing the evaluating of creative design and creative

pedagogy. It involves blurry boundaries between the learning system itself and the content within the system. Further, measuring engagement for young children was a significant challenge in S55. Three different CEMs were utilized because of conflicting ideas, but two of them were excluded due to the validity of the results. Consequently, the study could use only the adapted self-reporting of *engagement*.

In addition, the studies highlighted different CEMs. S40 highlighted divergence tests and considered divergence from the accepted norm as a significant creativity indicator. However, S81 used a binary measure that only qualify designs as creative if it exceeds 4.0 on a seven-point Likert scale on both the scales of the factors *originality* and *practicality*.

4.5.3 Outcome. The outcomes of most included studies was how an existing or self-developed CSTs improves creativity. There are different distribution of creativity topics for the products in these two CSTs categories, as depicted in figure 7. Games dominate the existing CSTs, while idea generating is the most frequently self-developed CSTs. In addition, a few studies investigated gender differences of creativity, or developed a new CEM to evaluate CSTs.

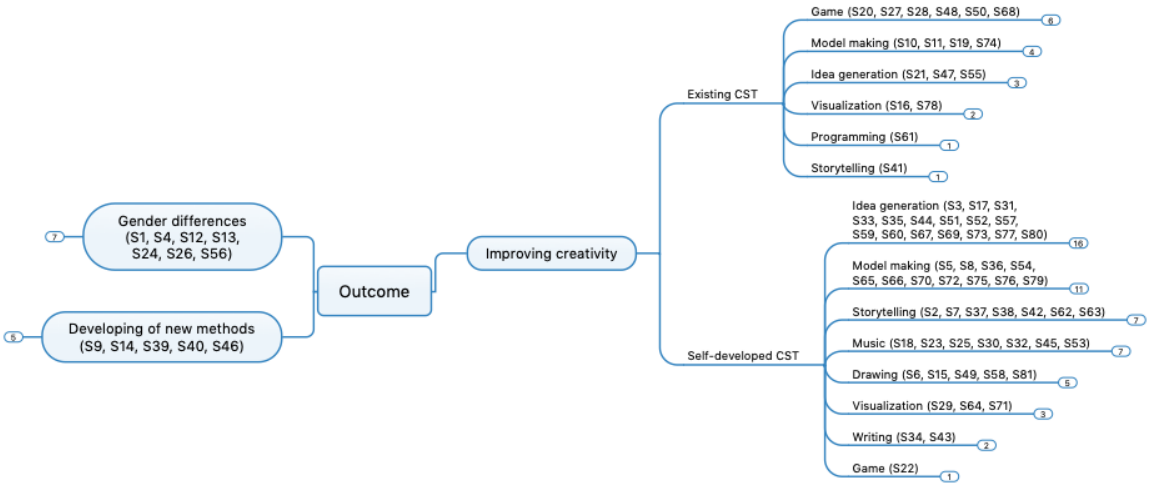


Fig. 7. The outcomes of the included studies.

S14 and S39 are mentioned in section 2 as Cherry and Latulipe's CSI and Kerne et al.'s evaluation of creativity components, respectively. Besides, S40's calculating of the divergence to a standard norm is mention in section 4.2. On the other hand, S9 developed a design model of platform for user creativity, and S46 developed a rubric for evaluating creativity in learning management systems based on the factors *novelty*, *affect*, and *importance*.

5 DISCUSSION

5.1 RQ1: What are the current Creativity Evaluation Methods of digital Creativity Support Tools for children?

The present SLR of 81 papers identified a wide specter of CEMs of digital CSTs for children due to the lack of a standardized CEM. Numerous factors have been considered in the CEMs, but the 10 most common factors were *valuable*, *novelty*, *fluency*, *enjoyment*, *user feeling*, *collaboration*,

expressiveness, *immersion*, *flexibility*, and *interaction*. The CEMs were also performed in different contexts using different products and differed in terms of approaches, data collection methods, and data analysis.

The target group's age impacted the CEMs. The distribution of the factors, products, and methods was different for children and adults. The factors *novelty*, *enjoyment*, and *collaboration* were more frequently utilized for targeting children, while the factors *valuable*, *fluency*, and *user feeling* were utilized more in the studies with adults as participants. Storytelling and programming products were more frequently utilized in the studies with children as participants, in contrast to idea generation and drawing products for adults. Children were evaluated more objectively than adults, with more quantitative psychophysical measurements with observation, while adults' CEMs had a larger amount of subjective qualitative questionnaire with self-reporting. The children's age also impacted the CEMs, as it is more important with quantitative observations than qualitative questionnaires and interviews for younger children.

It is difficult to determine the effectiveness of CEMs. Most of of the studies did not mention the threats to the validity to the study. The included studies did not have a uniform perception of creativity. Several studies did not define it and those that did used different definitions. In addition, the outcomes varied. Many of the studies showed how an existing or self-developed CST improved creativity. As discussed in the literature review by Remy et al. [52], there are three different aspects of evaluating the creativity of a CST; the usability of the CST, the creativity of the outcome, and the productivity of the process supported by the CST. Several of the included studies did not specify which of these aspects they used to evaluate creativity, and, thus, it is difficult to compare the studies to each other. Therefore, we support in section 5.5 Remy et al.'s [52] recommendation of clearly defining the goal of the CEM of the CST.

5.2 Implication for Research

This SLR contributes to researchers by identifying trends and opportunities for future works. It shows a need for more research on this topic by encompassing the wide specter of CEMs. The present study also identified knowledge gaps — for example, with regard to digital drawing and visualization CSTs for children. Further, the SLR identified that promising CEMs for children had not been explored for digital CSTs — for example, the Smileyometer or This or That. There is also a need for more structured CEMs of digital CSTs for children since the current CEMs vary widely. In section 5.5, we therefore support Remy et al.'s [52] recommendation of developing a toolbox for CEMs of CSTs. However, we want to change the focus of the recommendation from a toolbox to a tailored approach since the CEM of a digital CST has to be tailored to the specific CST.

5.3 Implication for Practice

The trends of CEMs identified in this research could help practitioners in companies who develop digital CSTs for children to make better solutions. A challenge is that there is no standardized CEM. However, the psychophysical measurement approach with observation occurred in 79% of the studies with children and is a promising method for companies. This is a good starting point for the CEM since it is easiest to use on the youngest children because they are only observed and do not need to participate in interviews or answer questionnaires. For older children, the included studies often used a combination of qualitative and quantitative methods. Thus, we recommend in section 5.5 to begin with a quantitative psychophysical measurement approach but consider

expanding to qualitative methods. The CEM could also be more structured by presenting the 10 proposed factors in section 4.2 to the companies, including definitions and specific examples of evaluating each factor. Further, the company could choose which factors to focus on after product category and target age. This is also one of our recommendations presented in section 5.5.

5.4 Limitations of the SLR

5.4.1 Completeness. A structured SLR was performed in relevant databases and journals following the guidelines by Kitchenham [39], which resulted in 81 peer-review articles with 77 different main authors. However, it could not be guaranteed that all literature in this area was captured due to continuously new research on this. Figure 3 showed relatively few publications in 2020, but this may be since the automatic search was conducted before the year was over and restriction in research due to the COVID-19 pandemic. Another limitation was the search strategy. It was difficult to find the relevant studies. The term *"assessment"* was not included in the search string in the automatic search, which could lead to missing relevant studies. On the other hand, the manual search compensated to this to a certain extent degree. The manual search opened for more relevant studies because the search string was not the same as the automatic search.

Studies with adults as participant were included in the review because there is scarce research on evaluating digital CSTs focused on children. This could be due to the ethical issues for research with children as participants, which are related to confidentiality, consent, and protection [15]. The inclusion of studies targeting adults allowed comparing the CEMs to the target group's age in RQ1.1.3 and seeing if there were CEMs for adults that could be transformed into children's CEMs.

5.4.2 Potential Biases. This SLR has multiple potential biases. First, a large number of the included studies do not focus on evaluating creativity. They focus on evaluating usability and technical aspects instead of creativity. This also occurred in the literature review by Remy et al., where it was a strong tendency that the most recent papers in the review used usability testing to evaluate CSTs [52]. Remy et al. recommend deciding whether to evaluate creativity or usability, but we will instead recommend focusing on both. Usability is an important product evaluation aspect, as it is one of the eight main quality characteristics in the software product quality model ISO/IEC 25010:2011 [34]. Nevertheless, creativity must not be forgotten as it is a very beneficial quality for children. Since the evaluation factor *valuable* is closely related to usability, we merge this recommendation with the recommendation of the 10 factors in section 5.5.

Further, a few of the studies lacked sufficient details. A large number of them did not ground their choice of CEM adequately in theory. This also occurred in the literature review by Remy et al. [52], where less than half of the included studies' CEMs built on identifiable theoretical foundations. Thus, we support the recommendation of using theoretical grounding in the CEM of the CST in section 5.5. Good examples of this are S6 and S34, which based their selection of evaluation factors on three and nine studies, respectively. However, this recommendation in section 5.5 is merged with the recommendation of the 10 factors because these factors and the definitions of these factors are based on theoretical groundings of evaluating creativity. The included studies also had different perceptions of creativity. A few of the studies did not even clarify the definition of creativity. This case also applied to most of the evaluation factors. Consequently, a direct comparison between the studies is because of different interpretations of the term. However, *flexibility* was the most

uniform factor where most studies provided a similar definition.

5.4.3 Data Synthesis. The data was categorized to identify themes that answer the RQs. However, there are issues with qualitative data analysis. There was some overlap between the categories. For example, the product in S53 is a self-developed educational digital music game. It was placed in the music category, but it also belonged to the game category. There is also a possibility of wrong interpretation in the data extraction process because a few studies did not describe the data to be extracted adequately. To reduce bias, the first author did this categorizing alone and discussed it with the second and third authors.

5.5 Recommendations for Creativity Evaluation Methods of digital Creativity Support Tools for children

To sum up the discussion, we support Remy et al.'s [52] recommendations of recommendation (R)1, R2, and R4 to also be applied to CEMs of digital CSTs for children because the same issues occurred in our SLR. Based on the findings, we changed R2's center from theoretical groundings to the 10 factors, including usability in the factor *valuable*. The focus of R4 was changed from a toolbox to a tailored approach because the findings indicated that the CEM has to be tailored to the software product. Based on the used methods in the included studies, we also added R3. Thus, we recommend the following four Rs for CEMs of digital CST for children:

- **R1)** Clearly define the goal;
- **R2)** Review and choose amongst the 10 evaluation factors *valuable*, *novelty*, *fluency*, *enjoyment*, *user feeling*, *collaboration*, *expressiveness*, *immersion*, *flexibility*, and *interaction*;
- **R3)** Start with a quantitative psychophysical measurement approach but consider expansion to qualitative methods;
- **R4)** Develop a tailored approach for the evaluation of creativity.

6 CONCLUSION

This SLR analyzed 81 peer-reviewed conference papers and articles from the last 10 years. The review aimed to investigate the state-of-the-art procedures of CEMs of digital CSTs for children. A wide variation of CEMs for children was found. Since a standardized CEM is lacking, it is difficult to reach a conclusion regarding the effectiveness of the CEMs. The 10 most common factors in the CEMs were the following categories: *valuable*, *novelty*, *fluency*, *enjoyment*, *user feeling*, *collaboration*, *expressiveness*, *immersion*, *flexibility*, and *interaction*. The CEMs was performed in different contexts, impacted by the age of the target group. This research contributes to both practitioners and researchers by identifying trends in CEMs and encompassing areas with a need for further investigation. Future research could focus on the knowledge gap in evaluating drawing and visualization CSTs for children and the promising CEMs Smileyometer and This or That, which have not yet been explored for digital CSTs for children. Based on the findings, we have six recommendations for CEMs of digital CST for children: R1) Clearly define the goal; R2) Review and choose amongst the 10 evaluation factors *valuable*, *novelty*, *fluency*, *enjoyment*, *user feeling*, *collaboration*, *expressiveness*, *immersion*, *flexibility*, and *interaction*; R3) Start with a quantitative psychophysical measurement approach but consider expansion to qualitative methods; R4) Develop a tailored approach for the evaluation of creativity.

REFERENCES

[1] Lola Prieto Almeida, Leandro S.and Prieto, Mercedes Ferrando, Emma Oliveira, and Carmen Ferrandiz. 2008. Torrance Test of Creative Thinking: The question of its construct validity. *Thinking Skills and Creativity* 3, 1 (April 2008), 53–58. https://doi.org/10.1016/j.tsc.2008.03.003

[2] Teresa M. Amabile. 1983. The social psychology of creativity: A componential conceptualization. *Journal of Personality and Social Psychology* 45, 2 (Nov. 1983), 357–376. https://doi.org/10.1037/0022-3514.45.2.357

[3] Teresa M. Amabile. 1996. *Creativity in Context: Update to the Social Psychology of Creativity.* Westview Press, Boulder, USA.

[4] Lorin W. Anderson, David R. Krathwohl, Peter W. Airasian, Kathleen A. Cruikshank, Richard E. Mayer, Paul R. Pintrich, James Raths, and Merlin C. Wittrock. 2001. *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives.* Longman, New York, USA.

[5] Susan P. Besemer and Donald J. Treffinger. 1981. Analysis of creative products: Review and synthesis. *The Journal of Creative Behavior* 15, 3 (Sept. 1981), 158–178. https://doi.org/10.1002/j.2162-6057.1981.tb00287.x

[6] Tony Buzan. 1993. *The Mind Map Book.* BBC Books, London, UK.

[7] Erin A. Carroll and Celine Latulipe. 2012. Triangulating the Personal Creative Experience: Self-report, External Judgments, and Physiology. In *Proceedings of the Graphics Interface Conference 2012* (Toronto, Canada, May 28–30) *(GI '12)*. Canadian Information Processing Society, Mississauga, Canada, 53–60. https://doi.org/10.5555/2305276.2305286

[8] Shelley H. Carson, Jordan B. Peterson, and Daniel M. Higgens. 2005. Reliability, validity, and factor structure of the creative achievement questionnaire. *Creativity Research Journal* 17, 1 (Dec. 2005), 37–50. https://doi.org/10.1207/s15326934crj1701_4

[9] Justine Cassell and Kimiko Ryokai. 2001. Making space for voice: Technologies to support childrens fantasy and storytelling. *Personal and Ubiquitous Computing* 5, 3 (Aug. 2001), 169–190. https://doi.org/10.1007/PL00000018

[10] Yushan Chang. 2014. 3D-CAD effects on creative design performance of different spatial abilities students. *Journal of Computer Assisted Learning* 30, 5 (Oct. 2014), 397–407. https://doi.org/10.1111/jcal.12051

[11] Erin Cherry and Celine Latulipe. 2014. Quantifying the Creativity Support of Digital Tools through the Creativity Support Index. *ACM Transactions on Computer-Human Interaction* 21, 4, Article 21 (June 2014). https://doi.org/10.1145/2617588

[12] Daniela Soares Cruzes and Tore Dybå. 2011. Recommended Steps for Thematic Synthesis in Software Engineering. In *Proceedings of the 2011 International Symposium on Empirical Software Engineering and Measurement* (Banff, Canada, September 22–23) *(ESEM '11)*. IEEE, Piscataway, USA, 275–284. https://doi.org/10.1109/ESEM.2011.36

[13] Mihaly Csikszentmihalyi. 1997. *Creativity: Flow and the Psychology of Discovery and Invention.* Harper Perennial, New York, USA.

[14] Tore Dybå and Torgeir Dingsøyr. 2008. Empirical studies of agile software development: A systematic review. *Information and Software Technology* 50, 9–10 (Aug. 2008), 833–859. https://doi.org/10.1016/j.infsof.2008.01.006

[15] Nahid Ferdousi. 2015. Children as Research Subjects: The Ethical Issues. *Bangladesh Journal of Bioethics* 6, 1 (Aug. 2015), 6–10. https://doi.org/10.3329/bioethics.v6i1.24398

[16] Jonas Frich, Lindsay MacDonald Vermeulen, Christian Remy, Michael Mose Biskjaer, and Peter Dalsgaard. 2019. Mapping the Landscape of Creativity Support Tools in HCI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland, May 4–9) *(CHI '19)*. ACM, New York, USA, 1–18 (Paper 389). https://doi.org/10.1145/3290605.3300619

[17] Andrea Gaggioli, Elvis Mazzoni, Luca Milani, and Giuseppe Riva. 2015. The creative link: Investigating the relationship between social network indices, creative performance and flow in blended teams. *Computers in Human Behavior* 42, 2015 (Jan. 2015), 157–166. https://doi.org/10.1016/j.chb.2013.12.003

[18] Michail N. Giannakos and Letizia Jaccheri. 2013. What motivates children to become creators of digital enriched artifacts?. In *Proceedings of the 9th ACM Conference on Creativity & Cognition* (Sydney, Australia, June 17–20) *(C&C '13)*. ACM, New York, USA, 104–113. https://doi.org/10.1145/2466627.2466634

[19] Michail N. Giannakos, Letizia Jaccheri, and Roberta Proto. 2013. Teaching Computer Science to Young Children Through Creativity: Lessons Learned from the Case of Norway. In *Proceedings of the 3rd Computer Science Education Research Conference on Computer Science Education Research* (Arnhem, Netherlands, April 4–5) *(CSERC'13)*. Open Universiteit, Heerlen, Netherland, 103–111. https://doi.org/10.5555/2541917.2541927

[20] Michail N. Giannakos, Zacharoula Papamitsiou, Panos Markopoulos, Janet Read, and Juan Pablo Hourcade. 2020. Mapping child-computer interaction research through co-word analysis. *International Journal of Child-Computer Interaction* 23–24, 2020 (June 2020), 100165. https://doi.org/10.1016/j.ijcci.2020.100165

[21] Vlad Petre Glăveanu. 2010. Paradigms in the study of creativity: Introducing the perspective of cultural psychology. *New Ideas in Psychology* 28, 1 (April 2010), 79–93. https://doi.org/10.1016/j.newideapsych.2009.07.007

[22] Kathy Goff and Ellis Paul Torrance. 2002. *Abbreviated Torrance Test for Adults: Manual.* Scholastic Testing Service, Bensenville, USA.

[23] Gerard Goggin. 2011. Ubiquitous apps: politics of openness in global mobile cultures. *Digital Creativity* 22, 3 (Sept. 2011), 148–159. https://doi.org/10.1080/14626268.2011.603733
[24] Joy Paul Guilford. 1950. Creativity. *American Psychologist* 5, 9 (1950), 444–454. https://doi.org/10.1037/h0063487
[25] Joy Paul Guilford. 1967. *The Nature of Human Intelligence.* McGraw-Hill, New York, USA.
[26] Jean-Xavier Guinard. 2000. Sensory and consumer testing with children. *Trends in Food Science & Technology* 11, 8 (Aug. 2000), 273–283. https://doi.org/10.1016/S0924-2244(01)00015-2
[27] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (task load index): Results of empirical and theoretical research. *Advances in Psychology* 52, 1 (1988), 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9
[28] Marc Hassenzahl, Michael Burmester, and Franz Koller. 2003. AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualitat. In *Mensch & Computer 2003: Interaktion in Bewegung*, Gerd Szwillus and Jurgen Ziegler (Eds.). Vieweg+Teubner Verlag, Wiesbaden, Germany, 187–196. https://doi.org/10.1007/978-3-322-80058-9_19
[29] Dennis Hocevar. 1981. Measurement of Creativity: Review and Critique. *Journal of Personality Assessment* 45, 5 (Oct. 1981), 450–464. https://doi.org/10.1207/s15327752jpa4505_1
[30] Robyn M. Holmes, Brianna Gardner, Kristen Kohm, Christine Bant, Anjelica Ciminello, Kelly Moedt, and Lynn Romeo. 2019. The relationship between young children's language abilities, creativity, play, and storytelling, Early Child Development and Care. *Early Child Development and Care* 189, 2 (2019), 244–254. https://doi.org/10.1080/03004430.2017.1314274
[31] Diana Horn and Gavriel Salvendy. 2006. Consumer-based assessment of product creativity: A review and reappraisal. *Human Factors and Ergonomics in Manufacturing & Service Industries* 16, 2 (March 2006), 155–175. https://doi.org/10.1002/hfm.20047
[32] Hsien-Sheng Hsiao, Kai-Hsin Wong, Meng-Jie Wang, Kuang-Chao Yu, Kuo-En Chang, and Yao-Ting Sung. 2006. Using Cognitive Affective Interaction Model to Construct On-Line Game for Creativity. In *First International Conference on Technologies for E-Learning and Digital Entertainment* (Hangzhou, China, April 16–19) *(Edutainment'06).* Technologies for E-Learning and Digital Entertainment, Hangzhou, China, 409–418. https://doi.org/10.1007/11736639_52
[33] Netta Iivari, Marianne Kinnula, and Leena Kuure. 2014. With best intentions: A Foucauldian examination on children's genuine participation in ICT design. *Information Technology & People* 28, 2 (June 2014), 246–280. https://doi.org/10.1108/ITP-12-2013-0223
[34] ISO/IEC 25010: 2011 2011. *Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – System and software quality models.* Standard. International Organization for Standardization, Geneva, Switzerland. https://webstore.iec.ch/publication/11245
[35] Anna Jordanous and Bill Keller. 2016. Modelling creativity: Identifying key components through a corpus-based approach. *PLOS ONE* 11, 10 (Oct. 2016), e0162959. https://doi.org/10.1371/journal.pone.0162959
[36] James C. Kaufman. 2012. Counting the muses: Development of the Kaufman Domains of Creativity Scale (K-DOCS). *Psychology of Aesthetics, Creativity, and the Arts* 6, 4 (Nov. 2012), 298–308. https://doi.org/10.1037/a0029751
[37] James C. Kaufman and Ronald A. Beghetto. 2009. Beyond big and little: The four c model of creativity. *Review of General Psychology* 13, 1 (March 2009), 1–12. https://doi.org/10.1037/a0013688
[38] 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 Transactions Computer-Human Interactions* 21, 3, Article 14 (June 2014). https://doi.org/10.1145/2591677
[39] Barbara A. Kitchenham. 2004. *Procedures for Performing Systematic Reviews.* Technical Report TR/SE-0401. Department of Computer Science, Keele University, UK. https://www.bibsonomy.org/bibtex/2e48137ec01b6308876e05ab1ecdf4bc4/wiljami74
[40] Carolyn Lamb, Daniel G. Brown, and Charles L. A. Clarke. 2018. Evaluating Computational Creativity: An Interdisciplinary Tutorial. *Comput. Surveys* 51, 2, Article 28 (Feb. 2018), 34 pages. https://doi.org/10.1145/3167476
[41] Ritter Simone M. and Rietzschel Eric F. 2004. Lay Theories of Creativity. In *The Science of Lay Theories*, Claire M. Zedelius, Barbara C. N. Muller, and Jonathan W. Schooler (Eds.). Springer, Cham, Germany, Chapter 7, 95–126. https://doi.org/10.1007/978-3-319-57306-9_5
[42] Karwowski Maciej McKay, Alexander S. and James C. Kaufman. 2017. Measuring the muses: Validating the kaufman domains of creativity scale (K-DOCS). *Psychology of Aesthetics, Creativity, the Arts* 11, 2 (May 2017), 216–230. https://doi.org/10.1037/aca0000074
[43] Martha T. Mednick, Sarnoff A. Mednick, and Edward V. Mednick. 1964. Incubation of creative performance and specific associative priming. *Journal of Abnormal and Social Psychology* 69, 1 (July 1964), 84–88. https://doi.org/10.1037/h0045994
[44] James D. Moran III, Roberta M. Milgram, Janet K. Sawyers, and Victoria R. Fu. 1983. Original thinking in preschool children. *Child Development* 54, 4 (Aug. 1983), 921–926. https://doi.org/10.2307/1129896
[45] Oxford Dictionary. 2021. *Oxford Learner's Dictionary.* Retrieved May 25, 2021 from https://www.oxfordlearnersdictionaries.com/definition/english

[46] Sofia Papavlasopoulou, Michail N. Giannakos, and Letizia Jaccheri. 2017. Empirical studies on the maker movement, a promising approach to learning: A literature review. *Entertainment Computing* 18, 1 (Jan. 2017), 57–78. https://doi.org/10.1016/j.entcom.2016.09.002

[47] Janet C. Read. 2008. Validating the Fun Toolkit: an instrument for measuring children's opinions of technology. *Cognition Technology and Work* 10, 2 (April 2008), 119–128. https://doi.org/10.1007/s10111-007-0069-9

[48] Janet C. Read, Daniel Fitton, and Matthew Horton. 2014. Giving ideas an equal chance: inclusion and representation in participatory design with children. In *Proceedings of the 2014 Conference on interaction Design and Children* (Aarhus, Denmark, June 17–20) *(IDC '14)*. ACM, New York, USA, 105–114. https://doi.org/10.1145/2593968.2593986

[49] Janet C. Read and Stuart MacFarlane. 2006. Using the Fun Toolkit and Other Survey Methods to Gather Opinions in Child Computer Interaction. In *Proceedings of the 2006 Conference on Interaction Design and Children* (Tampere, Finland, June 7–9) *(IDC '06)*. ACM Press, New York, USA, 81–88. https://doi.org/10.1145/1139073.1139096

[50] Janet C. Read, Stuart MacFarlane, and Christopher Casey. 2002. Endurability, engagement and expectations: Measuring children's fun. In *Proceedings of Interaction Design and Children 2* (Eindhoven, Netherlands, December 17) *(IDC'02)*. ACM, New York, USA, 1–23. https://www.researchgate.net/publication/228870976_Endurability_Engagement_and_Expectations_Measuring_Childrenaposs_Fun

[51] Janet C. Read and Panos P. Markopoulos. 2013. Child–computer interaction. *International Journal of Child-Computer Interaction* 1, 1 (Jan. 2013), 2–6. https://doi.org/10.1016/j.ijcci.2012.09.001

[52] Christian Remy, Lindsay MacDonald Vermeulen, Jonas Frich, Michael Mose Biskjaer, and Peter Dalsgaard. 2020. Evaluating Creativity Support Tools in HCI Research. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (Eindhoven, Netherlands, July 6–20) *(DIS '20)*. ACM, New York, USA, 457–476. https://doi.org/10.1145/3357236.3395474

[53] Mitchel Resnick. 2005. Seven issues for creativity support tool researchers. In *National science foundation Workshop Report on Creativity Support Tools*. Chapter 7, 67–70. http://www.cs.umd.edu/hcil/CST

[54] Marc A Runco and Garett J. Jaeger. 2012. The Standard Definition of Creativity. *Creativity Research Journal* 24, 1 (Jan. 2012), 92–96. https://doi.org/10.1080/10400419.2012.650092

[55] Prabir Sarkar and Amaresh Chakrabarti. 2008. Studying Engineering Design Creativity - Developing a Common Definition and Associated Measures. In *Invited Paper in the Proceedings of the NSF Workshop on Studying Design Creativity*, John Gero (Ed.). National Science Foundation, Alexandria, USA. https://cpdm.iisc.ac.in/cpdm/ideaslab/paper_scans/UID_17.pdf

[56] Prabir Sarkar and Amaresh Chakrabarti. 2011. Assessing design creativity. *Design Studies* 32, 4 (July 2011), 348–383. https://doi.org/10.1016/j.destud.2011.01.002

[57] Hessamoddin Sarooghi, Dirk Libaers, and Andrew Burkemper. 2015. Examining the relationship between creativity and innovation: A meta-analysis of organizational, cultural, and environmental factors. *Journal of Business Venturing* 30, 5 (Sept. 2015), 714–731. https://doi.org/10.1016/j.jbusvent.2014.12.003

[58] Semyon Savransky. 2000. *Engineering of Creativity: Introduction to TRIZ Methodology of Inventive Problem Solving*. CRC Press, Boca Raton, USA.

[59] Jami J. Shah, Steve M. Smith, and Noe Vargas-Hernandez. 2003. Metrics for measuring ideation effectiveness. *Design Studies* 24, 2 (March 2003), 111–134. https://doi.org/10.1016/S0142-694X(02)00034-0

[60] Ben Shneiderman, Gerhard Fischer, Mary Czerwinski, Mitch Resnick, Brad Myers, Linda Candy, Ernest Edmonds, Mike Eisenberg, Elisa Giaccardi, Tom Hewett, Pamela Jennings, Bill Kules, Kumiyo Nakakoji, Jay Nunamaker, Randy Pausch, Ted Selker, Elisabeth Sylvan, and Michael Terry. 2006. Creativity support tools: Report from a US National Science Foundation sponsored workshop. *International Journal of Human-Computer Interaction* 20, 2 (May 2006), 61–77. https://doi.org/10.1207/s15327590ijhc2002_1

[61] Yogesh Kumar Singh. 2006. *Fundamental of Research Methodology and Statistics*. Newage International (P) Ltd. Publisher, New Delhi, India.

[62] Robert J. Sternberg and Todd I. Lubart. 1999. The concept of creativity: Prospects and paradigms. In *Handbook of creativity*, Robert J. Sternberg (Ed.). Cambridge University Press, Cambridge, UK, Chapter 1, 3–15. https://doi.org/10.1017/CBO9780511807916.003

[63] Tim Tack and Juergen Jasperneite. 2013. Application specific Automation Devices-the App-based approach. In *4th Annual Colloquium Communication in Automation* (Magdeburg, Germany, November 13–14) *(KommA '13)*. 1–10. https://www.researchgate.net/publication/257403049_Application_specific_Automation_Devices_-_the_App-based_approach

[64] Ellis Paul Torrance. 1966. *Torrance tests of creative thinking: Norms technical manual (Research Ed.)*. Personnel Press, Princeton, USA.

[65] Ellis Paul Torrance. 1988. The nature of creativity as manifest in its testing. In *The nature of creativity: Contemporary psychological perspectives*, Robert J. Sternberg (Ed.). Cambridge University Press, New York, USA, Chapter 2, 43–75.

[66] Ellis Paul Torrance. 1998. *The Torrance Tests of Creative Thinking: Directions manual Figural (Streamlined) forms A & B.* Scholastic Testing Service, Bensenville, USA.

[67] United Nations. 1990. *Convention on the Rights of the Child.* Retrieved September 23, 2020 from https://www.ohchr.org/en/professionalinterest/pages/crc.aspx

[68] Ester van Laar, Alexander J. A. M. van Deursen, Jan A. G. M. van Dijk, and Jos de Haan. 2017. The relation between 21st-century skills and digital skills: A systematic literature review. *Computers in Human Behavior* 72, 2017 (July 2017), 577 – 588. https://doi.org/10.1016/j.chb.2017.03.010

[69] Jorick Vissers, Lode De Bot, and Bieke Zaman. 2013. MemoLine: evaluating long-term UX with children. In *Proceedings of the 12th International Conference on Interaction Design and Children* (New York, USA, June 24–27) *(IDC '13)*. ACM, New York, USA, 285–288. https://doi.org/10.1145/2485760.2485836

[70] Lev Semenovich Vygotsky. 2004. Imagination and Creativity in Childhood. *Journal of Russian and East European Psychology* 42, 1 (Jan. 2004), 7–97. https://doi.org/10.1080/10610405.2004.11059210

[71] Andrew Warr and Eamonn O'Neill. 2007. Tool Support for Creativity Using Externalizations. In *Proceedings of the 6th ACM SIGCHI Conference on Creativity & Cognition* (Washington, USA, June 13–15) *(C&C'07)*. ACM, New York, USA, 127–136. https://doi.org/10.1145/1254960.1254979

[72] Robert W. Weisberg. 2015. On the Usefulness of "Value" in the Definition of Creativity. *Creativity Research Journal* 27, 2 (June 2015), 111–124. https://doi.org/10.1080/10400419.2015.1030320

[73] Svetlana Yarosh, Iulian Radu, Seth Hunter, and Eric Rosenbaum. 2011. Examining values: an analysis of nine years of IDC research. In *Proceedings of the 10th International Conference on Interaction Design and Children* (Ann Arbor, USA, June 20–23) *(IDC '11)*. ACM, New York, USA, 136–144. https://doi.org/10.1145/1999030.1999046

[74] Bieke Zaman. 2009. Introducing a Pairwise Comparison Scale for UX Evaluations with Preschoolers. In *Human-Computer Interaction – INTERACT 2009. INTERACT 2009. Lecture Notes in Computer Science, vol 5727.*, Tom Gross, Jan Gulliksen, Paula Kotzé, Lars Oestreicher, Phillippe Palanque, Raquel Oliveira Prates, and Marco Winckler (Eds.). Springer, Berlin, Germany, 634–637. https://doi.org/10.1007/978-3-642-03658-3_68

[75] Bieke Zaman, Vero Vanden Abeele, and Dirk De Grooff. 2013. Measuring product liking in preschool children: An evaluation of the Smileyometer and This or That methods. *International Journal of Child-Computer Interaction* 1, 2 (May 2013), 61–70. https://doi.org/10.1016/j.ijcci.2012.12.001

INCLUDED STUDIES

S1: Zahra Al-Mousawi and Asmaa Alsumait. 2012. A Digital Storytelling Tool for Arab Children. In *Proceedings of the 14th International Conference on Information Integration and Web-Based Applications & Services* (Bali, Indonesia, December 3–5) (*IIWAS '12*). ACM, New York, USA, 26–35. https://doi.org/10.1145/2428736.2428746

S2: Patrícia Alves-Oliveira, Patrícia Arriaga, Ana Maria 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* (Stanford, USA, June 27–30) (*IDC '17*). ACM, New York, USA, 423–429. https://doi.org/10.1145/3078072.3084304

S3: 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 & Cognition* (Singapore, June 27–30) (*C&C '17*). ACM, New York, USA, 106–118. https://doi.org/10.1145/3059454.3059477

S4: Dana Atwood-Blaine, Audrey C. Rule, and Jordan Walker. 2019. Creative self-efficacy of children aged 9–14 in a science center using a situated Mobile game. *Thinking Skills and Creativity* 33, 2019 (Sept. 2019), 100580. https://doi.org/10.1016/j.tsc.2019.100580

S5: Daniele Bacciotti, Yuri Borgianni, and Federico Rotini. 2016. A CAD tool to support idea generation in the product planning phase. *Computer-Aided Design and Applications* 13, 4 (Jun. 2016), 490–502. https://doi.org/10.1080/16864360.2015.1131543

S6: Luca Benedetti, Holger Winnemöller, 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* (Honolulu, USA, October 5–8) (*UIST '14*). ACM, New York, USA, 419–428. https://doi.org/10.1145/2642918.2647415

S7: Xiang Cao, Siân 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 ACM Conference on Computer Supported Cooperative Work* (Savannah, USA, February 6–10) (*CSCW '10*). ACM, New York, USA, 251–260. https://doi.org/10.1145/1718918.1718967

S8: Alejandro Catala, Javier Jaen, Betsy van Dijk, and Sergi Jordà. 2012. Exploring Tabletops as an Effective Tool to Foster Creativity Traits. In *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction* (Kingston, Canada, February 19–22) (*TEI '12*). ACM, New York, USA, 143–150. https://doi.org/10.1145/2148131.2148163

S9: Chia-Ling Chang and Ding-Bang Luh. 2012. User as Designer: A Design Model of User Creativity Platforms. *Journal of Integrated Design and Process Science* 16, 4 (Jan. 2012), 19–30. https://doi.org/10.3233/jid-2012-0020

S10: Yu-Shan Chang, Yu-Hung Chien, Hung-Chang Lin, Mavis Yiching Chen, and Hui-Hsien Hsieh. 2016. Effects of 3D CAD applications on the design creativity of students with different representational abilities. *Computers in Human Behavior* 65, 2016 (Dec. 2016), 107–113. https://doi.org/10.1016/j.chb.2016.08.0240747-5632

S11: Yu-shan Chang, Mavis Yi-Ching Chen, Meng-Jung Chuang, and Chia-hui Chou. 2019. Improving creative self-efficacy and performance through computer-aided design application. *Thinking Skills and Creativity* 31, 2019 (Mar. 2019), 103–111. https://doi.org/10.1016/j.tsc.2018.11.007

S12: Hsin-Yun Chang, Hao-Chiang Lin, Ting-Ting Wu, and Yueh-Min Huang. 2019. The Influence of Interactive Art of Visual Music on the Creativity of Science and Engineering Students. In *Proceedings of the 2019 IEEE Global Engineering Education Conference* (Dubai, Canada, April 9–11) (*EDUCON '19*). IEEE, Piscataway, USA, 1087–1092. https://doi.org/10.1109/EDUCON.2019.8725233

S13: Julia Chatain, Olivier Bitter, Violaine Fayolle, Robert W. Sumner, and Stéphane Magnenat. 2019. A Creative Game Design and Programming App. In *Motion, Interaction and Games* (Newcastle upon Tyne, UK, October 28–30) (*MIG '19*). ACM, New York, USA, Article 6. https://doi.org/10.1145/3359566.3360056

S14: Erin Cherry and Celine Latulipe. 2014. Quantifying the Creativity Support of Digital Tools through the Creativity Support Index. *ACM Transactions on Computer-Human Interaction* 21, 4 (Jun. 2014), Article 21. https://doi.org/10.1145/2617588

S15: 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* (Sonoma, USA, March 7–10) (*IUI '16*). ACM, New York, USA, 196–207. https://doi.org/10.1145/2856767.2856795

S16: Steven P. Dow, Alana Glassco, Jonathan Kass, Melissa Schwarz, Daniel L. Schwartz, and Scott R. Klemmer. 2010. Parallel Prototyping Leads to Better Design Results, More Divergence, and Increased Self-Efficacy. *ACM Transactions on Computer-Human Interaction* 17, 4 (Dec. 2010), Article 18. https://doi.org/10.1145/1879831.1879836

S17: 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* (Atlanta, USA, April 10–15) (*CHI '10*). ACM, New York, USA, 1233–1236. https://doi.org/10.1145/1753326.1753511

S18: Michael P. Downton. 2015. The aesthetics, creativity and craftsmanship of fourth graders' compositions. *Journal of Music, Technology & Education* 8, 3 (Dec. 2015), 273–287. https://doi.org/10.1386/jmte.8.3.273_1

S19: Seth M. Feeman, Landon B. Wright, and John L. Salmon. 2018. Exploration and evaluation of CAD modeling in virtual reality. *Computer-Aided Design and Applications* 15, 6 (Apr. 2018), 892–904. https://doi.org/10.1080/16864360.2018.1462570

S20: Georgios Fessakis, Dimitrios Lappas, and Elisavet E. Mavroudi. 2015. Could Computer Games-based Problem Solving Positively Affect the Development of Creativity in Young Children? A Mixed Method Case Study. In *Young Children and Families in the Information Age: Applications of Technology in Early Childhood*, Kelly L. Heider and Mary Renck Jalongo (Eds.). Springer Netherlands, Dordrecht, Netherlands, Chapter 12, 207–225. https://doi.org/10.1007/978-94-017-9184-7_12

S21: Sylvain Fleury, Aurélien Agnès, Rishi Vanukuru, Emma Goumillout, Nicolas Delcombel, and Simon Richir. 2020. Studying the effects of visual movement on creativity. *Thinking Skills and Creativity* 36, 2020 (Jun. 2020), 100661. https://doi.org/10.1016/j.tsc.2020.100661

S22: Eugenio Gaeta, María Eugenia Beltrán-Jaunsaras, Gloria Cea, Bernadette Spieler, Andrew Burton, Rebeca Isabel García-Betances, María Fernanda Cabrera-Umpiérrez, David Brown, Helen Boulton, and María T. Arredondo Waldmeyer. 2019. Evaluation of the Create@School Game-Based Learning–Teaching Approach. *Sensors* 19, 15 (Jun. 2019), 3251. https://doi.org/10.3390/s19153251

S23: Jérémie 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* (Vancouver, Canada, June 21–25) (*DIS '14*). ACM, New York, USA, 199–208. https://doi.org/10.1145/2598510.2598512

S24: Michail N. Giannakos and Letizia Jaccheri. 2018. From players to makers: An empirical examination of factors that affect creative game development. *International Journal of Child-Computer Interaction* 18, 2018 (Jun. 2018), 27–36. https://doi.org/10.1016/j.ijcci.2018.06.002

S25: Garth Griffin and Robert Jacob. 2013. Priming Creativity through Improvisation on an Adaptive Musical Instrumen. In *Proceedings of the 9th ACM Conference on Creativity & Cognition* (Sydney, Australia, June 17–20) (*C&C '13*). ACM, New York, USA, 146–155. https://doi.org/10.1145/2466627.2466630

S26: Shuchi Grover, Satabdi Basu, and Patricia Schank. 2018. What We Can Learn About Student Learning From Open-Ended Programming Projects in Middle School Computer Science. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education* (Baltimore, USA, February 21–24) (*SIGCSE '18*). ACM, New York, USA, 999–1004. https://doi.org/10.1145/3159450.3159522

S27: Jérôme Guegan, Stéphanie Buisine, Fabrice Mantelet, Nicolas Maranzana, and Frédéric Segonds. 2016. Avatar-mediated creativity: When embodying inventors makes engineers more creative. *Computers in Human Behavior* 61, 2016 (Aug. 2016), 165–175. https://doi.org/10.1016/j.chb.2016.03.024

S28: Jérôme Guegan, Frédéric Segonds, Jessy Barré, Nicolas Maranzana, Fabrice Mantelet, and Stéphanie Buisine. 2017. Social identity cues to improve creativity and identification in face-to-face and virtual groups. *Computers in Human Behavior* 77, 2017 (Dec. 2017), 140–147. https://doi.org/10.1016/j.chb.2017.08.043

S29: Florian Güldenpfennig, Wolfgang Reitberger, and Geraldine Fitzpatrick. 2012. Capturing Rich Media through Media Objects on Smartphones. In *Proceedings of the 24th Australian Computer-Human Interaction Conference* (Melbourne, Australia, November 26–30) (*OzCHI '12*). ACM, New York, USA, 180–183. https://doi.org/10.1145/2414536.2414569

S30: 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* (Vancouver, Canada, May 7–12) (*CHI '11*). ACM, New York, USA, 557–560. https://doi.org/10.1145/1978942.1979020

S31: Adam Haar Horowitz, Ishaan Grover, Pedro Reynolds-Cuéllar, Cynthia Breazeal, and Pattie Maes. 2018. Dormio: Interfacing with Dreams. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Montréal, Canada, April 21–26) (*CHI '18*). ACM, New York, USA, 1–10. https://doi.org/10.1145/3170427.3188403

S32: 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* (Sonoma, USA, March 7–10) (*IUI '16*). ACM, New York, USA, 241–250. https://doi.org/10.1145/2856767.2856792

S33: 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 *The Eleventh ACM/IEEE International Conference on Human Robot Interaction* (Christchurch, New Zealand, March 7–10) (*HRI '16*). IEEE Press, Piscataway, USA, 173–180. https://doi.org/10.1109/HRI.2016.7451749

S34: Anna Kantosalo and Sirpa Riihiaho. 2019. Experience evaluations for human–computer co-creative processes – planning and conducting an evaluation in practice. *Connection Science* 31, 1 (Aug. 2019), 60–81. https://doi.org/10.1080/09540091.2018.1432566

S35: Ahmet Fatih Karakaya and Halime Demirkan. 2015. Collaborative digital environments to enhance the creativity of designers. *Computers in Human Behavior* 42, 2015 (Jan. 2015), 176–186. https://doi.org/10.1016/j.chb.2014.03.029

S36: Hafizoah Kassima, Howard H. Nicholas, and Wan Ng. 2014. Using a multimedia learning tool to improve creative performance. *Thinking Skills and Creativity* 13, 2014 (Sep. 2014), 9–19. https://doi.org/10.1016/j.tsc.2014.02.004

S37: Jun Kato, Tomoyasu Nakano, and Masataka Goto. 2015. TextAlive: Integrated Design Environment for Kinetic Typograph. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Crossings, Korea, April 18–23) (*CHI '15*). ACM, New York, USA, 3403–3412. https://doi.org/10.1145/2702123.2702140

S38: 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* (Vancouver, Canada, May 7–12) (*CHI '11*). ACM, New York, USA, 1283–1292. https://doi.org/10.1145/1978942.1979133

S39: 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 Transactions on Computer-Human Interaction* 21, 3 (Jun. 2014), Article 14. https://doi.org/10.1145/2591677

S40: Kyu Han Koh, Vicki Bennett, and Alexander Repenning. 2011. Computing Indicators of Creativity. In *Proceedings of the 8th ACM Conference on Creativity & Cognition* (Atlanta, USA, November 3–6) (*C&C '11*). ACM, New York, USA, 357–358. https://doi.org/10.1145/2069618.2069694

S41: Natalia Kucirkovaa and Mona Sakr. 2015. Child–father creative text-making at home with crayons, iPad collage & PC. *Thinking Skills and Creativity* 17, 2015 (Sep. 2015), 59–73. https://doi.org/10.1016/j.tsc.2015.05.003

S42: 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* (Vancouver, Canada, May 7–12) (*CHI '11*). ACM, New York, USA, 1919–1928. https://doi.org/10.1145/1978942.1979221

S43: Neil Maiden, Konstantinos Zachos, Amanda Brown, George Brock, Lars Nyre, Aleksander Nygård 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* (Montreal, Canada, April 21–26) (*CHI '18*). ACM, New York, USA, 1–11. https://doi.org/10.1145/3173574.3174049

S44: Nicolas Mangano, Alex Baker, Mitch Dempsey, Emily Navarro, and André van der Hoek. 2010. Software Design Sketching with Calico. In *Proceedings of the IEEE/ACM International Conference on Automated Software Engineering* (Antwerp, Belgium, September 20–24) (*ASE '10*). ACM, New York, USA, 23–32. https://doi.org/10.1145/1858996.1859003

S45: 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* (San Jose, USA, May 7–12) (*CHI '16*). ACM, New York, USA, 2295–2306. https://doi.org/10.1145/2858036.2858269

S46: Rudy McDaniel, Joseph R. Fanfarelli, and Robb Lindgren. 2017. Creative Content Management: Importance, Novelty, and Affect as Design Heuristics for Learning Management Systems. *IEEE Transactions on Professional Communication* 60, 2 (Jun. 2017), 183–200. https://doi.org/10.1109/TPC.2017.2656578

S47: Lawrence McGrath, Sabrina Bresciani, and Martin J. Eppler. 2016. We walk the line: Icons provisional appearances on virtual whiteboards trigger elaborative dialogue and creativity. *Computers in Human Behavior* 63, 2016 (Oct. 2016), 717–726. https://doi.org/10.1016/j.chb.2016.05.086

S48: Fatimah Mohamed, Nor Nafizah Mohd Noor, Raja Farhana, and Raja Khairuddin. 2020. Simple ideas creativity of pre-service teachers in constructing game materials for learning plant diversity of Malaysia. *Journal of Critical Reviews* 7, 6 (Apr. 2020), 511–515. https://doi.org/10.31838/jcr.07.06.94

S49: Brad A. Myers, Ashley Lai, Tam Minh Le, YoungSeok Yoon, Andrew Faulring, and Joel Brandt. 2015. Selective Undo Support for Painting Applications. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Crossings, Korea, April 18–23) (*CHI '15*). ACM, New York, USA, 4227–4236. https://doi.org/10.1145/2702123.2702543

S50: Fatemeh Habibian Naeini and Mona Masood. 2012. Effect of Educational Computer Games on Student Creativity. *Research Journal of Applied Sciences, Engineering and Technology* 4, 23 (Dec. 2012), 5280–5284. https://doi.org/10.1007/s10639-020-10216-1

S51: 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* (Baltimore, USA, February 15–19) (*CSCW '14*). ACM, New York, USA, 75–83. https://doi.org/10.1145/2531602.2531637

S52: Courtney Nelson, Bradley J. Brummel, Frank Grove, Noah Jorgenson, Sandip Sen, and Rose Gamble. 2010. Measuring creativity in software development. In *Proceedings of the International Conference on Computational Creativity* (Lisbon, Portugal, January 7–9) (*ICCC '10*). Association for Computational Creativity, Lisbon, Portugal, 205–214. https://doi.org/10.1.1.1056.712

S53: Marije Nouwena, Selina Schepers, Karen Mouws, Karin Slegers, Niek Kosten, and Pieter Duysburgh. 2016. Designing an educational music game: What if children were calling the tune?. *International Journal of Child-Computer Interaction* 9–10, 2016 (Dec. 2016), 20–32. https://doi.org/10.1016/j.ijcci.2016.10.001

S54: 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 Tenth International Conference on Tangible, Embedded, and Embodied Interaction* (Eindhoven, Netherlands, February 14–17) (*TEI '16*). ACM, New York, USA, 340–348. https://doi.org/10.1145/2839462.2839474

S55: Jessica Taylor Piotrowski and Laurian Meester. 2018. Can apps support creativity in middle childhood?. *Computers in Human Behavior* 85, 2018 (Aug. 2018), 23–33. https://doi.org/10.1016/j.chb.2018.03.030

S56: Theodosios Sapounidis, Stavros Demetriadis, Pantelis M. Papadopoulos, and Dimitrios Stamovlasis. 2019. Tangible and graphical programming with experienced children: A mixed methods analysis. *International Journal of Child-Computer Interaction* 19, 2019 (Mar. 2019), 67–78. https://doi.org/10.1016/j.ijcci.2018.12.001

S57: Lara Schmitt, StéPhanie Buisine, Jonathan Chaboissier, AméZiane Aoussat, and Vernier. 2012. Dynamic tabletop interfaces for increasing creativity. *Computers in Human Behavior* 28, 2012 (Sep. 2012), 1892–1901. https://doi.org/10.1016/j.chb.2012.05.007

S58: Maria Shugrina, Jingwan Lu, and Stephen Diverdi. 2017. Playful Palette: An Interactive Parametric Color Mixer for Artists. *ACM Transactions on Graphics* 36, 4 (Jul. 2017), Article 61. https://doi.org/10.1145/3072959.3073690

S59: 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 & Cognition* (Glasgow, UK, June 22–25) (*C&C '15*). ACM, New York, USA, 83–92. https://doi.org/10.1145/2757226.2757230

S60: Kam Star, Fotis Paraskevopoulos, Maria Taramigkou, Dimitris Apostolou, Marise Schot, and Gregoris N. Mentzas. 2015. A Playful Affinity Space for Creative Research. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity & Cognition* (Glasgow, UK, June 22–25) (*C&C '15*). ACM, New York, USA, 107–110. https://doi.org/10.1145/2757226.2757244

S61: SugHee Kim, KwangSik Chung, and HeonChang Yu. 2013. Enhancing Digital Fluency through a Training Program for Creative Problem Solving Using Computer Programming. *The Journal of Creative Behavior* 47, 3 (Sep. 2013), 171–199. https://doi.org/10.1002/jocb.30

S62: Cristina Sylla, Clara Coutinhoa, Pedro Branco, and Wolfgang Muller. 2015. Investigating the use of digital manipulatives for storytelling in pre-school. *International Journal of Child-Computer Interaction* 6, 2015 (Dec. 2015), 39–48. https://doi.org/10.1016/j.ijcci.2015.10.001

S63: Cristina Sylla, Íris Susana Pires Pereira, Eva Brooks, and Nelson Zagalo. 2018. *t-books*: A block interface for young children's narrative construction. *International Journal of Child-Computer Interaction* 18, 2018 (Nov. 2018), 59–66. https://doi.org/10.1016/j.ijcci.2018.07.002

S64: Sarah L. Thew, Alistair G. Sutcliffe, Oscar De Bruijn, John McNaught, Rob N Procter, Paul Jarvis, and Iain Buchan. 2011. Supporting Creativity and Appreciation of Uncertainty in Exploring Geo-coded Public Health Data. *Methods of Information in Medicine* 50, 2 (Jan. 2011), 158–165. https://doi.org/10.3414/ME09-01-0070

S65: Cesar Torres and Eric Paulos. 2015. MetaMorphe: Designing Expressive 3D Models for Digital Fabrication. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity & Cognition* (Glasgow, UK, June 22–25) (*C&C '15*). ACM, New York, USA, 73–82. https://doi.org/10.1145/2757226.2757235

S66: 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* (Brisbane, Australia, June 4–8) (*DIS '16*). ACM, New York, USA, 158–169. https://doi.org/10.1145/2901790.2901828

S67: Matthias Voigt, Ralf Plattfaut, Kevin Ortbach, Andrea Malsbender, Bjorn Niehaves, and Jorg Becker. 2013. Evaluating Business Modeling Tools from a Creativity Support System Perspective - Results from a Focus Group in the Software Development Industry. In *Proceedings of the 2013 Pacific Asia Conference on Information Systems* (Jeju Island, Korea, June 18–22) (*PACIS 13'*). Association for Information Systems, Atlanta, USA, Article 155. https://aisel.aisnet.org/pacis2013/155

S68: Alexander E. Voiskounsky, Tatiana D. Yermolova, Sergey R. Yagolkovskiy, and Valeria M. Khromova. 2017. Creativity in online gaming: Individual and dyadic performance in Minecraft. *Psychology in Russia: State of the Art* 10, 4 (Jan. 2017), 144–161. https://doi.org/10.11621/pir.2017.0413

S69: 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* (Savannah, USA, February 6–10) (*CSCW '10*). ACM, New York, USA, 103–106. https://doi.org/10.1145/1718918.1718938

S70: 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 Designing Interactive Systems Conference* (Hong Kong, China, June 9–13) (*DIS '18*). ACM, New York, USA, 1095–1106. https://doi.org/10.1145/3196709.3196791

S71: Yizhi Wang, Yue Gao, and Zhouhui Lian. 2020. Attribute2Font: Creating Fonts You Want From Attributes. *ACM Transactions on Graphics* 39, 4 (Jul. 2020), Article 69. https://doi.org/10.1145/3386569.3392456

S72: 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* (Cambridge, USA, January 24–27) (*TEI '10*). ACM, New York, USA, 5–12. https://doi.org/10.1145/1709886.1709890

S73: Maximilian Witt and Susanne Robra-Bissantz. 2012. Sparking Motivation and Creativity with "Online Ideation Games". In *INFORMATIK 2012*, Ursula Goltz, Marcus Magnor, Hans-Jurgen Appelrath, Herbert K. Matthies, Wolf-Tilo Balke and Lars Wolf (Eds.). Gesellschaft für Informatik e.V., Bonn, Germany, 1006–1023. https://subs.emis.de/LNI/Proceedings/Proceedings208/1006.pdf

S74: Alicja Wojtczuk and Nathalie Bonnardel. 2011. Designing and assessing everyday objects: Impact of externalisation tools and judges' backgrounds. *Interacting with Computers* 23, 2011 (Jul. 2011), 337–345. https://doi.org/10.1016/j.intcom.2011.05.004

S75: 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* (St. Andrews, Scotland, October 8–11) (*UIST '13*). ACM, New York, USA, 387–396. https://doi.org/10.1145/2501988.2502015

S76: 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 & Cognition* (Singapore, June 27–30) (*C&C '17*). ACM, New York, USA, 63–74. https://doi.org/10.1145/3059454.3059459

S77: Xiaozhe Yang, Lin Lin, Pei-Yu Cheng, Xue Yang, and Youqun Ren. 2019. Which EEG feedback works better for creativity performance in immersive virtual reality: The reminder or encouraging feedback?. *Computers in Human Behavior* 99, 2019 (Oct. 2019), 345–351. https://doi.org/10.1016/j.chb.2019.06.002

S78: 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 & Cognition* (Atlanta, USA, November 3–6) (*C&C '11*). ACM, New York, USA, 3–12. https://doi.org/10.1145/2069618.2069622

S79: 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* (Boston, USA, June 21–24) (*IDC '15*). ACM, New York, USA, 11–20. https://doi.org/10.1145/2771839.2771841

S80: 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* (San Jose, USA, May 7–12) (*CHI EA '16*). ACM, New York, USA, 1322–1328. https://doi.org/10.1145/2851581.2892418

S81: 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* (Vancouver, Canada, May 7–12) (*CHI '11*). ACM, New York, USA, 1393–1402. https://doi.org/10.1145/1978942.1979147

Appendix E

Conference Paper

Creativity as a Quality Factor for Children: A Case Study

No Author Given

No Institute Given

Abstract. Current practices in the software development process do not include specific needs to ensure the quality of child-specific software. Creativity is one of these needs as it is beneficial for children. This research investigated creativity as a quality factor in the software development process of software for children. We ran a case study from software company Company X which has a long track record of evaluating creativity during the children's software development process. We evaluated a tailored Creativity Evaluation Method at Company X. The evaluation indicates that the Creativity Evaluation Method had good usability, intuitiveness, and compatibility rate for Company X. Based on the results, we made five recommendations for integrating creativity evaluation in the software development process targeting children. Future research could further investigate these recommendations in other software companies and do deeper studies on integrating creativity as a quality factor in the software development process.

Keywords: Creativity · Creativity Support Tool · Creativity Evaluation Method · Children · Quality Factor · Evidence-Based Software Quality Engineering.

1 Introduction

Creativity originate from the Latin word *creō*, which means "to create, to make" [28, p. 97]. This quality contributes to individual personal development and improves skills in language, storytelling, problem-solving, programming, computer science, and play [9, 12, 15, 17].

Some software applications are supposed to improve the user's creativity [24]. Cherry and Latulipe define "any tool that can be used by people in the open-ended creation of new artifacts" as Creativity Support Tools (CSTs) [8, p. 2]. CSTs are difficult to assess because they have no single metrics [31]. According to Resnick [26], creativity depends greatly on human factors, which gives the software development process of CSTs atypical challenges and issues. Any software with children as target users should be aware of the need to foster children's creativity.

Current practice in the software development process with the software product quality model ISO/IEC 25010:2011 does not include specific needs for quality software for children [19]. Creativity is not a characteristic of the software product that the development team is aware of during the software development

process. Research of creativity in the software development process is focused on the creative ability of the developers [3]. In a review of the literature of Creativity Evaluation Methods (CEMs) of digital CSTs for children [14], we have not found research papers on how to integrate creativity aspects in the software development of software targeting children. Our perspective is that, with the current pace of development of software features, the whole software development team shall be aware and working actively on assuring that certain the children's needs and rights are addressed in the final product. Therefore, this research aims is to investigate how to include creativity as a quality factor in the software development process of software for children.

The remainder of this paper is organized as follows. Section 2 introduces the background of CEMs of CSTs for children as a basis for the approach tailored in this paper. Further, section 3 depicts the context of this case study regarding the collaboration with the software company Company X. Section 4 describes the research methods for collecting and analyzing data. Section 5 presents the results of the study. Lastly, section 6 discusses these findings and section 7 concludes the research.

2 Creativity Evaluation Methods of Creativity Support Tools for children

We have recently performed a systematic literature review (SLR) that investigated the current CEMs of digital CSTs for children [14]. The SLR included 81 peer-reviewed articles and conference papers from the last 10 years. Only one of these studies was in the software engineering area.

CEMs have been researched for decades [2]. According to Carroll and Latulipe [6], the CEMs could be divided into three basic approaches: psychophysical measures using biometrics, self-reporting by users, and external judges.

CSTs could be evaluated with different evaluation factors. For example, the CEMs Creativity Product Analysis Matrix and Creative Product Semantic Scale use external judges to evaluate the three factors *elaboration and synthesis* (i.e., beauty and aesthetic), *novelty* (i.e., intuitiveness and originality), and *resolution* (i.e., relevance, logic, and usefulness) [5, 7, 11]. Others have simplified these factors to the two evaluation factors *valuable* and *novelty* [16, 29, 30].

The Creativity Support Index (CSI) is a standardized quantitative self-reporting psychometric survey of CSTs developed by Cherry and Latulipe [8]. This CEM utilized the six evaluation factors *collaboration*, *enjoyment*, *exploration*, *expressiveness*, *immersion*, and *results satisfaction*. On the other hand, Kerne et al. [20] evaluated CSTs using self-reporting questionnaires and psychophysical observation of a user's development of new ideas, based on the evaluation factors *emergence*, *exposition*, *flexibility*, *fluency*, *novelty*, *relevance*, and *visual presentation*.

An example of a CEM designed for children is the self-reporting preference CEM called This or That. The child compares two different CSTs and selects the preferred one [34]. According to Guinard [13], the only reliable method for two to three-year-old children is such paired-preference testing.

Table 1: Distribution of factors, products, approaches (appr.), and methods for studies with children and adults as target age. Adapted from the review [14].

	Factors											Products								Appr.			Methods				
	Valuable	Novelty	Fluency	Enjoyment	User feeling	Collaboration	Expressiveness	Immersion	Flexibility	Interaction	Other	Idea generation	Model making	Storytelling	Game	Music	Drawing	Programming	Visualization	Writing	Psychophysical	Self-reporting	External judges	Observation	Questionnaire	Interview	Qualitative	Quantitative
Children	18	20	9	13	10	11	4	5	7	4	9	3	7	8	5	3	1	5	0	1	26	18	13	26	21	12	24	22
Adults	35	31	22	16	19	14	10	8	7	8	14	18	14	4	7	5	6	1	5	1	43	39	19	43	42	15	42	31

For older children, it can be considered preference ranking for more than two products and three- to nine-point acceptance scales [13]. Another well-used CEM for children is the self-reporting five-point acceptance scale Smileyometer, comprising smileys representing the perceived level of *enjoyment* [25].

The number of studies in the SLR with different evaluation factors, products, approaches, and methods distributed into the target age is presented in Table 1. Twelve of the studies had both children and adults as participants [14].

Table 1 indicates that the CEMs were performed in different contexts using various software products, approaches, data collection methods, and data analysis. This distribution was also different for the age groups children and adults.

The 10 most common factors that appeared in the SLR were *valuable*, *novelty*, *fluency*, *enjoyment*, *user feeling*, *collaboration*, *expressiveness*, *immersion*, *flexibility*, and *interaction* [14]. The definitions of these 10 evaluation factors are presented in Table 2.

Based on the findings, the SLR proposes four recommendations (Rs) for CEMs of digital CSTs for children [14]:

- R1) Define the goal of the evaluation;
- R2)Review and choose amongst the 10 evaluation factors *valuable*, *novelty*, *fluency*, *enjoyment*, *user feeling*, *collaboration*, *expressiveness*, *immersion*, *flexibility*, and *interaction*;
- R3) Start with a quantitative psychophysical measurement approach but consider expansion to qualitative methods;
- R4) Developing a tailored approach for evaluation of creativity.

Table 2: The definitions of the ten evaluation factors.

Factor	Definition
Valuable	The user's perceived usefulness of the software.
Novelty	The extent to which the user can create something unique with the software.
Fluency	The degree the user gets relevant ideas from the software.
Enjoyment	The pleasure the user gets from the software.
User feeling	The user's perception of the software.
Collaboration	The extent to which the user works with others on the software.
Expressiveness	The degree the software expresses the user's thoughts or feelings.
Immersion	The extent to which the user is absorbed into the software.
Flexibility	The degree the software can be adapted to suit new circumstances.
Interaction	How the user's movements affect the software.

3 The Case Study Context

The first three authors have collaborated with Company X since August 2020, which gave a good insight into the company. Company X is a Norwegian software company with long experience developing software for children. Together with their subsidiary company, Company X wants to improve their processes and products by testing, logging, and analyzing to target their audience's families more accurately. The software development team in Company X comprises 10 different employees. The team is very interdisciplinary with developers, designers, concept developers, and one business administrator. The fourth author is one of the designers who also has the role of a tester. The company is centered on creativity and has meetings about it during the whole development process. Company X is also heavily focused on iterations of developing prototypes and evaluate them. Their focus is not only on getting features out. There is also a constant discussion of the solution from the perspective of the children.

Examples of the CEM in Company X are presented in Table 3. The insight indicated that Company X already followed some of the recommendations in section 2 and has a good starting point. Company X has a clear goal of the CEM based on their design principles (R1). Table 3 indicates that Company X utilized some of the 10 evaluation factors in their CEM (R2). Only one of Company X's evaluation factors, *exploration*, was not included among these 10 factors. They were positive about using these concrete factors as a starting point. However, it is too much to use all the factors for a specific software, and the combination of these factors depends on the software, as Table 3 indicates. Company X has a psychophysical measurement approach with observations, but it is more qualitative than quantitative (R3). They also include interviews and questionnaires if necessary. They welcomed more use of interviews and questionnaires like Smileyometer and This or That. Moreover, Company X also needs to develop a tailored approach for evaluating creativity as they do not use a standardized CEM (R4).

To tailor the CEM of Company X, we focused on using the 10 factors more systematically and use more qualitative methods with the self-reporting questionnaire Smileyometer.

Table 3: The Creativity Evaluation Methods in the different documents of Ablemagic. (Appr. = Approaches)

	Factors											Products								Appr.			Methods						
	Valuable	Novelty	Fluency	Enjoyment	User feeling	Collaboration	Expressiveness	Immersion	Flexibility	Interaction	Other	Idea generation	Model making	Storytelling	Game	Music	Drawing	Programming	Visualization	Writing	Psychophysical	Self-reporting	External judges	Observation	Questionnaire	Interview	Qualitative	Quantitative	Adults	Children
Design principles																													
An observation guide																													
An observation note																													
	3	1	0	3	3	3	3	0	1	3	1	0	1	1	1	1	3	0	1	0	3	1	0	3	0	1	1	3	1	3

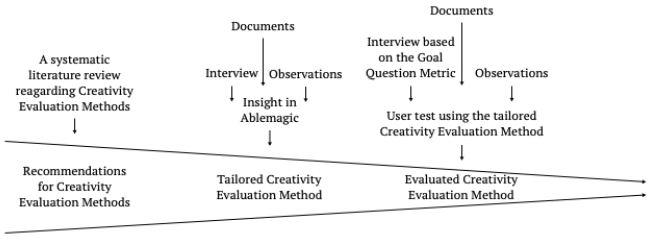


Fig. 1: The research process.

4 Research Methodology

The research process is presented in Fig. 2. The process went from unspecific recommendations for CEMs in the SLR [14] to develop a more specific tailored CEM for Company X based on the recommendations and insight. This development of the tailored CEM is explained in section 3. Lastly, the evaluation made the CEM even more specific. The study's research question was to investigate how to include creativity as a quality factor in the software development process of software for children by evaluating the tailored CEM of Company X. This evaluation was based on the Goal Question Metric (GQM), as it is an effective data collection method for evaluating the software development method [4]. The GQM for the evaluation of the tailored CEM is presented in Fig. 2. This measurement model comprises three levels:

- The conceptual level *Goal*: Defines a specific goal for an object;
- The operational level *Question*: A set of questions is used to characterize how the goal will be measured;
- The quantitative level *Metric*: A set of data associated with every question to answer it quantitatively.

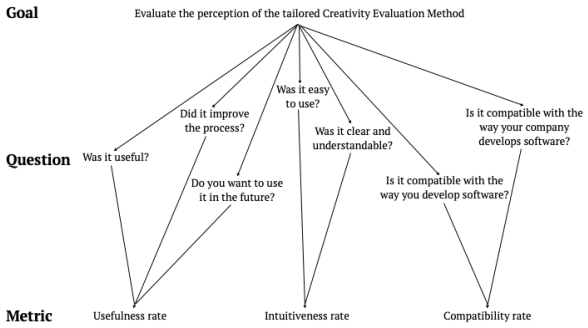


Fig. 2: Goal Question Metric of evaluation of the tailored Creativity Evaluation Method of Company X.

The questions and metrics were adapted from Riemenschneider, Hardgrave, and Davis's [27] examination of evaluating the individual acceptance of methodologies. Riemenschneider et al. [27] did this examination by investigating the five existing models of individual acceptance of information technology (IT) tools: Technology Acceptance Model (TAM), TAM2, Perceived Characteristics of Innovating (PCI), Theory of Planned Behavior (TPB), and Model of Personal Computer Utilization (MPCU).

TAM explains the adaption of IT apps in the organizational workplace [10], while TAM2 extended TAM to include voluntary and mandatory usage situations [33]. Further, PCI is an instrument measuring adaptation of IT innovations in personal workstations [21]. On the other hand, TPB is a theory linking beliefs to behavior with three primary determinants of intention [1], and MPCU refined TPB to predict personal computer utilization behavior [32].

These five models included different factors. For evaluating the tailored CEM, it is most relevant to measure the *usefulness*. According to Riemenschneider et al. [27], if a methodology should succeed, it must be regarded as useful (i.e., users perceived using it will enhance the job performance) by developers as all the models included usefulness. Further, the evaluation of the tailored CEM included the *intuitiveness* (ease of use) since all of the models without TPB included ease of use (i.e., the degree to which using it is perceived free of effort). Lastly, the evaluation of the tailored CEM included *compatibility* as Riemenschneider et al. [27] found that the compatibility of the methodology with how developers perform their work drives the methodology adoption intentions. Only PCI included compatibility (i.e., the extent to which it is perceived as consistent with the existing values, needs, and experience of potential adopters).

4.1 Data Collection Methods

To evaluate the quality of the tailored CEM of Company X, a combination of the data collection methods observation and interview were used. The interview was based on the GQM, while observation was used as a supplement because this data collection method reveals what people actually do instead of what they say they do when questioned [23].

Observation The main author observed a user test of a drawing software (software A) and a visualization software (software B) using the tailored CEM of Company X. The participants were 10 seven and eight years old children from the after-school program nearby. The leader of the after-school program was also allowed to try software A after a while. The user test took around one hour and 15 minutes. The author took the participant-observer role because it helped observe information through the experience as a participant in the user test following the tailored CEM. In that way, the author investigated how it was to use the CEM in practice. Besides, the author observed the software engineers while they conducted the user test. The data was collected using an observation template regarding the categories creativity, the software engineering team, the

communication, and other relevant things. The template was filled out by the main author after the test. The data sources also included the artifact of the evaluation template of the user test.

Interview A personal interview of the tester in Company X regarding the user test and the tailored CEM was conducted. The questions of the part about the tailored CEM was based on the GQM. The interview guide is attached in Appendix A. The duration of the interview was half an hour. The interviews were semi-structured, so the subject had the possibility to turn the conversation into themes the interviewee had not thought about. The video communications software Zoom 5.4.9[1] was used to record the interviews with the built-in recorder. Since the recording processed digital media files, a notification form was sent to the Norwegian Centre for Research Data (NSD) to get an ethical assessment of the project [22]. The tester also consented to participate in the interview. Lastly, the interviews were transcribed.

4.2 Data analysis

The study used primarily qualitative data analysis to answer the research question by exploring the tailored CEM of Company X in dept by summarizing, interpreting, and categorizing the collected data. The observation notes and the transcribed interview were analyzed inductively by coding themes with the specialist software package for qualitative analysis of textual data MAXQDA Analytics Pro 2020 (Release 20.2.1)[2] and the mind mapping software MindManager 13.0.181[3]. The observation was coded in an exploratory way with a total of three higher-order themes (the team, evaluation factors, and evaluation of the user test) and 26 sub-themes emerging from the observation. Besides, the CEMs of the different software were coded in the same way as as the documents in figure 3. On the other hand, the interview were categorized with a top-down approach based on the GQM. This resulted in four higher-order themes (the user test, usefulness rate, intuitiveness rate, and compatibility rate) and 31 sub-themes.

5 Results

5.1 Observation of the User Test

The atmosphere in Company X was a bit stressed before the start of the user test. Approximately half of the employees in Company X participated in the test. The tester observed and facilitated the test of software A. One of the designers collected the drawings to software A in the suitcase consecutively and scanned them in the office. A bachelor student and the main author observed software

[1] www.zoom.us
[2] www.maxqda.com
[3] www.mindmanager.com

Fig. 3: The user test setup of Software A.

A. One developer fixed the display of the scanned drawings on the screens. Another developer observed and facilitated software B, where the last developer and another bachelor student observed. The test setup of software A is presented in Fig. 3.

The CEM was based on the evaluation template of the user test. The distribution of evaluation factors, products, approaches, and methods in the two different software is presented in Table 4. The table depicted that the CEMs were quite similar for the two software. A Smileyometer comprised Company X's own figures was used to evaluate the factor *enjoyment*. The facilitator lay a poster with the Smileyometer on the table. At the same time, the children should point at the smiley representing their experience with the software. Based on this, the facilitator asked the children why they chose specific smileys. *Enjoyment* was also evaluated through different observation questions in the observation as the remainder of the factors.

Table 4: Distribution of the Creativity Evaluation Methods of the two different software for children in the user test. (Appr. = Approaches)

	Factors											Products									Appr.			Methods						
	Valuable	Novelty	Fluency	Enjoyment	User feeling	Collaboration	Expressiveness	Immersion	Flexibility	Interaction	Other	Idea generation	Model making	Storytelling	Game	Music	Drawing	Programming	Visualization	Writing	Psychophysical	Self-reporting	External judges	Observation	Questionnaire	Interview	Qualitative	Quantitative	Adults	Children
Software A																														
Software B																														
	2	0	2	2	2	1	1	0	0	0	0	0	0	0	0	0	1	0	1	0	2	2	0	2	2	2	2	2	1	2

Immediately after the test, the employees discussed the evaluation results in the office and on their Slack workspace. They were focused on how they should change the software based on the user test. After the children had left, the user test was discussed in plenary. They agreed the test was successful and that Smileyometer worked in their CEM of software for children. The main author also asked the two bachelor students how they experienced using the observation guide. They found it beneficial to have something as a starting point so one does not use a lot of time focusing on other things than you should.

5.2 Interview of the Evaluation of the CEM

The tester was interviewed regarding the evaluation of the CEM. The first part was about the user test, and the second part was about the experience of using the CEM.

The User Test Company X was satisfied with the user test of software A, according to the tester. The evaluation revealed what worked in the software and tested what Company X wanted, whether the animations made the software magical. It was mostly about being surprised by how the children reacted. The test was not located in a correct test environment, but it was still a bit realistic with other software the children could try after the drawing. This shows the balance between development and evaluation. The tester reflected on this balance: "*But it just has to be that way. I think it's best in this order because then you kind of avoid developing something that does not work at all*". The next evaluation of software A is about polishing the software, as claimed by the tester. They could evaluate animations and equipment internally, but they want to evaluate the experience flow of the software with the target group. Then, the prototype should be more executed, and they want a family workshop with a wider age group (four to 12 years old). As they do not yet have access to the correct test environment, this evaluation will, as the previous evaluation, be conducted in the test location.

Usefulness Rate Based on the user test planning meeting, the tester made an evaluation template containing the 10 evaluation factors, the goal of the evaluation, and the software's goal. One should for each evaluation determine the most important factors and further use these selected factors to make observation questions. There is no need to invent a new test frame for every evaluation. It is important and time-saving to have something standardized. Later, one could expand to conduct interviews based on the selected factors. The tester stated that with the use of the evaluation template, "*we knew for sure that we have somehow elaborated everything, all factors, and this is what is important to focus on in this test*". The observation questions reflected their preliminary thoughts. The tester pinpointed it was good assurance that the template was academically grounded and not just experience-based. It is helpful on how to test, evaluate and make progress. The tester concluded that the Smileyometer worked excellent to evaluate enjoyment after the children have used the software. On the other hand, the test exposed one must be very visual and concrete with the age group of seven to eight years old because they did not always understand what they should evaluate with the Smileyometer. Thus, the tester concluded: "*So I do not want to say the evaluation [the questionnaire and the interview] is weighted so highly with this age group. It is more the observation and such - what they are talking about, what they answer, and, yes*".

Intuitiveness Rate The tester experienced the CEM easy to use. The tester underlined it was easy to transfer to Company X's own design principles, some-

thing the tester used in the process of understanding it. The tester reflected that some of the factors did not fit the software evaluated in the user test: "*Some factors may not have been so immediate or fit so well, but they may fit later*".

Compatibility Rate The CEM was compatible with the way the tester evaluated software. The tester pinpointed: "*It is a method that I would not be able to use without working with it myself*". The user must understand the method; one could not only follow a guide. The tester experienced it good to go back to basics. The CEM was also compatible with the way Company X worked. The tester said it should be presented at the user test evaluation meeting and make part of what they call *the method of Company X*. The tester highlighted that it is important to stop and evaluate the method, and the whole team must be involved. The tester continued: "*So that, maybe I should set it up as a point, that I should bring it up, after all this. We'll see if time allows*".

6 Discussion

6.1 How to integrate evaluation of creativity in the software development process?

Company X is a software company aware of creativity as a quality factor in software for children. The company has regular creativity meetings during the software development process and targets integrating the quality creativity in their products. This could have made it easier for Company X to utilize the tailored CEM as the employees saw its value. Thus, making software companies aware of creativity as a quality factor is an important part of integrating evaluation of creativity in the software development process.

The tailored CEM of Company X using a mix of theory from the SLR [14] and experience from earlier work seemed useful, intuitive, and compatible. The observation of the user test indicated that evaluation of creativity helped find limitations of the prototype. In the interview, the tester expressed an intuitive transformation of the tailored CEM to the design principles of Company X originating from experience. The tester also highlighted the importance of the evaluation factors building on theory, securing a thoughtful evaluation.

It was very promising that the tester made an evaluation template based on the recommendations from the SLR [14]. The bachelor students liked to have a template as a basis for the observation of the user test. In the interview, the tester highlighted the importance of working with the method. This is supported by the fact that employee involvement is essential in change management [18]. Furthermore, this indicates the value of using the specific company's software development process as a starting point and build on that process to integrate creativity as a quality factor.

The process of using a tailored CEM of Company X is time-consuming. For example, This or That was not utilized in the user test even though the tester was aware of the benefits of the questionnaires in the same way as the Smileyometer.

It is wise not to do all the changes concurrently and use the time this creativity integration needs.

Thus, we have the following recommendations for integrating evaluation of creativity in the software development process targeting children:

- 1) Make software companies aware of creativity as a quality factor;
- 2) Analyze the specific company's process to find points to improve;
- 3) Use a mix of theory- and experience-based CEM;
- 4) Involve employees in changing the software development process;
- 5) Use the time this integration needs.

6.2 Limitations

This study is a case study of the software company Company X, a company that already was aware of creativity as a quality factor in software for children. It could be very different in other companies, and the findings could not be generalized to software companies in general. It had been advantageous to test this in several companies, but it was time-consuming to get into a company. Nevertheless, our recommendations could be transferable to other companies.

Another limitation is the fact that it was only one interviewee. It had been beneficial with more, but only this person had enough knowledge of the CEM at that time. On the other hand, several in the software development team was observed during the user test. The interviewee is also one of the authors, but this person has not participated in the data analysis. Besides, it is good to have a person to check if the information regarding Company X was correct.

7 Conclusion

The results indicated that the tailored CEM proposed in this research had good usability, intuitiveness, and compatibility rate for Company X. The tester in Company X indicated to adopt the evaluation template and discuss it with the rest of the team for further tests in Company X. Based on this research we made five recommendations for integrating creativity evaluation in the software development process targeting children: 1) Make software companies aware of creativity as a quality factor; 2) Analyze the specific company's process to find points to improve; 3) Use a mix of theory- and experience-based CEM; 4) Involve employees in changing the software development process; 5) Use the time this integration needs.

This research contributes to helping practitioners in software companies, improving their software developing process towards more systematic evaluation of creativity as a quality factor. This research has started the investigation of creativity as a quality, a topic that software engineer researchers shall investigate further. Integrating creativity as a quality factor can lead to better software for children and improve children's creative skills. Future research could build on our recommendations to build further knowledge about how to integrate creativity as a quality factor in agile software development.

A Appendix: Interview Guide

A.1 User test

- How do you feel regarding the user test?
- Did you notice something special before the test?
- Did you notice something special after the test?
- How was it to use the Smileyometer?
- What would you do differently next time?
- What is the further plan based on the results from the user test?
- Do you consider using more interviews or questionnaires next time?

A.2 Evaluation of the CEM

Usefulness rate

- Was it useful? Why/why not?
- Did it improve the process? Can you explain?
- Do you want to use it in the future? Do you have any comments on your answers?

Intuitiveness rate

- Was it easy to use? What do you think makes it easy/not easy to use? How was your process of trying to understand it?
- Was it clear and understandable? If no: What was not clear and/or understandable?
- Was the use of it frustrating at some point? If yes: Which parts?

Compatibility rate

- Is it compatible with the way you develop software? Why/why not?
- Is it compatible with the way your company develops software? Can you explain?

Other Comments

- Do you have any other comments? Would you like to talk about other advantages or challenges of using this method? Or what you think should be improved?

References

1. Ajzen, I.: From Intentions to Actions: A Theory of Planned Behavior. In: Kuhl, J., Beckmann, J. (eds.) Action Control, chapter 1, pp. 11–39. Springer, Berlin, Germany (1985). https://doi.org/10.1007/978-3-642-69746-3_2
2. Albert, R. S., Runco, M. A.: A history of research on creativity. In: Sternberg, R. J., Lubart, T. I. (eds.) Handbook of creativity, chapter 2, pp. 16–32. Cambridge University Press, Cambridge, UK (1998). https://doi.org/10.1017/CBO9780511807916.004

3. de la Barra, C. L., Crawford, B., Soto, R., Misra, S., Monfroy, E.: Agile Software Development: It Is about Knowledge Management and Creativity. In: Murgante B., Misra, S., Carlini, M. Torre, C. M., Nguyen, H.-Q. Taniar, D., Apduhan, B. O., Gervasi, O. (eds) Computational Science and Its Applications – ICCSA 2013. Lecture Notes in Computer Science, vol. 7973. Springer, pp. 98–113. Berlin, Germany (2013)
4. Basili, V. R., Weiss, D. M.: A Methodology for Collecting Valid Software Engineering Data. IEEE Transactions on Software Engineering **SE-10**(6), 728–738 (1984)
5. Besemer, S. P., Treffinger, D. J.: Analysis of creative products: Review and synthesis. The Journal of Creative Behavior **15**(3), 158–178 (1981)
6. Carroll, E. A., Latulipe, C.: Triangulating the Personal Creative Experience: Self-report, External Judgments, and Physiology. In: Proceedings of the Graphics Interface Conference 2012, pp. 53–60. Canadian Information Processing Society, Mississauga, Canada (2012)
7. Chang, Y.: 3D-CAD effects on creative design performance of different spatial abilities students. Journal of Computer Assisted Learning **30**(5), 397–407 (2014)
8. Cherry, E., Latulipe, C.: Quantifying the Creativity Support of Digital Tools through the Creativity Support Index. ACM Transactions on Computer-Human Interaction **21**(4), Article 21 (2014)
9. Csikszentmihalyi, M.: Creativity: Flow and the Psychology of Discovery and Invention. 1st edn. Harper Perennial, New York, USA (1997)
10. Davis, F. D.: Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly **13**(3), 319–340 (1989)
11. Gaggioli, A., Mazzoni, E., Milani, L., Riva, G.: The creative link: Investigating the relationship between social network indices, creative performance and flow in blended teams. Computers in Human Behavior **42**(2015), 157–166 (2015)
12. Giannakos, M. N. and Jaccheri, L. and Proto, R.: Teaching Computer Science to Young Children Through Creativity: Lessons Learned from the Case of Norway. In: Proceedings of the 3rd Computer Science Education Research Conference on Computer Science Education Research, pp. 103-111. Open Universiteit, Heerlen, Netherland (2013)
13. Guinard, J.-X.: Sensory and consumer testing with children. Trends in Food Science & Technology **11**(8), 273–283 (2000)
14. No Author Given: Evaluation of Creativity Support Tools for Children: A Systematic Literature Review. Submitted to Computing Surveys (2021)
15. Holmes, R. M., Gardner, B., Kohm, K., Bant, C., Ciminello, A., Moedt, K., Romeo, L.: The relationship between young children's language abilities, creativity, play, and storytelling. Early Child Development and Care **189**(2), 244–254 (2019)
16. Horn, D., Salvendy, G.: Consumer-based assessment of product creativity: A review and reappraisal.Human Factors and Ergonomics in Manufacturing & Service Industries **16**(2), 155–175 (2006)
17. Hsiao, H.-S., Wong, K.-H., Wang, M.-J., Yu, K.-C., Chang, K.-E., Sung, Y.-T.: Using Cognitive Affective Interaction Model to Construct On-Line Game for Creativity. In: First International Conference on Technologies for E-Learning and Digital Entertainment, pp. 409–418. Technologies for E-Learning and Digital Entertainment, Hangzhou, China (2006)
18. Hussain, S. T., Lei, S., Akram, T., Haider, M. J., Hussain, S. H., Ali, M.: Kurt Lewin's change model: A critical review of the role of leadership and employee involvement in organizational change. Journal of Innovation & Knowledge **3**(3), 123–127 (2018)

19. International Organization for Standardization: ISO/IEC 25010: 2011 - Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – System and software quality models. Geneva, Switzerland (2011)
20. Kerne, A., Webb, A. M., Smith, S. M., Linder, R., Lupfer, N., Qu, Y., Moeller, J., Damaraju, S.: Using Metrics of Curation to Evaluate Information-Based Ideation. ACM Transactions Computer-Human Interactions **21**(3), Article 14 (2014)
21. Moore, G. C., Benbasat, I.: Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation. Information Systems Research **2**(3), 192–222 (1991)
22. Notification Form for projects processing personal data, NSD, https://nsd.no/personvernombud/en/notify/meldeskjema?eng. Last accessed 28 Sep 2020
23. Oates, B. J.: Researching information systems and computing. 1st edn. SAGE Publications Ltd., London, UK (2006)
24. Papavlasopoulou, S., Giannakos, M. N., Jaccheri, L.: Empirical studies on the maker movement, a promising approach to learning: A literature review. Entertainment Computing **18**(1), 57–78 (2017)
25. Read, J. C., MacFarlane, S.: Using the Fun Toolkit and Other Survey Methods to Gather Opinions in Child Computer Interaction. In: Proceedings of the 2006 Conference on Interaction Design and Children, pp. 81–88. ACM Press, New York, USA (2006)
26. Resnick, M.: Seven issues for creativity support tool researchers. In: Shneiderman, B., Fischer, G., Czerwinski, M., Myers, B., Resnick, M. (eds.) National science foundation Workshop Report on Creativity Support Tools, chapter 7, pp. 67–70 (2005)
27. Riemenschneider, C. K., Hardgrave, B. C., Davis, F. D.: Explaining Software Developer Acceptance of Methodologies: A Comparison of Five Theoretical Models. IEEE Transactions on Software Engineering **28**(12), 1135–1145 (2002)
28. Ritter S. M., Rietzschel E. F.: Lay Theories of Creativity. In: Zedelius, C. M., Müller, B. C. N., Schooler, J. W. (eds.) The Science of Lay Theories, chapter 7, pp. 95–126. Springer, Cham, Germany (2004). https://doi.org/10.1007/978-3-319-57306-9_5
29. Sarkar, P., Chakrabarti, A.: Assessing design creativity. Design Studies **32**(4), 348–383 (2011)
30. Sarooghi, H., Libaers, D., Burkemper, A.: Examining the relationship between creativity and innovation: A meta-analysis of organizational, cultural, and environmental factors. Journal of Business Venturing **30**(5), 714–731 (2015)
31. Shneiderman, B., Fischer, G., Czerwinski, M., Resnick, M., Myers, B., Candy, L., Edmonds, E., Eisenberg, M., Giaccardi, E., Hewett, T., Jennings, P., Kules, B., Nakakoji, K., Nunamaker, J., Pausch, R., Selker, T., Sylvan, E., Terry, M.: Personal Creativity support tools: Report from a US National Science Foundation sponsored workshop. International Journal of Human-Computer Interaction **20**(2), 61–77 (2006)
32. Thompson, R. L., Higgins, C. A., Howell, J. M.: Personal Computing: Toward a Conceptual Model of Utilization. MIS Quarterly **15**(1), 125–143 (1991)
33. Venkatesh, V., Davis, F.: A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. Management Science **46**(2), 186–204 (2000)
34. Zaman, B., Vanden Abeele, V., De Grooff, D.: Measuring product liking in preschool children: An evaluation of the Smileyometer and This or That methods. International Journal of Child-Computer Interaction **1**(2), 61–70 (2013)

Marte Hoff Hagen

Creativity as a Quality Factor in Software for Children

NTNU
Norwegian University of
Science and Technology

ablenagic