Contents lists available at ScienceDirect

International Journal of Child-Computer Interaction

Review article

Evaluating digital creativity support for children: A systematic literature review

Marte Hoff Hagen [a],[*], Daniela Soares Cruzes [a], Letizia Jaccheri [a], Jerry Alan Fails [b]

[a] *Norwegian University of Science and Technology (NTNU), Høgskoleringen 1, Trondheim, 7491, Norway*
[b] *Boise State University (BSU), 1910 University Drive, Boise, 83725, ID, USA*

ARTICLE INFO

ABSTRACT

Creativity, the process of creating something new and valuable, benefits children by improving their skills and development, encouraging interaction and engagement, and enabling the generation and expression of novel ideas. In recent years, interactive digital tools have emerged to support the user's creativity in the open-ended creation of new artifacts. However, the question of evaluating the creativity happening in the interplay between children, digital tools, and products is still open. This systematic literature review investigated the evaluations of digital creativity support tools for children and identified 81 peer-reviewed relevant articles from the last 10 years. This research contributes to practitioners and researchers by providing an overview of the evaluations in a framework based on 10 factors (value, novelty, fluency, enjoyment, user feeling, collaboration, expressiveness, immersion, flexibility, and interaction), nine product areas, three approaches, and five methods. The review demonstrated that the evaluations differ widely, and the area lacks a standard evaluation framework. We propose the dimensions of our analysis as an initial framework for situating the evaluation of digital creativity support tools for children that the child–computer interaction community can further refine.

1. Introduction

Creativity is an essential element of human nature (Vidal, 2012), linked to innovation and growth in society (Amabile, 1996). Creativity is beneficial for children because it improves their problem-solving, computer science, programming, storytelling, play, and language skills and contributes to individual personal development (Csikszentmihalyi, 1997; Giannakos, Jaccheri, & Proto, 2013; Holmes et al., 2019; Hsiao et al., 2006). It is one of the seven core dimensions of digital skills for students of the 21st century and central in Article 31 of the United Nations Convention on the Rights of the Child regarding the right to play (United Nations, 1990; van Laar, van Deursen, van Dijk, & de Haan, 2017).

Varied research within the child–computer interaction (CCI) field builds on the theory of *constructionism* to provide children opportunities to learn by building, exploring, and sharing (Hourcade, 2015). CCI aims to actively involve children in designing technologies that contribute to their intellectual, social, and creative growth (Yarosh, Radu, Hunter, &

Rosenbaum, 2011). This involvement is beneficial because children contribute different ideas and perspectives than adult designers (Read, Fitton, & Horton, 2014). However, it can be challenging to collaborate with children in design processes because they can be afraid of talking (Iivari, Kinnula, & Kuure, 2014). The CCI research field continuously grows and evolves by linking various approaches, techniques, methodologies, end-user groups, technologies, and scientific disciplines, including creativity (Giannakos, Papamitsiou, Markopoulos, Read, & Hourcade, 2020). The field encompasses the entire age spectrum of children from 0 to 18 years old (Read & Markopoulos, 2013), which aligns with Article 1 in the United Nations Convention on the Rights of the Child defining a child as a "human being below the age of eighteen years unless under the law applicable to the child, majority is attained earlier" (United Nations, 1990).

Certain tools are intended to enhance people's creativity (Papavlasopoulou, Giannakos, & Jaccheri, 2017), called creativity support tools (CSTs). Cherry and Latulipe define a CST as "any tool that can be used by people in the open-ended creation of new artifacts" (Cherry & Latulipe,

2014 p. 2). CSTs are apps that are often used in the process of completing an artifact or used to create digital artifacts (Cherry & Latulipe, 2014). Examples include the photo editor program Adobe Photoshop, the drawing program Paint, and the visual programming language Scratch.

It is challenging to evaluate CSTs because no single metric is appropriate to measure them (Nakakoji, 2005). Nakakoji (2005) indicated that the evaluations of CSTs will have non-obvious challenges and issues for the traditional human–computer interaction (HCI) field. While much technology for children is often intended for entertaining rather than supporting creativity (Cassell & Ryokai, 2001), a significant amount of work focuses on building creativity support for children.

This article aims to explore the state-of-the-art evaluations of digital CSTs for children. To accomplish this, 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 (Kitchenham, 2004). Remy, Mac-Donald Vermeulen, Frich, Biskjaer, and Dalsgaard (2020) performed a recent and relevant in-depth literature review on evaluating CSTs but did not investigate the age of the target user groups in the studies they reviewed. Thus, there is a need for an SLR focusing on the evaluation of digital CSTs for children. The article investigates this knowledge gap in the intersection between creativity and CCI by aiming to answer the following research questions (RQs):

- **RQ 1:** How evaluations for creativity support tool (CST) for children are performed?
  - **RQ 1.1:** Which factors have been considered during the evaluations?
  - **RQ 1.2:** In which context are the evaluations performed?
  - **RQ 1.3:** What differences exist based on the age of the participants in the evaluations?
  - **RQ 1.4:** How credible are the evaluations?

2. Background

2.1. Creativity

According to Ritter and Rietzschel (2004), the English word *creativity* stems from the Latin word *creō*, which means "to create, to make" (Ritter & Rietzschel, 2004 p. 97). The term has evolved from psychology theory and is defined in many ways (Kaufman & Beghetto, 2009). Oxford Dictionary defines it as "the use of skill and imagination to produce something new or to produce art" (Oxford Dictionary, 2021). This definition is similar to the common definition of creativity as the ability to generate ideas (often as a solution to a problem) that are both *novel* (i. e., perceived unique or original) and *valuable* (i.e., perceived useful or functional) (Amabile, 1983; Runco & Jaeger, 2012; Sternberg & Lubart, 1999). 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" (Sarkar & Chakrabarti, 2008 p. 11).

By examining the meaning of the word creativity, Jordanous and Keller (2016) 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 (2015), *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" (Weisberg, 2015 p. 119). Another approach to creativity by Vygotsky (2004) is the process of constructing these novel ideas or behaviors by revising old memories. This approach is similar to Papert's (1980) learning theory *constructionism*, which creates mental models to acquire knowledge.

2.2. Measurements of the creative experience

According to Carroll and Latulipe (2012), the three basic approaches to creativity measurements that can be used to evaluate CSTs are *physiological measurement* using biometrics, user *self-reporting*, and *external judges*. An overview of the different creativity measurements discussed later in this section is presented in Table 1.

2.2.1. Physiological measurement

Physiological measurements focus on the users' interaction with the CSTs. They are perhaps the most objective creativity measurement with the least potential for bias, conducted by observers in the form of the physiological measurement with biometrics such as the user's behavior with the product. Guilfords' development of the physiological measurement Alternative Uses Task based on cognitive psychology initiated modern creativity research in 1950 (Guilford, 1950). The Alternative Uses Task measures creativity in the form of the *divergent thinking ability*, revealing how many alternative uses for an object a participant managed to mention within a specific amount of time (Guilford, 1967).

Torrance built on Guilford's theories 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*) (Torrance, 1988). These four factors could be measured by the physiological measurement Torrance Tests of Creative Thinking, a standardized measure of children's creativity (Torrance, 1998). The Torrance Tests of Creative Thinking consists of a verbal test with writing and a figural test with drawings, each with different subtests (Torrance, 1966). Based on the principle of the Torrance Tests of Creative Thinking, the abbreviated Torrance test for adults required a shorter test time (Goff & Torrance, 2002). However, the Torrance Tests of Creative Thinking has been criticized for being uncertain since the factor *elaboration* contributes substantially to the score despite being less important (Almeida, Leandro, Prieto Ferrando, Oliveira, & Ferrandiz, 2008).

A similar physiological measurement is the Multidimensional Stimulus Fluency Measure creativity test, which evaluates the factors *fluency* and *originality* (Moran, Milgram, Sawyers, & Fu, 1983). It can be applied to children as young as four years old as it requires verbal responses and uses tactile and visual stimuli (Moran et al., 1983). On the other hand, the physiological measurement Remote Associates Test is a convergent creativity test where the participants are presented with three words and have to guess the word these words are semantically related to (Mednick, Mednick, and Mednick (1964).

2.2.2. Self-reporting

Self-reporting is considered a subjective creativity measurement since the user indicates their self-perceived notions of the experience. An example is the Creativity Achievement Questionnaire, a questionnaire of

Table 1
Overview of the approaches to the existing creativity measurements.

Physiological measurement	Self-reporting	External judges
● Alternative Uses Task	● Likert scale	● Theory of inventive problem solving
● Torrance Tests of Creative Thinking	● Creativity Achievement Questionnaire	● Consensual Assessment Technique
● Abbreviated Torrance test for adults	● Kaufman Domains of Creative Scale	● Creative Product Semantic Scale
● Multidimensional Stimulus Fluency Measure	● AttrakDiff	
● Remote Associate Test	● Fun-Sorter	
	● Memoline	
● Creativity Support Index		
● Level of creativity components		

creativity scores across 10 domains (writing, culinary arts, architecture, visual arts, dance, music, theater, scientific discovery, inventions, and humor) (Carson, Peterson, & Higgens, 2005). This creativity measurement is similar to the Kaufman Domains of Creativity Scale, a five-point Likert scale with 50 subdivisions in the following domains: everyday, academic, mechanical or science, artistic, and performance (Kaufman, 2012). The Likert scale is a self-reporting questionnaire with a set of statements where the participants are asked to evaluate the level of agreement on a scale from strongly disagree to strongly agree (Singh, 2006). Kaufman Domains of Creativity Scale has been proved to be a valid and reliable measurement for evaluating domain-specific creativity (McKay, Alexander, & Kaufman, 2017).

An example of the acceptance self-reporting creativity measurement for products is AttrakDiff, which evaluates both the pragmatic (*functionality*), the hedonic perception (*emotional experience*), and attractiveness (*aesthetic*) on a seven-point scale (Hassenzahl, Burmester, & Koller, 2003). This scale focus on where the product lies along a continuum of two opposites (e.g., *ordinary* and *novel*).

For measuring CSTs, Cherry and Latulipe (2014) developed the self-reporting standardized physiological measurement Creativity Support Index, designed to evaluate a CST's ability to assist a user engaged in creative work. This method measures six factors of creativity support: *collaboration*, *expressiveness*, *exploration*, *enjoyment*, *immersion*, and *results satisfaction*. The Creativity Support Index enables comparable and quantifiable results by providing a rating scale section of two agreement statements for each factor and a paired-factor comparison. A similar measurement by Kerne et al. (2014) shows how a CST supports creative engagement can be revealed by measuring the levels of creativity components – based on the factors *fluency*, *flexibility*, *novelty*, *emergence*, *relevance*, *visual presentation*, and *exposition* – using physiological observations of a person's development of new ideas combined with self-report post-evaluative questionnaires. Such measurements may be too complicated for children because the language and constructs used do not match children's literacy and cognition levels. For evaluating CSTs designed for children, it could be valuable to better understand several aspects from their perspective (Giannakos & Jaccheri, 2013). There exist self-reported measurements explicitly designed for children that are used to evaluate CSTs but do not necessarily measure the creative experience. Examples are the Fun-Sorter, assessing the level of the factor *fun* by ranking statements for a series of connected activities from the most to the least *fun* (Read, MacFarlane, & Casey, 2002), and the MemoLine, measuring long-term user experience by asking the child to color a timeline according to their experience with the product over a predefined timespan (Vissers, De Bot, & Zaman, 2013).

2.2.3. External judges

External judges are used to evaluate creativity by evaluating a user's creative outcome. External judges can be more objective than self-reporting because of the observational approach taken by people outside of the direct interactive experience. However, these judgments are subject to personal biases of the judge. The systematic human-oriented knowledge-based method of inventive problem-solving called the theory of inventive problem solving is a creativity measurement that uses external judges to compare the creative process with this method (Savransky, 2000).

The Consensual Assessment Technique uses external expert judges to evaluate products based on their subjective perception of creativity (Amabile, 1983). As reported by Hocevar (1981), an issue with external judges is their inability to distinguish between creativity, aesthetics, and technical skills. According to Glăveanu (2010), this judgment depends on the environment since members of different communities could assess the same artifact differently.

The evaluation rubrics in Creativity Product Analysis Matrix and Creative Product Semantic Scale use external judges to assess three factors: *novelty* (i.e., originality and intuitiveness), *resolution* (i.e., usefulness, logic, and relevance), and *elaboration and synthesis* (i.e.,

aesthetic and beauty) (Besemer & Treffinger, 1981; Chang, 2014; Gaggioli, Mazzoni, Milani, & Riva, 2015). Others have simplified these aspects to the two factors *novelty* and *value* (Horn & Salvendy, 2006; Sarkar & Chakrabarti, 2011; Sarooghi, Libaers, & Burkemper, 2015).

3. Review method

The SLR presented herein was conducted using the original guidelines for an SLR in software engineering provided by Kitchenham (2004), which has recently been adopted and used by other researchers in the CCI area (Bakala, Gerosa, Hourcade, & Tejera, 2021; Quayyum, Cruzes, & Jaccheri, 2021; Sharma & Giannakos, 2021; Theodoropoulos & Lepouras, 2021; Verbruggen, Depaepe, & Torbeyns, 2021). A review protocol was developed in the initial phase to plan the SLR (Kitchenham, 2004). The protocol contained the rationale for the survey, the RQs, 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 was conducted in five stages as illustrated in Fig. 1. The remainder of this section describes each step in detail.

3.1. Identification of research

The first step of conducting the SLR was the identification of research (Kitchenham, 2004). A pilot search was used to find the most optimal search string and databases to identify the evaluations of digital CSTs for children. This search also found the mentioned literature review by Remy et al. (2020), which reused the corpus of their earlier in-depth literature review (Frich, MacDonald Vermeulen, Remy, Biskjaer, & Dalsgaard, 2019) by selecting all papers in this sample that evaluated a CST. Frich et al. (2019) used the search string *"creativity" OR "creativity support tool"* in the *ACM Digital Library*. The sample size was further constrained to those with at least above-average citations of 0.669 per year since they were published (Frich et al., 2019). Remy et al. (2020) used the same method for papers after 2016 but constrained the sample size based on the above-average download counts per year (Remy et al., 2020). The corpus of 113 studies from this literature review was added to the automatic search.

3.1.1. Automatic search

The automatic literature search was conducted in October 2020. The search strategy included the electronic bibliographic databases *Scopus*, *ACM Digital Library*, and *IEEE Xplore* and a journal hand search in the *International Journal of Child Computer Interaction* to include as many relevant studies as possible. It was intentionally decided not to include the term *"children"* in the search string, despite the focus on children in RQ 1. This choice broadened the inclusion by exploring whether the evaluations of digital CSTs for adults could be transferable to children, particularly in helping to reveal any overarching evaluation framework. An overview of this initial identification of research is presented in Table 2, resulting in **593** hits. The bibliographic package EndNote X9.3.3[1] was used to manage this large number of references.

3.1.2. Manual search

To include additional papers not discovered by the limited search string in the automatic search, manual searches were conducted in November 2020 in CCI, HCI, and creativity journals that were identified in the automatic search. The documenting of the manual search is presented in Table 3, resulting in **2081** hits.

---

[1] endnote.com.

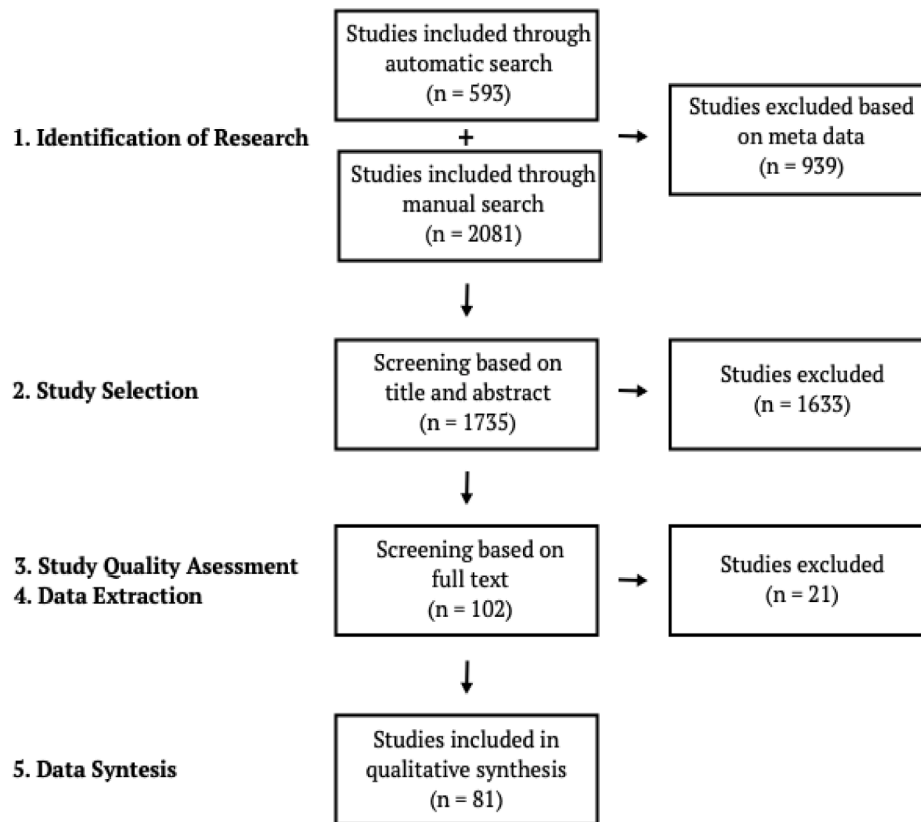**Fig. 1.** Study selection process.

Table 2
Overview of the automatic search.

Data source	Search string	Date	Hits
Recent and relevant literature review (Remy et al., 2020)	Based on *"creativity"* OR *"creativity support tool"* in the *ACM Digital Library* (Frich et al., 2019)	02.10.20 (pilot search)	113
Scopus	*(TITLE AND ABS)("creativity"*	09.10.20	259
ACM Digital Library	*AND ("measure" OR*	10.10.20	89
IEEE Xplore	*"measuring" OR "metrics" OR*	13.10.20	91
	"evaluate" OR "evaluating")		
	AND ("software" OR "app"))		
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**

Table 3
Documenting of the manual search.

Data source	Search string	Date	Hits
International Journal of Child–Computer Interaction	*"creativity"*	16.11.20	60
Interaction with Computers			310
ACM Transactions on Computer Human Interaction			94
AIS Transaction on Human–Computer Interaction			34
Computers in Human Behavior			552
Computer-Aided Design and Application			210
International Journal of Design Creativity and Innovation	*("software" OR "app")*	18.11.20	39
Journal of Creativity Behavior			124
Thinking Skills and Creativity			169
Digital Creativity			489
		Total:	**2081**

3.2. Study selection

The potentially relevant primary studies from the search were then assessed with exclusion criteria (EC) for their actual relevance (Kitchenham, 2004). These criteria are presented in Table 4, resulting in 2572 excluded studies, and **102** remained primary studies.

Due to the focus on current evaluations in RQ 1, EC 1 excluded non-recent studies. The year 2010 was selected as the lower time limit because this year represented the start of significant growth in the number of apps since Apple released its Software Development Kit in 2009 (Goggin, 2011). Following the method used in Kitchenham (2004), if a duplicate of a study was published more recently than the other duplicate(s), the most recent duplicate publication was kept in EC 2. Duplicates exclusion was conducted in the repository in Endnote. Even though Kitchenham (2004) recommends avoiding exclusion based on

Table 4
The exclusion criteria (EC).

	Exclusion criterion	Excluded studies	Screening type
EC 1	Publishing date before 2010	898	Meta data
EC 2	Duplicated studies	40	
EC 3	Non-English language	1	
EC 4	No evaluation of a creativity support tool	1633	Title and abstract
	Total:	**2572**	

the language, it was decided to include EC 3 as an exclusion criterion. The English abstract in the single non-English paper did not give as much detail as was required for the stages three to five in this study's selection process. EC 4 was based on the RQs, and the exclusion in this criterion was based on reading titles and abstracts.

3.3. Assessment of study quality

Further, quality assessments (QAs) were made to evaluate each primary study's quality based on the full-text reading (Kitchenham, 2004). Since this quality assessment step included reading the full text of the studies, it was combined with the next step of data extraction to save time. Each primary study had to pass three QA questions regarding relevance, rigor, and credibility given in Table 5, adapted from the checklist to Dybå and Dingsøyr (2008). Overall, the QAs excluded 21 studies, resulting in an inclusion of **81** studies in the qualitative data synthesis.

3.4. Data extraction

As proposed by Kitchenham (2004), we designed a systematic data extraction protocol to collect the required information from the primary studies to address the RQs and QAs. The data were extracted using a specialized software package for qualitative analysis of textual data MAXQDA Analytics Pro 2020 (Release 20.2.1)[2]. The extraction was done independently by the first author but discussed with the second and third authors to ensure it was done systematically. The data extraction form is presented in Table 6.

3.5. Data synthesis

Thematic synthesis was used to collect and summarize the results of the included primary studies to answer the RQs (Kitchenham, 2004). The five recommended steps for thematic synthesis proposed by Cruzes and Dybå (2011) 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 3.4. The following two steps were to translate the codes within each data extraction point into themes (e.g., Table 11 in Section 4) and later categorize these sub-themes into a model of higher-order themes (presented in Table 10) by exploring the relationships between the themes based on the RQs. For example, the higher-order theme *interaction* within the factors in Table 10 was derived from the two sub-themes *interaction* and *behavior* in Table 11. This process was also performed in MAXQDA Analytics Pro 2020 and later in the mind mapping software MindManager 13.0.181[3]. The last step was to assess the trustworthiness of the interpretation leading to the

Table 5
The quality assessments (QAs).

	Quality assessment	Excluded studies
QA 1	Does the study refer to a digital CST?	6
QA 2	Is the evaluation appropriate in terms of using at least one concrete factor to justify the evaluation of the digital CST?	15
QA 3	Do the presented findings have validity by clarifying how the evaluation was conducted?	0
	Total:	**21**

(CST = creativity support tool).

Table 6
Data extraction form.

Focus area	Data	Mapping to QAs	Mapping to RQs
Meta data	• Publication channel • Publication year		
Factors	• Factors evaluating the product	QA 2	RQ 1.1, RQ 1.3
Products	• Software type • Main functionality • Creativity outcome • Creativity topic • Characteristics improving creativity	QA 1	RQ 1.2

Methods	• Data gathering method
	• Data analysis
	• Participants:
	  - Sample size
	  - Gender distribution
	  - Age
	• Approach
	• Validity of method
	• Study environment
		QA 2, QA 3	RQ 1, RQ 1.2, RQ 1.3, RQ 1.4

Research	• Objective • RQs • Findings	QA 3	
Creativity	• Definition of creativity • Unclear words describing creativity	QA 3	RQ 1.4

(RQ = Research Question, QA = Quality Assessment).

synthesis by looking for counter-evidence.

4. Results

4.1. Meta data

To contextualize the results more clearly throughout this article, all included studies in the SLR are identified with different colors according to the age of the target user group. The referencing of the studies targeting *children* are **red** ($n = 21$), *adults* **blue** ($n = 46$), *both children and adults* **purple** ($n = 12$), and the studies *not specifying the age of the target group* **black** ($n = 2$). Fig. 2 illustrates the included studies' publication frequency between 2010 and 2020. The publication frequency for the included studies increased from 2013 until 2016, with 12 published studies at the peak and then decreased. Adult-focused studies follow this trend of all the included studies, while child-focused studies expand.

The distribution of the publication channels according to the publication years is presented in Table 7. Table 8 gives a more detailed presentation of the distribution of the different publication channels. Overall, there were more conference papers than journal articles. Many adult-focused studies were conference papers, while most child-focused studies were journal articles. In addition, Fessakis, Lappas, and Mavroudi (2015) published their research as a chapter in the book entitled *Young Children and Families in the Information Age* and Witt and Robra-Bissantz (2012) as a chapter in *Gesellschaft für Informatik e.V's lecture notes in informatics*.
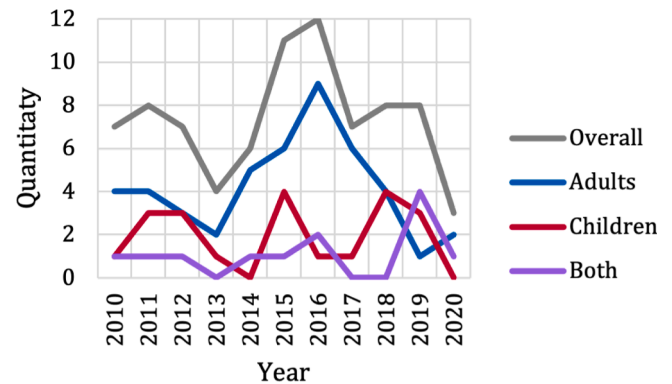
4.2. Evaluation factors

To answer the main RQ 1 regarding evaluations of digital CSTs for children, this section addresses the RQ 1.1 regarding the factors used to evaluate digital CSTs for both children and adults. A total of 53 factors were considered in the evaluations of the included studies. These were categorized into higher-order themes based on similarity. The categorizing yielded 11 different categories, including the 10 most frequently occurring categories presented in Table 9 that covered 90% of the

Table 7
Publication frequency.

Year	Journal	Conference	Book
2020	Fleury et al. (2020), Mohamed, Noor, Farhana, and Khairuddin (2020), Wang, Gao, and Lian (2020)		
2019	Atwood-Blaine, Rule, and Walker (2019), Kantosalo and Riihiaho (2019), Sapounidis, Demetriadis, Papadopoulos, and Stamovlasis (2019), Chang, Chen, Chuang and Chou (2019), Gaeta et al. (2019), Yang, Lin, Cheng, Yang, and Ren (2019)	Chatain, Bitter, Fayolle, Sumner, and Magnenat (2019), Chang, Lin, Wu and Huang (2019)	
2018	Giannakos and Jaccheri (2018), Piotrowski and Meester (2018), Sylla, Pereira, Brooks, and Zagalo (2018), Feeman, Wright, and Salmon (2018)	Grover, Basu, and Schank (2018), Haar Horowitz, Grover, Reynolds-Cuéllar, Breazeal, and Maes (2018), Maiden et al. (2018), Wang et al. (2018)	
2017	Guegan et al. (2017), McDaniel, Fanfarelli, and Lindgren (2017), Shugrina, Lu, and Diverdi (2017), Voiskounsky, Yermolova, Yagolkovskiy, and Khromova (2017)	Alves-Oliveira, Arriaga, Paiva, and Hoffman (2017), Andolina et al. (2017), Zheng, Do, and Budd (2017)	
2016	Nouwena et al. (2016), Chang, Chien, Lin, Chen, and Hsieh (2016), Bacciotti, Borgianni, and Rotini (2016), Guegan, Buisine, Mantelet, Maranzana, and Segonds (2016), McGrath, Bresciani, and Eppler (2016)	Oh et al. (2016), Davis, Hsiao, Yashraj Singh, Li, and Magerko (2016), Huang, Duvenaud, and Gajos (2016), Kahn et al. (2016), Martin, Gardner, Swift, and Martin (2016), Torres, Li, and Paulos (2016), Yoshida, Fukushima, Aida, and Naemura (2016)	
2015	Downton (2015), Sylla, Coutinhoa, Branco, and Muller (2015), Kucirkovaa and Sakr (2015), Karakaya and Demirkan (2015)	Yoon, Verma, Peppler, and Ramani (2015), Kato, Nakano, and Goto (2015), Myers et al. (2015), Siangliulue, Chan, Gajos, and Dow (2015), Star et al. (2015), Torres and Paulos (2015)	Fessakis et al. (2015)
2014	Cherry and Latulipe (2014), Kassima, Nicholas, and Ng (2014), Kerne et al. (2014)	Benedetti, Winnemöller, Corsini, and Scopigno (2014), Garcia, Tsandilas, Agon, and Mackay (2014), Nakazato et al. (2014)	
2013	Kim, Chung, and Yu (2013)	Voigt et al. (2013), Zhang et al. (2013), Griffin and Jacob (2013)	
2012	Chang and Luh (2012), Habibian Naeini and Masood (2012), Schmitt, Buisine, Chaboissier, Aoussat, and Vernier (2012)	Al-Mousawi and Alsumait (2012), Catala, Jaen, van Dijk, and Jordà (2012), Güldenpfennig, Reitberger, and Fitzpatrick (2012)	Witt and Robra-Bissantz (2012)
2011	Thew et al. (2011), Wojtczuk and Bonnardel (2011)	Koh, Bennett, and Repenning (2011), Lu et al. (2011), Chu Yew Yee, Quek, and Xiao (2011), Halpern et al. (2011), Kazi, Chua, Zhao, Davis, and Low (2011), Yu and Nickerson (2011)	

Table 7 (*continued*)

Year	Journal	Conference	Book
2010	Dow et al. (2010)	Cao, Lindley, Helmes, and Sellen (2010), Willis, Lin, Mitani, and Igarashi (2010), Bao, Gerber, Gergle, and Hoffman (2010), Mangano, Baker, Dempsey, Navarro, and van der Hoek (2010), Nelson et al. (2010), Wang, Cosley, and Fussell (2010)	
Sum:	36 (45%)	43 (53%)	2 (2%)

Fig. 2. Publication frequency. (Color codes for references based on the target group's age: children, adults, both, not specifying.)

factors identified in the evaluations of digital CSTs and one *Other* category that enveloped the remaining factors. The distribution of these categories in the included studies is illustrated in Table 10. Products, approaches, and methods are discussed in Section 4.3, addressing RQ 1.2. Table 11 identifies the presence of individual factors in studies along with the higher-order factor categories summarized in Table 10. The remainder of this section will investigate each of these factor categories in detail.

4.2.1. Value

As identified in Table 11, 47 studies (58%) evaluated *value*, utilizing 13 different terms. *Quality*, *usefulness*, and *intuitiveness* were the three most frequently used terms in this review to indicate *value*. The terms with the most significant proportion of child and adult-focused studies were *quality* and *usefulness*, respectively. *Quality* was identified in 13 studies. Four studies explicitly defined *quality*, but their definitions and focus varied. Griffin and Jacob (2013, p. 150) defined it outcome-focused as "the length of the participant response", Voigt et al. (2013, p. 8) process-focused 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", Chu Yew Yee et al. (2011, p. 7) process focused as "coherence, continuation and completion", and Yoon et al. (2015, p. 13) process-focused as "the kit is to be used by people of different age groups, the device should be safe and should work reliably". Despite this, all these definitions agree with the Oxford Dictionary's definition of *value* as "the quality of being useful or important" (Oxford Dictionary, 2021) since the term depends on its context.

Two of the 12 studies that used the term *usefulness* explicitly defined what they meant by the term. Chang, Chen et al. (2019, p. 105) used the definition "functions, practicality, and attractiveness", while Kassima et al. (2014, p. 14) defined it as "how well the product does what it is supposed to do". In addition, four of the 11 studies focusing on *intuitiveness* defined it. They did it with a similarity; Benedetti et al. (2014, p. 21) as "easy to use", Chang and Luh (2012, p. 26) as "user-friendliness of

Table 8
Publication channels.

Publication channel	Study	Sum
ACM SIGCHI Conference on Human Factors in Computing Systems	Lu et al. (2011), Halpern et al. (2011), Bao et al. (2010), Haar Horowitz et al. (2018), Kato et al. (2015), Kazi et al. (2011), Maiden et al. (2018), Martin et al. (2016), Myers et al. (2015), Yoshida et al. (2016), Yu and Nickerson (2011)	11
Computers in Human Behavior	Piotrowski and Meester (2018), Chang et al. (2016), Yang et al. (2019), Guegan et al. (2016, 2017), Karakaya and Demirkan (2015), McGrath et al. (2016), Schmitt et al. (2012)	8
ACM SIGCHI Conference on Creativity & Cognition	Koh et al. (2011), Chu Yew Yee et al. (2011), Andolina et al. (2017), Siangliulue et al. (2015), Star et al. (2015), Torres and Paulos (2015), Zheng et al. (2017), Griffin and Jacob (2013)	8
International Journal of Child–Computer Interaction	Giannakos and Jaccheri (2018), Nouwena et al. (2016), Sapounidis et al. (2019), Sylla et al. (2015, 2018)	5
Thinking Skills and Creativity	Atwood-Blaine et al. (2019), Chang, Chen et al. (2019), Fleury et al. (2020), Kucirkovaa and Sakr (2015), Kassima et al. (2014)	5
ACM Conference on Computer Supported Cooperative Work	Cao et al. (2010), Nakazato et al. (2014), Wang et al. (2010)	3
ACM International Conference on Tangible, Embedded and Embodied Interaction	Catala et al. (2012), Oh et al. (2016), Willis et al. (2010)	3
ACM SIGCHI Conference on Designing Interactive Systems	Garcia et al. (2014), Torres et al. (2016), Wang et al. (2018)	3
ACM Transactions on Computer–Human Interaction	Cherry and Latulipe (2014), Dow et al. (2010), Kerne et al. (2014)	3
ACM SIGCHI Conference on Interaction Design and Children	Alves-Oliveira et al. (2017), Yoon et al. (2015)	2
ACM Symposium on User Interface Software and Technology	Benedetti et al. (2014), Zhang et al. (2013)	2
Computer-Aided Design and Applications	Bacciotti et al. (2016), Feeman et al. (2018)	2
ACM Conference on Intelligent User Interfaces	Davis et al. (2016), Huang et al. (2016)	2
ACM Transactions on Graphics	Shugrina et al. (2017), Wang et al. (2020)	2
Journal of Integrated Design and Process Science	Chang and Luh (2012)	1
Journal of Creative Behavior	Kim et al. (2013)	1
Research Journal of Applied Sciences, Engineering and Technology	Habibian Naeini and Masood (2012)	1
Connection Science	Kantosalo and Riihiaho (2019)	1
Journal of Music, Technology & Education	Downton (2015)	1
ACM Technical Symposium on Computer Science Education	Grover et al. (2018)	1
International Conference on Information Integration and Web-Based Applications & Services	Al-Mousawi and Alsumait (2012)	1
Young Children and Families in the Information Age: Applications of Technology in Early Childhood	Fessakis et al. (2015)	1

Table 8 (*continued*)

Publication channel	Study	Sum
Sensors	Gaeta et al. (2019)	1
ACM SIGGRAPH Conference on Motion, Interaction and Games	Chatain et al. (2019)	1
Interacting with Computers	Wojtczuk and Bonnardel (2011)	1
Methods of Information in Medicine	Thew et al. (2011)	1
Psychology in Russia: State of the Art	Voiskounsky et al. (2017)	1
Journal of Critical Reviews	Mohamed et al. (2020)	1
ACM/IEEE International Conference on Human Robot Interaction	Kahn et al. (2016)	1
IEEE/ACM International Conference on Automated Software Engineering	Mangano et al. (2010)	1
IEEE Transactions on Professional Communication	McDaniel et al. (2017)	1
IEEE Global Engineering Education Conference	Chang, Lin et al. (2019)	1
International Conference on Computational Creativity	Nelson et al. (2010)	1
International Conference on Information Systems	Voigt et al. (2013)	1
Australian Computer–Human Interaction Conference	Güldenpfennig et al. (2012)	1
Gesellschaft für Informatik e.V's lecture notes in informatics	Witt and Robra-Bissantz (2012)	1

Table 9
Definitions and focus areas of the 10 most frequent evaluation factors.

Factor	Definition	Focus
Value	The perceived usefulness of the digital CST.	Process/outcome
Novelty	The extent to which the user can create unique artifacts using the digital CST.	Outcome
Fluency	The degree the user creates relevant artifacts by using digital CST.	Process
Enjoyment	The pleasure the user gets from using the digital CST.	Process
User feeling	The user's perception of using the digital CST.	Process/outcome
Collaboration	The extent to which the user works with others on the digital CST.	Process
Expressiveness	The degree the artifacts created with the digital CST expresses the user's thoughts or feelings.	Outcome
Immersion	The extent to which the user is engaged with, immersed in, or attentive to the digital CST.	Process
Flexibility	The degree the artifacts created with the digital CST can be adapted to suit new circumstances.	Outcome
Interaction	How the user's actions changed the artifact created in the digital CST.	Process

(CST = creativity support tool).

product", Giannakos and Jaccheri (2018, p. 35) as "the degree to which an individual believes that attending the activity is easy", and Voigt et al. (2013, p.8) elaborating these definitions to "the property of a tool to foster a rapid and clear understanding of the artifacts employed for idea development". These definitions of *usefulness* and *intuitiveness* correlate well with each other and align well with the Oxford Dictionary's definition of *value* (Oxford Dictionary, 2021) — perhaps even to a larger degree than the definitions of *quality*.

An example of an evaluation of *quality* was Griffin and Jacob (2013), specifying an objective physiological measurement based on the Alternative Uses Task by examining the participant's response length for alternative uses of a common household object after using a digital music CST. On the other hand, Martin et al. (2016, p. 2299) used a subjective self-reporting questionnaire that included questions like

Table 10

Overview of the included studies' distribution of the 10 factors, nine products, three approaches (appr.), and five methods. (Color codes for references based on the target group's age: children, adults, both, not specifying.)

	Factors											Products									Appr.			Methods				
	Value	Novelty	Fluency	Enjoyment	User feeling	Collaboration	Expressiveness	Immersion	Flexibility	Interaction	Other	Idea generation	Model making	Storytelling	Game	Music	Drawing	Programming	Visualization	Writing	Physiological	Self-reporting	External judges	Observation	Questionnaire	Interview	Qualitative	Quantitative
(Chang & Luh, 2012)																												
(Kim et al., 2013)																												
(Habibian Naeini & Masood, 2012)																												
(Lu et al., 2011)																												
(Chu Yew Yee et al., 2011)																												
(Kantosalo & Riihiaho, 2019)																												
(Giannakos & Jaccheri, 2018)																												
(Sapounidis et al., 2019)																												
(Al-Mousawi & Alsumait, 2012)																												
(Yoon et al., 2015)																												
(Fessakis et al., 2015)																												
(Piotrowski & Meester, 2018)																												
(Alves-Oliveira et al., 2017)																												
(Cao et al., 2010)																												
(Grover et al., 2018)																												
(Downton, 2015)																												
(Koh et al., 2011)																												
(Atwood-Blaine et al., 2019)																												
(Sylla et al., 2018)																												
(Sylla et al., 2015)																												
(Nouwena et al., 2016)																												
(Catala et al., 2012)																												
(Fleury et al., 2020)																												
(Benedetti et al., 2014)																												
(Kucirkovaa & Sakr, 2015)																												
(Chang et al., 2019b)																												
(Chang et al., 2016)																												
(Yang et al., 2019)																												
(Oh et al., 2016)																												
(Willis et al., 2010)																												
(Chatain et al., 2019)																												
(Halpern et al., 2011)																												
(Gaeta et al., 2019)																												
(Kassima et al., 2014)																												
(Andolina et al., 2017)																												
(Kerne et al., 2014)																												
(Nelson et al., 2010)																												
(Chang et al., 2019a)																												
(Siangliulue et al., 2015)																												
(Bao et al., 2010)																												
(McDaniel et al., 2017)																												
(Davis et al., 2016)																												
(McGrath et al., 2016)																												
(Witt & Robra-Bissantz, 2012)																												
(Wojtczuk & Bonnardel, 2011)																												
(Mohamed et al., 2020)																												
(Yu & Nickerson, 2011)																												
(Star et al., 2015)																												
(Martin et al., 2016)																												
(Wang et al., 2018)																												
(Thew et al., 2011)																												
(Voigt et al., 2013)																												
(Zhang et al., 2013)																												
(Torres et al., 2016)																												
(Kato et al., 2015)																												
(Kazi et al., 2011)																												
(Zheng et al., 2017)																												
(Wang et al., 2020)																												
(Maiden et al., 2018)																												
(Myers et al., 2015)																												
(Schmitt et al., 2012)																												
(Guegan et al., 2016)																												
(Guegan et al., 2017)																												
(Yoshida et al., 2016)																												
(Bacciotti et al., 2016)																												
(Voiskounsky et al., 2017)																												
(Haar Horowitz et al., 2018)																												
(Huang et al., 2016)																												
(Dow et al., 2010)																												
(Kahn et al., 2016)																												
(Feeman et al., 2018)																												
(Karakaya & Demirkan, 2015)																												
(Nakazato et al., 2014)																												
(Cherry & Latulipe, 2014)																												
(Shugrina et al., 2017)																												
(Güldenpfennig et al., 2012)																												
(Mangano et al., 2010)																												
(Torres & Paulos, 2015)																												
(Garcia et al., 2014)																												
(Wang et al., 2010)																												
(Griffin & Jacob, 2013)																												
	47 (58%)	45 (56%)	29 (36%)	26 (32%)	24 (30%)	22 (27%)	12 (15%)	12 (15%)	11 (14%)	11 (14%)	20 (25%)	20 (25%)	16 (20%)	11 (14%)	10 (12%)	8 (10%)	6 (7%)	6 (7%)	5 (6%)	2 (2%)	62 (77%)	53 (65%)	26 (32%)	62 (77%)	57 (70%)	26 (32%)	60 (74%)	48 (59%)

"How would you rate the overall quality of that performance?" regarding the outcome of a digital music CST. The evaluations of *intuitiveness* also included both objective and subjective measures. For example, Oh et al. (2016) conducted objective physiological observations of the participants' outcomes after using a digital model-making CST, and Zhang et al. (2013) did subjective self-reporting interviews where the participants self-reported their experiences of the process of using a digital model-making CST. However, the evaluations of *usefulness* were only subjective. An example is Bao et al. (2010), using two external expert judges to evaluate how an idea that a participant produced by using a digital idea generation CST helped solve the problem using a five-point Likert scale.

Table 11

Distribution of the terms used to denote the 10 factors presented in Table 10. (Flu. = Fluency, Imm. = Immersion, Int. = Interaction, Color codes for references based on the target group's age: children, adults, both, not specifying.)

The table columns are grouped under the following factors, with the term variants listed below each:

- **Value**: Quality, Usefulness, Intuitiveness, Practicality, Logical, Elaboration, Feasibility, Aesthetics, Relevance, Adequate, Value, Completeness, Marketability
- **Novelty**: Originality, Novelty, Variance, Uniqueness, Innovation, Unexpecting
- **Flu. (Fluency)**: Fluency, Idea generation, Frequency
- **Enjoyment**: Enjoyment, Duration, Fun, Playfulness, Interest, Result satisfaction
- **User feeling**: User feeling, Preference, Ownership, Self-efficacy, Affect
- Collaboration, Expressiveness
- **Imm. (Immersion)**: Engagement, Immersion, Presence
- Flexibility
- **Int. (Interaction)**: Interaction, Behavior

Rows (references):

(Chang & Luh, 2012)
(Kim et al., 2013)
(Habibian Naeini & Masood, 2012)
(Lu et al., 2011)
(Chu Yew Yee et al., 2011)
(Kantosalo & Riihiaho, 2019)
(Giannakos & Jaccheri, 2018)
(Sapounidis et al., 2019)
(Al-Mousawi & Alsumait, 2012)
(Yoon et al., 2015)
(Fessakis et al., 2015)
(Piotrowski & Meester, 2018)
(Alves-Oliveira et al., 2017)
(Cao et al., 2010)
(Grover et al., 2018)
(Downton, 2015)
(Koh et al., 2011)
(Atwood-Blaine et al., 2019)
(Sylla et al., 2018)
(Sylla et al., 2015)
(Nouwena et al., 2016)
(Catala et al., 2012)
(Fleury et al., 2020)
(Benedetti et al., 2014)
(Kucirkovaa & Sakr, 2015)
(Chang et al., 2019b)
(Chang et al., 2019b)
(Yang et al., 2019)
(Oh et al., 2016)
(Willis et al., 2010)
(Chatain et al., 2019)
(Halpern et al., 2011)
(Gaeta et al., 2019)
(Kassima et al., 2014)
(Andolina et al., 2017)
(Kerne et al., 2014)
(Nelson et al., 2010)
(Chang et al., 2019a)
(Siangliulue et al., 2015)
(Bao et al., 2010)
(McDaniel et al., 2017)
(Davis et al., 2016)
(McGrath et al., 2016)
(Witt & Robra-Bissantz, 2012)
(Wojtczuk & Bonnardel, 2011)
(Mohamed et al., 2020)
(Yu & Nickerson, 2011)
(Star et al., 2015)
(Martin et al., 2016)
(Wang et al., 2018)
(Thew et al., 2011)
(Voigt et al., 2013)
(Zhang et al., 2013)
(Torres et al., 2016)
(Kato et al., 2015)
(Kazi et al., 2011)
(Zheng et al., 2017)
(Wang et al., 2020)
(Maiden et al., 2018)
(Myers et al., 2015)
(Schmitt et al., 2012)
(Guegan et al., 2016)
(Guegan et al., 2017)
(Yoshida et al., 2016)
(Bacciotti et al., 2016)
(Voiskounsky et al., 2017)
(Haar Horowitz et al., 2018)
(Huang et al., 2016)
(Dow et al., 2010)
(Kahn et al., 2016)
(Feeman et al., 2018)
(Karakaya & Demirkan, 2015)
(Nakazato et al., 2014)
(Cherry & Latulipe, 2014)
(Shugrina et al., 2017)
(Güldenpfennig et al., 2012)
(Mangano et al., 2010)
(Torres & Paulos, 2015)
(Garcia et al., 2014)
(Wang et al., 2010)
(Griffin & Jacob, 2013)

Column totals (count and percentage):

Term	Total
Quality	13 (16%)
Usefulness	12 (15%)
Intuitiveness	11 (14%)
Practicality	7 (9%)
Logical	7 (9%)
Elaboration	6 (7%)
Feasibility	5 (6%)
Aesthetics	5 (6%)
Relevance	4 (5%)
Adequate	3 (4%)
Value	3 (4%)
Completeness	2 (2%)
Marketability	1 (1%)
Originality	24 (30%)
Novelty	14 (17%)
Variance	6 (7%)
Uniqueness	4 (5%)
Innovation	2 (2%)
Unexpecting	2 (2%)
Fluency	15 (19%)
Idea generation	12 (15%)
Frequency	2 (2%)
Enjoyment	12 (15%)
Duration	9 (11%)
Fun	5 (6%)
Playfulness	5 (6%)
Interest	1 (1%)
Result satisfaction	9 (11%)
User feeling	7 (9%)
Preference	6 (7%)
Ownership	3 (4%)
Self-efficacy	3 (4%)
Affect	1 (1%)
Collaboration	22 (27%)
Expressiveness	12 (15%)
Engagement	8 (10%)
Immersion	3 (4%)
Presence	1 (1%)
Flexibility	11 (14%)
Interaction	8 (10%)
Behavior	3 (4%)

4.2.2. Novelty

Novelty is frequently used in evaluations of digital CSTs, as discussed in Section 2. As indicated in Table 11, 45 studies (56%) looked at *novelty*, using six different terms to evaluate it. Fourteen studies used exactly the word *novelty*, but approximately half did not specify how they defined it. Those who defined the term explicitly defined it differently. Andolina et al. (2017) focused on *surprising*; Chang, Chen et al. (2019) on *uniqueness*; Kerne et al. (2014) on *rarity*; McDaniel et al. (2017) on *infrequency* and *rarity*; Catala et al. (2012) on *unusual*, *uniqueness*, and *surprising*; and Siangliulue et al. (2015) on *novel*, *originality*, and *surprising*.

Nevertheless, 24 of the 45 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 define the term had slightly different

focuses. Alves-Oliveira et al. (2017), Bao et al. (2010), and Nelson et al. (2010) focused on *novelty*; Benedetti et al. (2014), Fleury et al. (2020) and Kim et al. (2013) on *uniqueness*; Piotrowski and Meester (2018) on *unusual*; Fessakis et al. (2015) on *unusual, relevance* and *rarity*; Kassima et al. (2014) on *surprising, unusual, uniqueness,* and *rarity*; McGrath et al. (2016) on *novelty, unusual,* and *infrequent*; and Habibian Naeini and Masood (2012) on *innovative, unusual,* and *fresh*. Most of these terms correspond with the mentioned terms used to define *novelty*, which also is found in the Oxford Dictionary's definition as "the quality of being new, different and interesting" (Oxford Dictionary, 2021). *Originality* was the term with the most considerable amount of child-focused studies, while *novelty* was mainly used in adult-focused studies. *Originality* was evaluated both subjectively and objectively in the included studies. Witt and Robra-Bissantz (2012) used a subjective evaluation with external expert judges using a five-point Likert scale to rate the degree of originality of the ideas the users made by using a digital idea generation CST. On the other hand, Kim et al. (2013) objectively and physiologically observed the number of unique responses the users had in creative problem-solving tests after using a digital programming CST. *Variance* was also measured objectively. An example is Koh et al. (2011), which physiologically compared the nine-element vector *v* of a user's programming artifact produced with the help of a digital programming CST 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

Table 11 illustrates that 29 studies (36%) evaluated *fluency*, utilizing three different terms. In contrast to *value* and *novelty*, none of the studies used several terms. Most of the studies used exactly the word *fluency*, while 12 used *idea generation* and two *frequency*. *Fluency* was the term with the highest proportion of child-focused studies, while mainly adult-focused studies included *idea generation* or *frequency*. Most of the studies that used the term *fluency* defined it. They did so uniformly, corresponding with the definition provided in the Oxford Dictionary as "the quality of doing something in a way that is smooth and shows skill" (Oxford Dictionary, 2021). In fact, Fleury et al. (2020, p. 3), Guegan et al. (2016, p. 170), and Guegan et al. (2017, p. 143) used the same definition: "the number of ideas generated". However, there were some variations. For example, Kassima et al. (2014, p. 14) defined it as "the number of interpretable, meaningful and relevant responses" and Nelson et al. (2010, p. 206) as "the frequency of ideas generated".

Fluency is mostly measured objectively by physiologically counting the number of relevant artifacts (e.g., ideas) the user generates using the digital CST. Most studies did not specify a theoretical grounding for choosing *idea generation* as an evaluation factor. On the other hand, Alves-Oliveira et al. (2017, p. 424) said: "According to literature, idea generation (or fluency) is one fundamental creative ability that tends to lead to originality and novelty (Guilford, 1967). With YOLO, we aim to stimulate fluency during the creative process of storytelling".

4.2.4. Enjoyment

Table 11 shows that 26 studies (32%) evaluated *enjoyment* using five different terms. The word *enjoyment* had the highest proportion of the child-focused studies, while *duration* was mainly used in the studies targeting adults. Twelve studies used the word *enjoyment*, but only Giannakos and Jaccheri (2018, p. 35) defined it. This definition was: "the degree to which the activity is perceived to be personally enjoyable", which corresponds with the definition provided by the Oxford Dictionary as "the pleasure that you get from something" (Oxford Dictionary, 2021).

Most of the studies used subjective self-reporting evaluations of *enjoyment*. Atwood-Blaine et al. (2019, p. 6) used a questionnaire for users of a digital game CST with the question: "How much did you enjoy

the creativity project? (Circle a number on the scale below.)", while Thew et al. (2011, p. 163) conducted interviews with users of a digital visualization CST including the question "Do you enjoy using the system?". However, Cao et al. (2010) used physiological observation in the form of a camera to record the user's interactions around a digital storytelling CST, including *enjoyment. Duration* could also be measured objectively by physiologically observing users' interaction time with a digital storytelling CST like Sylla et al. (2015). On the other hand, Star et al. (2015, p. 109) evaluated *duration* subjectively for the user's process with a digital idea generation CST with a self-reporting questionnaire, including the statement: "Time appeared to go by quickly when I was interacting with the activity".

4.2.5. User feeling

As illustrated in Table 11, 24 studies (30%) evaluated *user feeling*, utilizing six different terms. Seven studies used the process-focused term *user feeling*, but nine used the more outcome-focused term *result satisfaction. Result satisfaction* had a high proportion of adult-focused studies. On the other hand, *user feeling* included a more significant amount of child-focused studies. Few studies defined what they meant by *result satisfaction*. However, Guegan et al. (2016, p. 170), Kantosalo and Riihiaho (2019, p. 64) and Torres et al. (2016, p. 165) defined it similarly, corresponding to the Oxford Dictionary's definition of *feeling* as "an attitude or opinion about something" (Oxford Dictionary, 2021). The definitions were respectively: "requires the users to evaluate the final result of the creative process", "[the users'] perceptions of [...] their avatar", and "happy with the final design".

User feeling was evaluated subjectively by self-reporting, both with questionnaires and interviews. For example, Giannakos and Jaccheri (2018, p. 35) used 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" to evaluate the user's perception of using a digital programming CST. In addition, Kassima et al. (2014) interviewed the subjects about their opinions of using a digital model making CST, which resulted in responses that aligned with the above definition of *feeling*.

4.2.6. Collaboration

As presented in Table 11, 22 studies (27%) evaluated this factor, and each used the word *collaboration*. There was slightly more adult- than child-focused studies. Very few defined *collaboration*, but Kantosalo and Riihiaho (2019, p. 65) defined it as "mutual influence, sharing and feedback with different agents", and Catala et al. (2012, 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". Both of these definitions correspond with the definition provided by the Oxford Dictionary (2021) as "the act of working with another person or group of people to create or produce something".

Different evaluations were used to evaluate *collaboration*. Kucirkovaa and Sakr (2015) observed it objectively by physiologically identifying episodes when both participants contributed to an idea in a digital storytelling CST. Sapounidis et al. (2019, p. 71) subjectively measured the *collaboration* in a digital programming CST by self-reporting with the question "With what system did you collaborate the most?" in the Fun-Sorter and the Likert question "How much did you collaborate with the Graphical/Tangible interface?". Further, Schmitt et al. (2012) calculated *collaboration* for a digital idea generation CST physiologically through the inequity index $I = \left| \frac{1}{N} - \frac{O_i}{\sum_{i=1}^{N} O_i} \right| \times 100$, where N is the group size, and O_i is the observed collaborative behavior for each participant.

4.2.7. Expressiveness

Table 11 illustrates that 12 studies (15%) evaluated this factor, all utilizing the word *expressiveness*. These studies were mainly targeting adults. Few defined it, and those who did use different definitions.

Kantosalo & Riihiaho's definition (Kantosalo & Riihiaho, 2019 p. 64) stating "how well the users are able to be creative and express themselves in the creative process" matches Oxford Dictionary's definition (Oxford Dictionary, 2021) as "the quality of expressing somebody's thoughts and feelings". On the other hand, the definitions of Karakaya and Demirkan (2015, p. 181) ("the success in producing a variety of drawings for a specified task"), Oh et al. (2016, p. 346) ("the wide range of constructions that the kit makes possible for both adults and kids"), and Yoon et al. (2015, p. 13) ("encourage users to explore topics through a new form of storytelling medium") are aligned with the idea of *variance*. The studies used both subjective and objective evaluations. For example, Cherry and Latulipe (2014, p. 6) included a self-reporting questionnaire to evaluate digital storytelling, drawing, and visualization CSTs, 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". Alternatively, Oh et al. (2016) observed physiologically the range of constructions created with a digital model making CST by both children and adults.

4.2.8. Immersion

As indicated in Table 11, 12 studies (15%) evaluated *immersion* using three different terms. Only three studies used the actual word *immersion*, while eight used the term *engagement*. *Immersion* was mainly used in the studies targeting adults, while *engagement* had a higher proportion in child-focused studies. Few studies specified their interpretation or definition of the term *engagement*. Those who defined it described it differently. Torres et al. (2016) identified it as *absorption with the activity*, which is most like the definition of *immersion* provided by the Oxford Dictionary (2021) as 'the state of being completely involved in something". On the other hand, Sylla et al. (2018) described it as *curiosity* and Yoon et al. (2015) as *interest*.

A majority of the studies used self-reporting subjective evaluations for *immersion*. For example, Yoon et al. (2015) interviewed the participants regarding their engagement in the process of using a digital model making CST. Further, Star et al. (2015, p. 109) conducted a qualitative questionnaire to evaluate the *immersion* of a digital idea generation CST that included 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, Torres et al. (2016) physiologically calculated the difference between perceived time and completion time for a user that utilized a digital model making CST. Other studies used objective physiological observation, like Zhang et al. (2013) asking the participants to think aloud throughout their use of a digital model making CST and observed their engagement behavior through what they described in the think-aloud.

4.2.9. Flexibility

Flexibility is "the ability to change to suit new conditions or situations" (Oxford Dictionary, 2021). As illustrated in Table 11, 11 studies (14%) evaluated this factor, all utilizing the word *flexibility*. It was an equal distribution of child- and adult-focused studies. Most studies specified how they interpreted the term. *Flexibility* was generally interpreted in the same manner but explained slightly differently in the studies. In short, Kassima et al. (2014, p. 14) defined *flexibility* as a "variety of categories of relevant responses".

This definition was elaborated to be more similar to the definition in the Oxford Dictionary (2021) as "the ability to change to suit new conditions or situations" by Kerne et al. (2014, 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 of the studies that evaluated *flexibility* did it like Kerne et al. (2014), by physiologically observing the number of different categories for the artifacts (ideas) the user produced using a digital idea generation CST. A few studies specified the evaluation in detail. For example, Kim

et al. (2013) calculated *flexibility* by the number of tools, principles, and procedures used to explain ideas produced by users of a digital programming CST. On the other hand, Catala et al. (2012) used external expert judges to rate the *flexibility* of the users' outcomes of a digital model-making CST.

4.2.10. Interaction

"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" (Oxford Dictionary, 2021). Table 11 shows that 11 studies (14%) evaluated this factor, where eight used the word *interaction* and three used the word *behavior* (as in behaviors that affect the digital CST usage). The word *interaction* had a high amount of adult-focused studies, while *behavior* had a higher proportion of child-focused studies. Interestingly, none of the included studies explained how they understood these terms. The evaluations used by all studies in this category to identify *interaction* were physiologically observing the participants' interaction with the digital CST.

4.2.11. Other

Twelve different terms did not fit the 10 previous factors and were included in the *Other* category. As presented in Table 11, 20 studies (25%) included this category. The factors *persistence*, *risk-taking*, *self-determination*, *start help*, *success guarantee*, and *accessibility* were present in one study each. On the other hand, *emergence* occurred in two studies; *inspiration*, *motivation*, and *imagination* in three; and *knowledge* in four. *Exploration* was the term included most in this *Other* category, mentioned in seven studies. Overall, most studies in this category targeted adults, which also applied to the *exploration* factor. On the other hand, the factors *knowledge* and *imagination* had an equal distribution in adult-and child-focused studies.

4.3. Context

This section addresses RQ 1.2 regarding the context of the evaluation. In this case, the context is the evaluated products, the approach used to evaluate creativity, and the research methods used for data collection and data analysis.

4.3.1. Product

The distribution of the creativity topics for the products in the included studies is presented in Table 10 and further specified in Table 12. A total of nine different creativity topics appeared. The ones that appeared the most frequently were *idea generation*, *model making*, and *games*. The *idea generation product category* is brainstorming and brainwriting tools, and not precisely the same as the term *idea generation* used to describe *fluency* in Section 4.2.3 as the number of generated ideas. The product category with the most significant proportion of child- and adult-focused studies was *programming* and *visualization*, respectively.

4.3.2. Approaches

Fig. 3 illustrates the included studies' distribution according to Carroll and Latulipe (2012)'s three basic approaches to creativity measurement, described in Section 2. This distribution is also illustrated in Table 10. The most frequently utilized approach is *physiological measurements*, with 62 studies (77%). *Self-reporting* was also a commonly-used approach, used by 53 studies (65%). *External judges* were used only in 26 studies (32%). Fig. 3 illustrates that most studies used a combination of these approaches. *Physiological measurement* and *self-reporting* were the most common co-occurring approaches used by 33 studies (41%). In addition, seven of the studies employed all three approaches. It was exclusively adult-focused studies that only used *self-reporting*. On the other hand, a large proportion of the studies that used only *physiological measurement* were child-focused.

Table 12
Overview of the products in the included studies. (Color codes for references based on the target group's age: children, adults, both, not specifying.)

Topic	Creativity support tool	Sum
Idea generation	• The idea generation apps *AppLab* and *Common Sense Media* (Piotrowski & Meester, 2018) • Virtual reality (VR) idea generation tools (Fleury et al., 2020; Yang et al., 2019) • Self-developed digital idea-generation tools (Andolina et al., 2017; Haar Horowitz et al., 2018; Kahn et al., 2016; Karakaya & Demirkan, 2015; Kerne et al., 2014; Mangano et al., 2010; Nelson et al., 2010; Siangliulue et al., 2015; Voigt et al., 2013; Yoshida et al., 2016) • Self-developed digital brainstorming tools (Bao et al., 2010; Star et al., 2015; Wang et al., 2010) • Self-developed digital brainwriting tool (Schmitt et al., 2012) • Self-developed online multiplayer ideation game (Witt & Robra-Bissantz, 2012) • Brainstorming in a video conference system (Nakazato et al., 2014) • *Adobe Connect* (McGrath et al., 2016)	20 (25%)
Model making	• Self-developed digital model making kits (Chang & Luh, 2012; Yoon et al., 2015; Oh et al., 2016) • Self-developed 3D-sketch app (Willis et al., 2010) • 3D-computer aided design (CAD) (Chang, Chen et al., 2019; Chang et al., 2016) • Self-developed digital prototype tools (Catala et al., 2012; Torres et al., 2016; Wang et al., 2018; Zheng et al., 2017) • Self-developed digital multimedia learning tool for mechanism design (Kassima et al., 2014) • Self-developed digital 3D-model tools (Torres & Paulos, 2015; Zhang et al., 2013) • CAD (Bacciotti et al., 2016) • The CAD *Rhino* (Wojtczuk & Bonnardel, 2011) • VR CAD (Feeman et al., 2018)	16 (20%)
Storytelling	• The digital storytelling software *Frames* combined with *Powerpoint* (Chu Yew Yee et al., 2011) • Self-developed digital storytelling tools (Al-Mousawi & Alsumait, 2012; Alves-Oliveira et al., 2017; Cao et al., 2010; Lu et al., 2011; Sylla et al., 2015, 2018; Kato et al., 2015; Kazi et al., 2011) • Storytelling with the software *Tuxpaint* combined with the *Our Story* app (Kucirkovaa & Sakr, 2015) • Short story writing in *Google Docs* (Cherry & Latulipe, 2014)	11 (14%)
Game	• The game development platforms *ARIS* (Atwood-Blaine et al., 2019) • The digital problem-solving game *Crayon Physics Deluxe* (Fessakis et al., 2015) • The educational computer game *I Spy Treasure Hunt* (Habibian Naeini & Masood, 2012) • Self-developed game creation apps (Chatain et al., 2019; Gaeta et al., 2019) • Self-developed digital gamified learning management system (McDaniel et al., 2017) • The game development platforms *Construct, Powerpoint, Powtoom*, and *Roar* (Mohamed et al., 2020) • The digital constructing game *Minecraft* (Voiskounsky et al., 2017) • The VR idea generation game *Second Life* (Guegan et al., 2016, 2017)	10 (12%)
Music	• Self-developed digital educational music game (Nouwena et al., 2016) • Self-developed digital music creation tools (Downton, 2015; Garcia et al., 2014; Huang et al., 2016; Martin et al., 2016) • Self-developed app for music creation and manipulation through movement (Halpern et al., 2011) • Visual music interactive art system with *Processing* and *Leap Motion* (Chang, Lin et al., 2019)	8 (10%)

Table 12 (*continued*)

Topic	Creativity support tool	Sum
	• Self-developed adaptive digital musical instrument (Griffin & Jacob, 2013)	
Drawing	• Self-developed digital painting tool (Benedetti et al., 2014) • Self-developed digital co-creative drawing agent (Davis et al., 2016) • Self-developed digital color picker (Shugrina et al., 2017) • Sketch combination system integrating *Mechanical Turk* and *Google Docs* (Yu & Nickerson, 2011) • *OdoScetch, AutoDesk Scetchbook*, and Mac OS's color picker *BiCEP* (Cherry & Latulipe, 2014) • The Java app *Pixelitor* (Myers et al., 2015)	6 (7%)
Programming	• *Scratch* (Giannakos & Jaccheri, 2018; Grover et al., 2018; Kim et al., 2013) • *App Inventor* (Grover et al., 2018) • Unspecified programming tools (Koh et al., 2011; Sapounidis et al., 2019) • *Scratch* and *Unity* (Mohamed et al., 2020)	6 (7%)
Visualization	• Self-developed digital photography tool (Güldenpfennig et al., 2012) • Self-developed digital data visualization tool (Thew et al., 2011) • Self-developed digital font visualization tool (Wang et al., 2020) • *Adobe Photoshop* (Cherry & Latulipe, 2014) • MySpace's Flash-based *AdBuilder Tool* (Dow et al., 2010)	5 (6%)
Writing	• Self-developed digital poetry tool (Kantosalo & Riihiaho, 2019) • Self-developed digital journalism tool (Maiden et al., 2018)	2 (2%)

4.3.3. Methods

Table 10 indicates that the studies used different research methods and a combination of different research methods. A majority of the studies used multiple data collection methods. *Observations* were the most frequently used data collection method with 62 studies (77%), while *questionnaires* and *interviews* followed with 57 (70%) and 26 (32%), respectively. *Questionnaire* and *interview* were the data collection methods with the largest proportion of adult and child-focused studies, respectively. The data analysis performed in the studies was mostly *qualitative*, with qualitative methods used in 60 studies (74%). In contrast, the *quantitative* studies had the largest proportion of child-focused studies. However, 48 studies (59%) utilized *quantitative* data analysis, and 27 studies (33%) used a *mixed method*.

Some studies also used existing creativity measurements presented in Section 2. Known creativity measurements that appeared in the studies were Alternative Uses Task (Fleury et al., 2020; Haar Horowitz et al., 2018; Griffin & Jacob, 2013), Torrance Tests of Creative Thinking (Habibian Naeini & Masood, 2012; Kim et al., 2013; Kassima et al., 2014), abbreviated Torrance test for adults (Chang, Lin et al., 2019), Multidimensional Stimulus Fluency Measure (Fessakis et al., 2015), Remote Associates Test (Fleury et al., 2020), Creativity Achievement Questionnaire (Dow et al., 2010), Kaufman Domains of Creativity Scale (Yang et al., 2019), theory of inventive problem solving (Mohamed et al., 2020), Consensual Assessment Technique (Chu Yew Yee et al., 2011; Downton, 2015), AttrakDiff (Gaeta et al., 2019), Creative Product Semantic Scale (Kassima et al., 2014), and Creativity Support Index (Cherry & Latulipe, 2014; Shugrina et al., 2017). Fun-Sorter (Sapounidis et al., 2019) and MemoLine (Nouwena et al., 2016) were also used.

A few studies were missing information regarding the methods. Griffin and Jacob (2013) did not mention anything about the participants. Further, 27 studies (33%) mentioned participants but lacked demographic information regarding the gender distribution among the participants, while Mohamed et al. (2020) and Wang et al. (2010) did
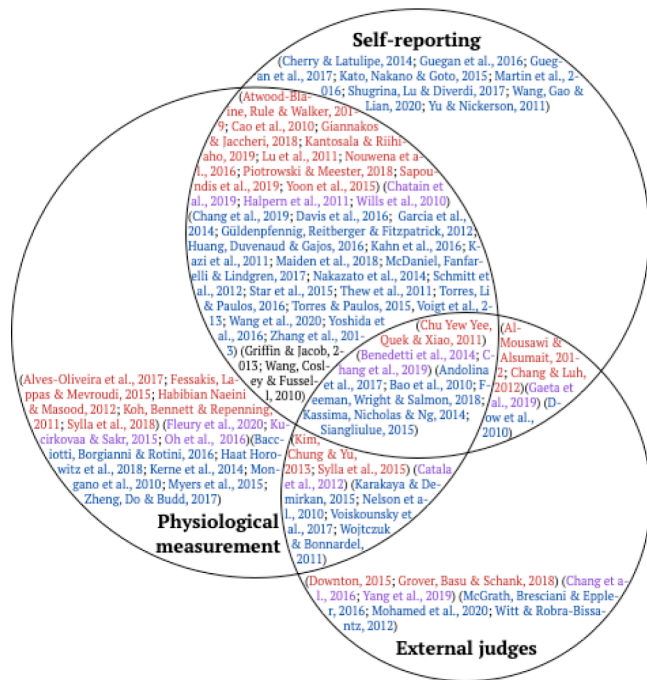
Fig. 3. The distribution of the approaches in the included studies. (Color codes for references based on the target group's age: children, adults, both, not specifying.)

not have information regarding sample size and age, respectively.

4.4. Age of the target group

This section addresses RQ 1.3 regarding how the target group's age impacts the evaluations of digital CSTs in the included studies. Table 10 indicates that the included studies' target group is mostly adults. Further, 58 studies (72%) had adult participants, and 33 studies (41%) had child participants, while 12 studies (15%) used both types of participants; 2 studies (2%) (Griffin & Jacob, 2013; Wang et al., 2010) did not specify the participant's age.

Table 13 illustrates how the age of the target group impacts the evaluation factors, products, and methods. Table 14 presents the distribution of the different terms used to evaluate the factors that used more than one term. This table does not include collaboration, expressiveness, and flexibility because only one term was used to evaluate these factors, as indicated in Table 11.

Table 13 indicates the most frequent factors are not similar for the different age groups. For example, enjoyment and collaboration have a coverage of 39% and 33% for children compared to 28% and 24% for adults, respectively. In contrast, fluency has a coverage of 38% for adults and 27% for children. Additionally, there were differences in the occurrences and use of flexibility, with a coverage of 21% for papers with child participants and only 12% for papers with adults. Moreover, the

different terms used to evaluate the different factors vary for children and adults. For example, only studies targeting adults use the terms feasibility, marketability, unexpecting, frequency, affect, and presence. Feasibility is the most remarkable term since it has a coverage of 9% for adults. For children, elaboration, and enjoyment with a coverage of 15% and 24% are the terms with the most significant differences in the occurrences compared to adults with a coverage of 5% and 10%, respectively. The products in the studies are also different for adults and children. The most common product for children is storytelling, with a coverage of 24%, which only has a coverage of 7% for adults. In addition, programming is the third most common product for children with a coverage of 15% but only 2% for adults. For adults, the most common product is idea generation with a coverage of 31%, which only has a coverage of 9% for children. Another interesting fact is that there are no visualization products for children and only coverage of 3% of drawing products, which are the fifth and fourth most common product for adults with a coverage of 9% and 10%, respectively.

The approaches in the included studies depend on the participant's age. For adults, there was slightly more physiological measurement (74%) than self-reporting (67%), while the studies with children had much more physiological measurement (79% against 55%). This aspect is reflected in the evaluations of digital CST in the studies with children with slightly more qualitative (73%) than quantitative (67%) data analysis. In contrast, studies with adults have more qualitative data analysis (72% against 53%). Concerning the data collection method, there is almost an equal number of observations (74%) and questionnaires (72%) for adults. In comparison, the studies with children have more observations of them interacting with the product (79%) instead of questionnaires (64%).

4.5. Credibility

This section addresses RQ 1.4 on the credibility of the evaluation of the digital CSTs in the included studies by first presenting how the included studies interpreted the term creativity and then the validity of the methods in the included studies.

4.5.1. Creativity

As illustrated in Fig. 4, 42 studies (52%) do not define creativity and nine studies (11%) that defined it used unclear definitions. An example is Wang et al. (2010, p. 103) stating: "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 unanimous opinion on creativity. There are a few overlaps among the different definitions, but 17 studies (21%) include a definition of something with the two characteristics novel and valuable. Most of these studies were adult-focused. Only two studies (2%) (Guegan et al., 2017; Schmitt et al., 2012) used the same definition of creativity by Sternberg as "the ability to produce work that is both novel and appropriate" (Sternberg & Lubart, 1999 p. 3).

Thirteen studies (16%) focus more on the process of creating something. A considerable amount of these studies were child-focused. For

Table 13
The coverage percentage of factors, products, approaches (appr.), and methods for children and adults in the included studies.

	Overall	Factors											Products									Appr.			Methods			
		Value	Novelty	Fluency	Enjoyment	User feeling	Collaboration	Expressiveness	Immersion	Flexibility	Interaction	Other	Idea generation	Model making	Storytelling	Game	Music	Drawing	Programming	Visualization	Writing	Physiological	Self-reporting	External judges	Observation	Questionnaire	Interview	Qualitative	Quantitative
Children	41	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	72	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

Table 14

The coverage percentage of the different terms of the evaluation factors in Table 13 uses more than one term to evaluate the factor. (Flu. = Fluency, Imm. = Immersion, Int. = Interaction).

	Overall	Value													Novelty						Flu.			Enjoyment					User feeling						Imm.			Int.	
		Quality	Usefulness	Intuitiveness	Practicality	Logical	Elaboration	Feasibility	Aesthetics	Relevance	Adequate	Value	Completeness	Marketability	Originality	Novelty	Variance	Uniqueness	Innovation	Unexpecting	Fluency	Idea generation	Frequency	Enjoyment	Duration	Fun	Playfulness	Interest	Result satisfaction	User feeling	Preference	Ownership	Self-efficacy	Affect	Engagement	Immersion	Presence	Interaction	Behavior
Children	41	12	6	15	6	9	15	0	3	3	6	3	3	0	33	18	9	3	6	0	24	3	0	24	9	6	6	3	6	9	9	9	6	0	12	3	0	9	3
Adults	72	12	19	14	10	9	5	9	9	7	2	5	2	2	24	21	7	5	3	3	17	17	3	10	12	5	7	2	12	10	7	3	3	2	7	5	2	10	3
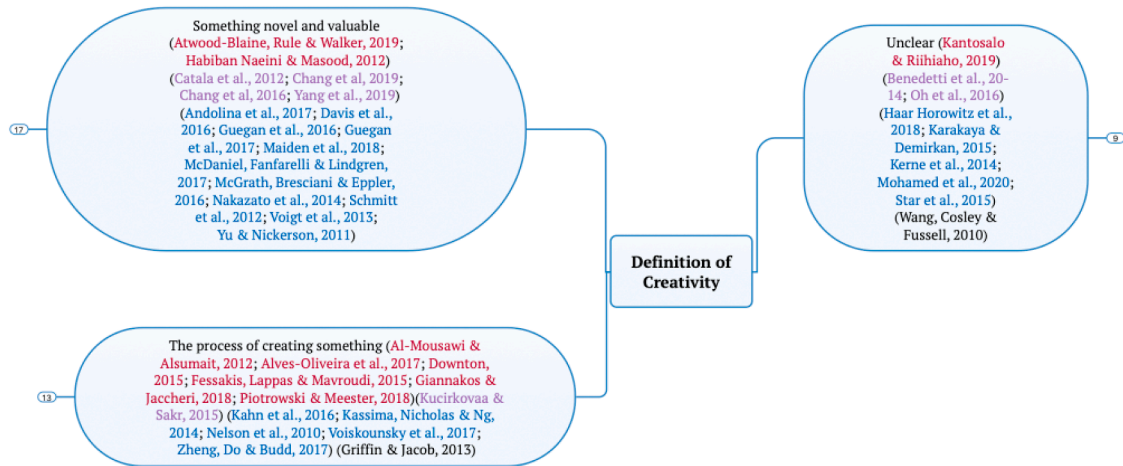
Fig. 4. Definitions of creativity in the 39 studies (48%) defined the term. (Color codes for references based on the target group's age: children, adults, both, not specifying.)
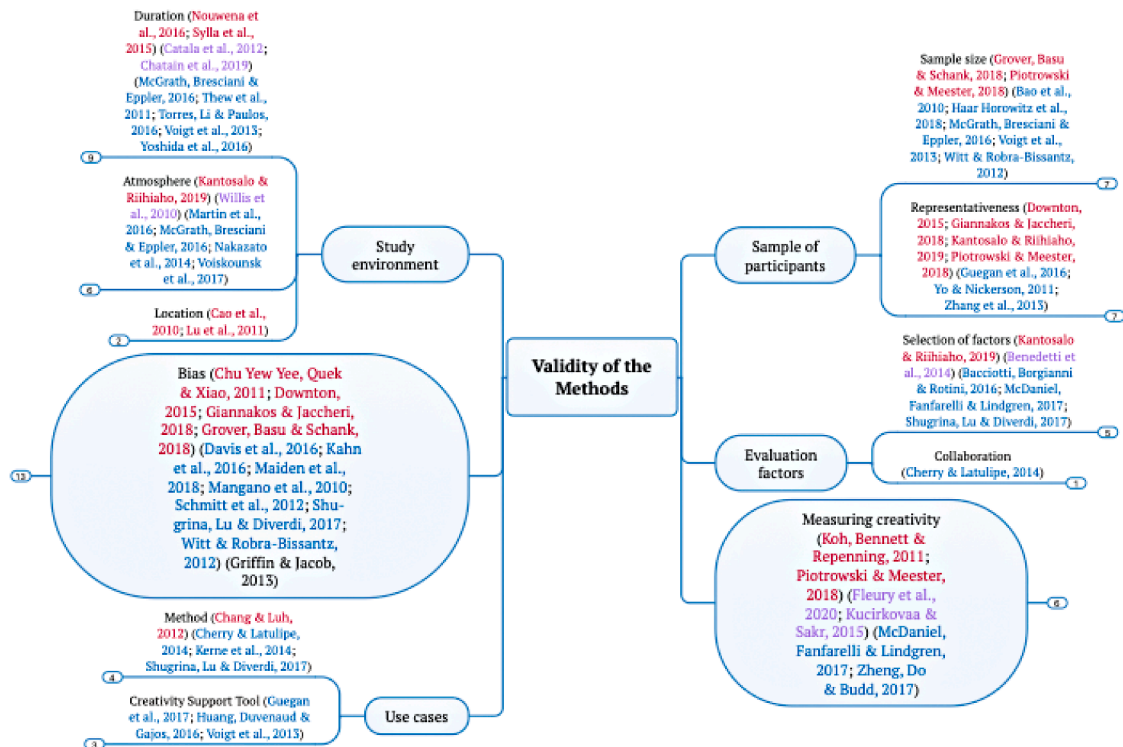
Fig. 5. The validity of methods in the 44 studies (54%) included details of the validity. (Color codes for references based on the target group's age: children, adults, both, not specifying.)

example, Fessakis et al. (2015, p. 209) 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". Both of these examples of definitions correspond to the definition provided by the Oxford Dictionary: "the use of skill and imagination to produce something new or to produce art" (Oxford Dictionary, 2021).

4.5.2. Validity of the methods

The validity of the methods was not mentioned in 37 (46%) of the included studies. The distribution among the 44 studies (54%), including details of the validity, is depicted in Fig. 5: 17 studies (21%) discussed the *study environment* regarding *duration*, *atmosphere*, or *location*; 14 studies (17%) the *sample of participants* regarding *sample size* or *representativeness*; 12 studies (15%) potential *biases*; seven studies (9%) the *evaluation factors*; seven *use cases* of the method or the CST; and six studies (7%) *measuring creativity*. A few of the studies included several of these six categories. The *use case* category had the most significant amount of adult-focused studies, while the category *sample of participants* had a larger proportion of child-focused studies.

Regarding *evaluation factors*, Cherry and Latulipe (2014) emphasized the difficulty of including *collaboration* in a standardized creativity index because many creative activities are not collaborative. The studies have different approaches for the selection of *evaluation factors*. Benedetti et al. (2014) and Kantosalo and Riihiaho (2019) chose the evaluation metrics by partially overlapping metrics from three and nine prior research articles, respectively, and Shugrina et al. (2017) used the well-tested Creativity Support Index. On the other hand, Bacciotti et al. (2016) recommended including *quality* and *novelty* as evaluation factors, and McDaniel et al. (2017) indicated how *importance*, *novelty*, and *affect* could be used to assess creativity in a web-based information design system.

Further, the studies have different limitations in *measuring creativity*. Fleury et al. (2020) indicated that the study's main limitation was their measuring of creativity with interpretations of Alternative Uses Task and Remote Associates Test scores. On the other hand, Kucirkovaa and Sakr (2015) focused on the aspect mentioned by Glăveanu (2010, p. 91) regarding the difference in how "members of different communities assess the creativity of one and the same artifact". One limitation of the study by McDaniel et al. (2017) is the difficulty in distinguishing the evaluation of the creative design from the creative pedagogy. It involves blurring the boundaries between the learning system itself and the content within the system. Further, measuring engagement for young children was a significant challenge in the study by Piotrowski and Meester (2018). Three different evaluations of digital CSTs were utilized because of conflicting ideas, but two of them were excluded due to the validity of the results. Consequently, the study could only use an adapted self-reporting of *engagement*.

In addition, the studies highlighted different evaluations of digital CSTs in the *use cases* category. Koh et al. (2011) highlighted divergence tests and considered divergence from the accepted norm a significant creativity indicator. However, Yu and Nickerson (2011) used a binary measure that only qualifies designs as creative if it exceeds 4.0 on a seven-point Likert scale on both the scales for the factors *originality* and *practicality*.

5. Discussion

5.1. Evaluations of digital creativity support tools for children

The present SLR of 81 papers identified a wide spectrum of evaluations of digital CSTs. These evaluations are summarized in Table 10 as an evaluation framework for digital CSTs for children based on different factors, products, approaches, and methods. Numerous factors have been considered in the evaluations, but the 10 most common ones were *value*, *novelty*, *fluency*, *enjoyment*, *user feeling*, *collaboration*,

expressiveness, *immersion*, *flexibility*, and *interaction*, as defined in Table 9. The evaluations were also performed in different contexts using a variation of nine products, three approaches, three data collection methods, and two data analyses. The target group's age impacted the choice of the evaluation. The factors, products, and methods distributions differed for children and adults. The factors *novelty*, *enjoyment*, and *collaboration* were more frequently utilized in research targeting children. In contrast, *value*, *fluency*, and *user feeling* were utilized more in the studies with adults as participants. These concepts seem to reveal a difference between more playful factors for children (regarded as valuable in the field (Kawas et al., 2020)) and more work-oriented factors for adults, aligning with differences noted in products as well. *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. While these products align with the factors, it is somewhat surprising as the focus of much of CCI research values design and idea generation, generative activities, and participatory design (Kawas et al., 2020; Yarosh et al., 2011). Children were evaluated more objectively than adults, with more *quantitative physiological measurements* with *observation*, while evaluations targeting adults included a larger amount of *subjective qualitative questionnaires* that employed *self-reporting*.

It is difficult to determine the *credibility* of evaluations of digital CSTs. Most of the included studies did not mention the threats to the study's validity. Many studies did not define creativity, and those studies defined it did not have a uniform perception. Besides, the evaluation of digital CSTs was multi-faceted and did not only focus on just *creativity*. The SLR revealed five categories; *creativity features of the digital CST*, *the perception of users that are utilizing the digital CST*, *the creativity of the process*, *the skills that the users of the digital CST are developing or improving while using the CST*, and *the creativity of the outcome*. These categories overlap with the three aspects of evaluating the creativity of a CST identified in the literature review by Remy et al. (2020); *the usability of the CST*, *the productivity of the process supported by the CST*, and *the creativity of the outcome*. Several of the included studies did not specify which of these aspects they used to evaluate creativity; thus, it is difficult to compare the studies to each other in a more quantitative way.

5.2. Implications

5.2.1. Implications for research

This SLR contributes to researchers by identifying trends and opportunities for future work in the evaluation framework for digital CSTs for children presented in Table 10. This evaluation framework should be investigated further by the CCI community. The article shows a need for more research on this topic by encompassing a wide variety of evaluations of digital CST. There is also a need for more structured evaluations of digital CSTs for children. Remy et al. (2020) recommended in their literature review to develop a toolbox for evaluations of CSTs. Lamb, Brown, and Clarke (2018) highlighted in their interdisciplinary tutorial on evaluating computational creativity the need for such a toolbox — indicating that existing evaluation standards of CSTs should be used to evaluate co-creative systems without citing which evaluation standards should be used.

The present study identified knowledge gaps in the evaluation framework for digital CSTs for children — for example, concerning digital *drawing* and *visualization* CSTs. Further, the SLR identified that promising evaluations for children had not been explored for digital CSTs. One example is the paired-preference evaluation *This or That*, where a child compares two different options and selects which product (s)he prefers most (Zaman, 2009; Zaman, Vanden Abeele, & De Grooff, 2013). According to Guinard (2000), such paired-preference evaluation is the only valid measurement for two to three-year-old children. It can be considered preference ranking for older children comparing more than two products, and acceptance scales between three- to nine-point (Guinard, 2000). One example of such a self-reporting acceptance scale is the Smileyometer (Read, 2008) which comprises a five-point

Likert scale where the response items are replaced by smiley faces representing the user's perceived *enjoyment* level (Read & MacFarlane, 2006). Smileyometer is a broadly used evaluation for children (Read, 2008).

The SLR revealed that not all included studies were related to digital CSTs. There is also an interplay between people and products as creativity happens in a context where children cooperate by using a digital CST to produce a product. If one wants to study the degree of creativity support of a CST, future research should investigate the interplay of creativity between tools, individuals, and products. To do so, there is a need for more interdisciplinary and intersectoral research on creativity and CCI. As the CCI community lacks a common definition of *a digital CST*, we suggest an interactive digital version of Cherry and Latulipe's definition of a CST as "any tool that can be used by people in the open-ended creation of new artifacts" (Cherry & Latulipe, 2014 p. 2). The CCI community also needs to adopt a clear definition of *creativity* since the included studies defined creativity differently. Based on Fig. 4, we propose: *The process of creating something new and valuable*, similar to Sarkar and Chakrabarti's definition: "Creativity occurs through a process by which an agent uses its ability to generate ideas, solutions or products that are novel and valuable" (Sarkar & Chakrabarti, 2008 p. 11). However, further work should look more into these definitions.

5.2.2. Implication for practice

The trends of evaluations of digital CSTs identified in this evaluation framework in Table 10 could enable practitioners in companies who develop digital CSTs for children to understand how to design better solutions. The evaluation framework can allow researchers and practitioners to customize evaluations by choosing which factors and methods to focus on after identifying their product category and the target age of their users. A challenge is that there is no standardized evaluation within the evaluation framework, and we recommend the selected evaluation should be more structured. However, the *physiological measurement* approach with *observation* occurred in 79% of the studies with children and is a promising method for companies. This evaluation is a good starting point for digital CSTs for children 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.

5.2.3. Contributions and recommendations

To summarize, there are several contributions of this SLR that lead to recommendations:

- **Evaluation framework:** This SLR emphasizes the need for researchers and practitioners to find, and use, a common framework for evaluating digital CSTs for children. Our evaluation framework identified trends, knowledge gaps, and opportunities for future work that researchers and practitioners should investigate further, particularly for children but also for adults. Additionally, the framework and overview provide an easy way to cross-reference methods that have been used for particular product domains. We recommend adoption, adaptation, and expansion of our framework.
- **Interdisciplinary and intersectoral research:** The review spotlights the significant need for interdisciplinary and intersectoral research on creativity and CCI to understand how to evaluate digital CSTs for children. We recommend continued – even increased – interdisciplinary research as it relates to developing and evaluating CSTs.
- **Common definitions:** The article pinpoints the need to find common definitions of *digital CSTs* and *creativity* in the CCI community. Further work should investigate our proposed definitions. We recommend adoption and refinement of the definitions we have presented pertaining to the various categories identified in the framework we present.

5.3. Limitations of the review

5.3.1. Completeness

A structured SLR was performed in relevant databases and journals following the guidelines by Kitchenham (2004), which resulted in 81 peer-review articles with 77 different main authors. While this systematic, rigorous approach is standard practice, it does not guarantee that all literature in this area was captured. A potential limitation was the focus of the search could have missed some relevant studies. For example, the term *"assessment"* was not included in the search string in the automatic search, which could lead to missing relevant studies, although our reasoning for not including it was to broaden the scope.

Additionally, some studies of digital CSTs may not clearly identify that the tool being investigated is intended to support creativity. For example, some of the tools mentioned in the review by Tsvyatkova and Storni (2019), like DisCo — a digital tool supporting co-design with children (Walsh et al., 2012), do not refer to the tool directly as a digital CST. Some systems, like narrative systems, may not self-identify as a digital CST (e.g., Howland, Good, and Boulay (2015)'s system that gives narrative support to children's writing). On the other hand, the manual search compensated for this to a certain extent by including 20 more studies using another search string than the automatic search. Nevertheless, some significant HCI journals were missing in the manual search (e.g., *International Journal of Human–Computer Interaction, International Journal of Human–Computer Studies*, and *Behavior & Information Technology*) because the manual search was based on the publication channels found in the automatic search. Studies with adults as participants were included in the review because there is scarce research on evaluating digital CSTs focused on children. This could be due to the mentioned issue regarding some communities of relevant tools and systems not using the term CST. Besides, it is also ethical issues for research with children as participants, related to confidentiality, consent, and protection (Ferdousi, 2015). The inclusion of studies targeting adults allowed comparing the evaluations of digital CSTs to the target group's age in RQ 1.1.3 and see if there were evaluations for adults that could be transformed into evaluations for children.

5.3.2. Potential biases

Even though this SLR did not focus on *usability*, the data synthesis found that many of the included studies focused on this. Multi-faceted evaluations of CSTs that do not only focus on creativity also occurred in the literature review by Remy et al. (2020), where it was a strong tendency that the most recent papers in the review used usability testing to evaluate CSTs. Remy et al. recommended 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 (2011). The SLR revealed that creativity support and usability overlap for some factors (e.g., *value, enjoyment, user feeling*, and *interaction*). Besides, Remy et al. (2020) considered the usability of the CST as one aspect of evaluating the creativity support of a CST. Each of these aspects pointed out by Remy et al. (2020) has benefits for a digital CST. The usability of the digital CST is an essential part of how well the user manages to get creative support from it in the interaction process. Further, the productivity of the process supported by the digital CST is an important part of the creative interaction between the user and the CST. Lastly, the creativity of the outcome is a significant part of the result of this creative interaction. Thus, a good evaluation of a digital CST may include more than one of these three aspects. Commonly used usability evaluations for children's systems such as *This or That* and Smileyometer should be investigated for evaluating digital CSTs for children. There is a need for a more cross-over and holistic evaluation approach. Moreover, some of the studies lacked sufficient details. Many of them did not ground their choice of evaluation of digital CSTs adequately in theory. This aspect also occurred in the literature review by Remy et al. (2020), where less than

half of the included studies' evaluations were built on identifiable theoretical foundations. Remy et al. (2020) recommended using theoretical grounding in the evaluation of the CST. Good examples found in the SLR were Benedetti et al. (2014) and Kantosalo and Riihiaho (2019), which based their selection of evaluation factors on three and nine studies, respectively. 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. Some of the included studies also used the term creativity to define the factor of *expressiveness*.

5.3.3. Data synthesis

The data was categorized to identify themes that answered the RQs. However, there are issues with qualitative data analysis. Some sub-themes could be categorized in the same higher-order theme based on similarity even though they are separated in other literature (e.g., *usefulness* and *intuitiveness* for the factor of *value*). There was also some overlap between the categories. For example, the *self-developed educational digital music game* in the study by Nouwena et al. (2016) was placed in the *music* product category but 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 adequately describe the data to be extracted. To reduce bias, the first author did this categorizing alone and discussed it with the second and third authors.

5.3.4. Broadening inclusion

As indicated in Section 3.1.1, the omission of the term *children* from the SLR was intentional to completely understand the evaluations to better situate relevant work with children within the broader context of research conducted with adults as participants. One of the contributions of this SLR is a better understanding of the language and concepts used to describe and approach evaluations of digital CSTs. As noted in Section 5.3.1, some child research may have been missed because the language used for children is different than that for research for adults. Understanding the terminology of creativity and digital CSTs for children and adults can better help researchers – in CCI and otherwise – to relate their CSTs work within the broader context. For example, one-third of the child-focused papers used the terms *constructivism* and/or *constructionism*, but none of the adult-oriented studies used these terms (although many used the root 'construct'). Thus, this SLR is essential in the CCI research community to set focus on the term digital CST to bring it closer to the creativity research community.

6. Conclusion

This SLR analyzed 81 peer-reviewed conference papers and journal articles from the last decade, aiming to investigate the state-of-the-art procedures for evaluating digital CSTs for children. A wide variation of evaluations was found, but it is difficult to conclude their credibility since a standardized evaluation is lacking. The evaluations were performed in different contexts, impacted by the age of the target group. This research contributes to practitioners and researchers by identifying the current evaluation framework for digital CSTs for children based on 10 factors, nine products, three approaches, and five methods. The research also reveals the trends and areas needing further investigation. Future work should continue to refine common definitions of digital CSTs and creativity for the CCI community and should use and examine this evaluation framework in an interdisciplinary and intersectoral way, aiming to create a common framework to use in evaluating novel digital CSTs.

7. Selection and participation

There were no participants in this systematic literature review.

References

Al-Mousawi, Z., & Alsumait, A. (2012). A digital storytelling tool for arab children. In *Proceedings of the 14th international conference on information integration and web-based applications & services IIWAS '12* (pp. 26–35). New York, USA: ACM. https://doi.org/10.1145/2428736.2428746.

Almeida, L. P., Leandro, S., Prieto Ferrando, M., Oliveira, E., & Ferrandiz, C. (2008). Torrance test of creative thinking: The question of its construct validity. *Thinking Skills and Creativity, 3*, 53–58. https://doi.org/10.1016/j.tsc.2008.03.003

Alves-Oliveira, P., Arriaga, P., Paiva, A. M., & Hoffman, G. (2017). YOLO, a robot for creativity: A co-design study with children. In *Proceedings of the 2017 conference on interaction design and children IDC '17* (pp. 423–429). New York, USA: ACM. https://doi.org/10.1145/3078072.3084304.

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

Amabile, T. M. (1996). *Creativity in context: update to the social psychology of creativity*. Boulder, USA: Westview Press.

Andolina, S., Schneider, H., Chan, J., Klouche, K., Jacucci, G., & Dow, S. (2017). Crowdboard: Augmenting in-person idea generation with real-TimeCrowds. In *Proceedings of the 2017 ACM SIGCHI conference on creativity & cognition C & C '17* (pp. 106–118). New York, USA: ACM. https://doi.org/10.1145/3059454.3059477.

Atwood-Blaine, D., Rule, A. C., & Walker, J. (2019). Creative self-efficacy of children aged 9–14 in a science center using a situated mobile game. *Thinking Skills and Creativity, 33*, Article 100580. https://doi.org/10.1016/j.tsc.2019.100580

Bacciotti, D., Borgianni, Y., & Rotini, F. (2016). A CAD tool to support idea generation in the product planning phase. *Computer Aided Design and Applications, 13*, 490–502. https://doi.org/10.1080/16864360.2015.1131543

Bakala, E., Gerosa, A., Hourcade, J. P., & Tejera, G. (2021). Preschool children, robots, and computational thinking: A systematic review. *International Journal of Child-Computer Interaction, 29*, Article 100337. https://doi.org/10.1016/j.ijcci.2021.100337

Bao, P., Gerber, E., Gergle, D., & Hoffman, D. (2010). Momentum: Getting and staying on topic during a brainstorm. In *Proceedings of the SIGCHI conference on human factors in computing systems CHI '10* (pp. 1233–1236). New York, NY, USA: ACM. https://doi.org/10.1145/1753326.1753511.

Benedetti, L., Winnemöller, H., Corsini, M., & Scopigno, R. (2014). Painting with bob: Assisted creativity for novices. In *Proceedings of the 27th Annual ACM symposium on user interface software and technology UIST '14* (pp. 419–428). New York, USA: ACM. https://doi.org/10.1145/2642918.2647415.

Besemer, S. P., & Treffinger, D. J. (1981). Analysis of creative products: Review and synthesis. *The Journal of Creative Behavior, 15*, 158–178. https://doi.org/10.1002/j.2162-6057.1981.tb00287.x

Cao, X., Lindley, S. E., Helmes, J., & Sellen, A. (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 CSCW '10* (pp. 251–260). New York, USA: ACM. https://doi.org/10.1145/1718918.1718967.

Carroll, E. A., & Latulipe, C. (2012). Triangulating the personal creative experience: Self-report, external judgments, and physiology. In *Proceedings of the graphics interface conference 2012 GI '12* (pp. 53–60). Mississauga, Canada: Canadian Information Processing Society. https://doi.org/10.5555/2305276.2305286.

Carson, S. H., Peterson, J. B., & Higgens, D. M. (2005). Reliability, validity, and factor structure of the creative achievement questionnaire. *Creativity Research Journal, 17*, 37–50. https://doi.org/10.1207/s15326934crj1701_4

Cassell, J., & Ryokai, K. (2001). Making space for voice: Technologies to support childrens fantasy and storytelling. *Personal and Ubiquitous Computing, 5*, 169–190. https://doi.org/10.1007/PL00000018

Catala, A., Jaen, J., van Dijk, B., & Jordà, S. (2012). Exploring tabletops as an effective tool to foster creativity traits. In *Proceedings of the sixth international conference on tangible, embedded and embodied interaction TEI '12* (pp. 143–150). New York, USA: ACM. https://doi.org/10.1145/2148131.2148163.

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

Chang, Y.-s., Chen, M. Y.-C., Chuang, M.-J., & Chou, C.-h. (2019). Improving creative self-efficacy and performance through computer-aided design application. *Thinking Skills and Creativity, 31*, 103–111. https://doi.org/10.1016/j.tsc.2018.11.007

Chang, Y.-S., Chien, Y.-H., Lin, H.-C., Chen, M. Y., & Hsieh, H.-H. (2016). Effects of 3D CAD applications on the design creativity of students with different representational abilities. *Computers in Human Behavior, 65*, 107–113. https://doi.org/10.1016/j.chb.2016.08.0240747-5632

Chang, H.-Y., Lin, H.-C., Wu, T.-T., & Huang, Y.-M. (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 EDUCON '19* (pp. 1087–1092). Piscataway, USA: IEEE. https://doi.org/10.1109/EDUCON.2019.8725233.

Chang, C.-L., & Luh, D.-B. (2012). User as designer: A design model of user creativity platforms. *Journal of Integrated Design and Process Science, 16*, 19–30. https://doi.org/10.3233/jid-2012-0020

Chatain, J., Bitter, O., Fayolle, V., Sumner, R. W., & Magnenat, S. (2019). A creative game design and programming app. In *Motion, interaction and games MIG '19*. New York, USA: ACM. https://doi.org/10.1145/3359566.3360056.

Cherry, E., & Latulipe, C. (2014). Quantifying the creativity support of digital tools through the creativity support index. *ACM Transactions on Computer-Human Interaction*, 21. https://doi.org/10.1145/2617588

Chu Yew Yee, S. L., Quek, F. K., & Xiao, L. (2011). Studying medium effects on children's creative processes. In *Proceedings of the 8th ACM conference on creativity & cognition C & C '11* (pp. 3–12). New York, USA: ACM. https://doi.org/10.1145/2069618.2069622.

Cruzes, D. S., & Dybå, T. (2011). Recommended steps for thematic synthesis in software engineering. In *Proceedings of the 2011 international symposium on empirical software engineering and measurement ESEM '11* (pp. 275–284). Piscataway, USA: IEEE. https://doi.org/10.1109/ESEM.2011.36.

Csikszentmihalyi, M. (1997). *Creativity: Flow and the psychology of discovery and invention.* New York, USA: Harper Perennial.

Davis, N., Hsiao, C.-P., Yashraj Singh, K., Li, L., & Magerko, B. (2016). Empirically studying participatory sense-making in abstract drawing with a co-creative cognitive agent. In *Proceedings of the 21st international conference on intelligent user interfaces IUI '16* (pp. 196–207). New York, USA: ACM. https://doi.org/10.1145/2856767.2856795.

Dow, S. P., Glassco, A., Kass, J., Schwarz, M., Schwartz, D. L., & Klemmer, S. R. (2010). Parallel prototyping leads to better design results, more divergence, and increased self-efficacy. *ACM Transactions on Computer-Human Interaction*, 17. https://doi.org/10.1145/1879831.1879836

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

Dybå, T., & Dingsøyr, T. (2008). Empirical studies of agile software development: A systematic review. *Information and Software Technology, 50*, 833–859. https://doi.org/10.1016/j.infsof.2008.01.006

Feeman, S. M., Wright, L. B., & Salmon, J. L. (2018). Exploration and evaluation of CAD modeling in virtual reality. *Computer Aided Design and Applications, 15*, 892–904. https://doi.org/10.1080/16864360.2018.1462570

Ferdousi, N. (2015). Children as research subjects: The ethical issues. *Bangladesh Journal of Bioethics, 6*, 6–10. https://doi.org/10.3329/bioethics.v6i1.24398

Fessakis, G., Lappas, D., & Mavroudi, E. E. (2015). Could computer gamesbased problem solving positively affect the development of creativity in Young children? A mixed method case study. In K. L. Heider, & M. Renck Jalongo (Eds.), *Young children and families in the information age: applications of technology in early childhood* (pp. 207–225). Dordrecht, Netherlands: Springer Netherlands. https://doi.org/10.1007/978-94-017-9184-7_12.

Fleury, S., Agnès, A., Vanukuru, R., Goumillout, E., Delcombel, N., & Richir, S. (2020). Studying the effects of visual movement on creativity. *Thinking Skills and Creativity, 36*, Article 100661. https://doi.org/10.1016/j.tsc.2020.100661

Frich, J., MacDonald Vermeulen, L., Remy, C., Biskjaer, M. M., & Dalsgaard, P. (2019). Mapping the landscape of creativity support tools in HCI. In *Proceedings of the 2019 CHI conference on human factors in computing systems CHI '19* (pp. 1–18). New York, USA: ACM. https://doi.org/10.1145/3290605.3300619 (Paper 389).

Gaeta, E., Beltrán-Jaunsaras, M. E., Cea, G., Spieler, B., Burton, A., García-Betances, R. I., et al. (2019). Evaluation of the create@school game-based learning–teaching approach. *Sensors, 19*(3251). https://doi.org/10.3390/s19153251

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

Garcia, J., Tsandilas, T., Agon, C., & Mackay, W. E. (2014). Structured observation with polyphony: A multifaceted tool for studying music composition. In *Proceedings of the 2014 conference on designing interactive systems DIS '14* (pp. 199–208). New York, USA: ACM. https://doi.org/10.1145/2598510.2598512.

Giannakos, M. N., & Jaccheri, L. (2013). What motivates children to become creators of digital enriched artifacts?. In *Proceedings of the 9th ACM conference on creativity & cognition C & C '13* (pp. 104–113). New York, USA: ACM. https://doi.org/10.1145/2466627.2466634.

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

Giannakos, M. N., Jaccheri, L., & Proto, R. (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 CSERC'13 (pp. 103–111). Heerlen, Netherland: Open Universiteit. https://doi.org/10.5555/2541917.2541927.

Giannakos, M. N., Papamitsiou, Z., Markopoulos, P., Read, J., & Hourcade, J. P. (2020). Mapping child-computer interaction research through co-word analysis. *International Journal of Child-Computer Interaction, (1001), 2*, 3–2465. https://doi.org/10.1016/j.ijcci.2020.100165

Glăveanu, V. P. (2010). Paradigms in the study of creativity: Introducing the perspective of cultural psychology. *New Ideas in Psychology, 28*, 79–93. https://doi.org/10.1016/j.newideapsych.2009.07.007

Goff, K., & Torrance, E. P. (2002). *Abbreviated torrance test for adults: manual.* Bensenville, USA: Scholastic Testing Service.

Goggin, G. (2011). Ubiquitous apps: politics of openness in global mobile cultures. *Digital Creativity, 22*, 148–159. https://doi.org/10.1080/14626268.2011.603733

Griffin, G., & Jacob, R. (2013). Priming creativity through improvisation on an adaptive musical instrument. In *Proceedings of the 9th ACM conference on creativity & cognition C & C '13* (pp. 146–155). New York, USA: ACM. https://doi.org/10.1145/2466627.2466630.

Grover, S., Basu, S., & Schank, P. (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 SIGCSE '18* (pp. 999–1004). New York, USA: ACM. https://doi.org/10.1145/3159450.3159522.

Guegan, J., Buisine, S., Mantelet, F., Maranzana, N., & Segonds, F. (2016). Avatar-mediated creativity: When embodying inventors makes engineers more creative. *Computers in Human Behavior, 61*, 165–175. https://doi.org/10.1016/j.chb.2016.03.024

Guegan, J., Segonds, F., Barré, J., Maranzana, N., Mantelet, F., & Buisine, S. (2017). Social identity cues to improve creativity and identification in face-to-face and virtual groups. *Computers in Human Behavior, 77*, 140–147. https://doi.org/10.1016/j.chb.2017.08.043

Guilford, J. P. (1950). Creativity. *American Psychologist, 5*, 444–454. https://doi.org/10.1037/h0063487

Guilford, J. P. (1967). *The nature of human intelligence.* New York, USA: McGraw-Hill.

Guinard, J.-X. (2000). Sensory and consumer testing with children. *Trends in Food Science & Technology, 11*, 273–283. https://doi.org/10.1016/S0924-2244(01)00015-2

Güldenpfennig, F., Reitberger, W., & Fitzpatrick, G. (2012). Capturing rich media through media objects on smartphones. In *Proceedings of the 24th Australian computer-human interaction conference OzCHI '12* (pp. 180–183). New York, USA: ACM. https://doi.org/10.1145/2414536.2414569.

Haar Horowitz, A., Grover, I., Reynolds-Cuéllar, P., Breazeal, C., & Maes, P. (2018). Dormio: Interfacing with dreams. In *Proceedings of the SIGCHI conference on human factors in computing systems CHI '18* (pp. 1–10). New York, USA: ACM. https://doi.org/10.1145/3170427.3188403.

Habibian Naeini, F., & Masood, M. (2012). Effect of educational computer games on student creativity. *Research Journal of Applied Sciences, Engineering and Technology, 4*, 5280–5284. https://doi.org/10.1007/s10639-020-10216-1

Halpern, M. K., Tholander, J., Evjen, M., Davis, S., Ehrlich, A., Schustak, K., et al. (2011). Moboogie: Creative expression through whole body musical interaction. In *Proceedings of the SIGCHI conference on human factors in computing systems CHI '11* (pp. 557–560). New York, USA: ACM. https://doi.org/10.1145/1978942.1979020.

Hassenzahl, M., Burmester, M., & Koller, F. (2003). AttrakDiff: Ein fragebogen zur messung wahrgenommener hedonischer und pragmatischer qualitat. In G. Szwillus, & J. Ziegler (Eds.), *Mensch & computer 2003: interaktion in bewegung* (pp. 187–196). Wiesbaden, Germany: Vieweg+Teubner Verlag. https://doi.org/10.1007/978-3-322-80058-9_19.

Hocevar, D. (1981). Measurement of creativity: Review and critique. *Journal of Personality Assessment, 45*, 450–464. https://doi.org/10.1207/s15327752jpa4505_1

Holmes, R. M., Gardner, B., Kohm, K., Bant, C., Ciminello, A., Moedt, K., et al. (2019). The relationship between young children's language abilities, creativity, play, and storytelling, early child development and care. *Early Child Development and Care, 189*, 244–254. https://doi.org/10.1080/03004430.2017.1314274

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

Hourcade, J. P. (2015). *Child-computer interaction.* Iowa, USA: Self.

Howland, K., Good, J., & Boulay, B.du. (2015). Narrative support for young game designers' writing. In *Proceedings of the 14th international conference on interaction design and children IDC '15* (pp. 178–187). New York, USA: Association for Computing Machinery. https://doi.org/10.1145/2771839.2771858.

Hsiao, H.-S., Wong, K.-H., Wang, M.-J., Yu, K.-C., Chang, K.-E., & Sung, Y. T. (2006). Using cognitive affective interaction model to construct on-line game for creativity. In *First international conference on technologies for elearning and digital entertainment Edutainment'06* (pp. 409–418). Hangzhou, China: Technologies for E-Learning and Digital Entertainment. https://doi.org/10.1007/11736639_52.

Huang, C.-Z. A., Duvenaud, D., & Gajos, K. Z. (2016). ChordRipple: Recommending chords to help novice composers go beyond the ordinary. In *Proceedings of the 21st international conference on intelligent user interfaces, IUI '16* (pp. 241–250). New York, USA: ACM. https://doi.org/10.1145/2856767.2856792.

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

ISO/IEC 25010:2011. (2011). *Systems and software engineering – systems and software quality requirements and evaluation (square) – system and software quality models.* Switzerland: Standard International Organization for Standardization Geneva. URL: https://webstore.iec.ch/publication/11245.

Jordanous, A., & Keller, B. (2016). Modelling creativity: Identifying key components through a corpus-based approach. *PLoS One, 11*, Article e0162959. https://doi.org/10.1371/journal.pone.0162959

Kahn, P. H., Kanda, T., Ishiguro, H., Gill, B. T., Shen, S., Ruckert, J. H., et al. (2016). Human creativity can be facilitated through interacting with a social robot. In *The eleventh ACM/IEEE international conference on human robot interaction HRI '16* (pp. 173–180). Piscataway, USA: IEEE Press. https://doi.org/10.1109/HRI.2016.7451749.

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

Karakaya, A. F., & Demirkan, H. (2015). Collaborative digital environments to enhance the creativity of designers. *Computers in Human Behavior, 42*, 176–186. https://doi.org/10.1016/j.chb.2014.03.029

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

Kato, J., Nakano, T., & Goto, M. (2015). TextAlive: Integrated design environment for kinetic typography. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems CHI '15* (pp. 3403–3412). New York, USA: ACM. https://doi.org/10.1145/2702123.2702140.

Kaufman, J. C. (2012). Counting the muses: Development of the kaufman domains of creativity scale (k-DOCS). *Psychology of Aesthetics, Creativity, and the Arts, 6*, 298–308. https://doi.org/10.1037/a0029751

Kaufman, J. C., & Beghetto, R. A. (2009). Beyond big and little: The four c model of creativity. *Review of General Psychology, 13*, 1–12. https://doi.org/10.1037/a0013688

Kawas, S., Yuan, Y., DeWitt, A., Jin, Q., Kirchner, S., Bilger, A., et al. (2020). Another decade of IDC research: Examining and reflecting on values and ethics. In *Proceedings of the interaction design and children conference IDC '20* (pp. 205–215). New York, USA: Association for Computing Machinery. https://doi.org/10.1145/3392063.3394436.

Kazi, R. H., Chua, K. C., Zhao, S., Davis, R., & Low, K.-L. (2011). Sandcanvas: A multi-touch art medium inspired by sand animation. In *Proceedings of the SIGCHI conference on human factors in computing systems CHI '11* (pp. 1283–1292). New York, USA: ACM. https://doi.org/10.1145/1978942.1979133.

Kerne, A., Webb, A. M., Smith, S. M., Linder, R., Lupfer, N., Qu, Y., et al. (2014). Using metrics of curation to evaluate InformationBased ideation. *ACM Transactions on Computer-Human Interaction, 21*. https://doi.org/10.1145/2591677

Kim, S., Chung, K., & Yu, H. (2013). Enhancing digital fluency through a training program for creative problem solving using computer programming. *The Journal of Creative Behavior, 47*, 171–199. https://doi.org/10.1002/jocb.30

Kitchenham, B. A. (2004). *Procedures for performing systematic reviews: Technical report TR/SE-0401*. UK: Department of Computer Science Keele University. URL: https://www.bibsonomy.org/bibtex/2e48137ec01b6308876e05ab1ecdf4bc4/wiljami74.

Koh, K. H., Bennett, V., & Repenning, A. (2011). Computing indicators of creativity. In *Proceedings of the 8th ACM conference on creativity & cognition C & C '11* (pp. 357–358). New York, USA: ACM. https://doi.org/10.1145/2069618.2069694.

Kucirkovaaa, N., & Sakr, M. (2015). Child–father creative text-making at home with crayons, ipad collage & PC. *Thinking Skills and Creativity, 17*, 59–73. https://doi.org/10.1016/j.tsc.2015.05.003

Lamb, C., Brown, D. G., & Clarke, C. L. A. (2018). Evaluating computational creativity: An interdisciplinary tutorial. *ACM Computing Surveys, 51*. https://doi.org/10.1145/3167476

Lu, F., Tian, F., Jiang, Y., Cao, X., Luo, W., Li, G., et al. (2011). ShadowStory: Creative and collaborative digital storytelling inspired by cultural heritage. In *Proceedings of the SIGCHI conference on human factors in computing systems CHI '11* (pp. 1919–1928). New York, USA: ACM. https://doi.org/10.1145/1978942.1979221.

Maiden, N., Zachos, K., Brown, A., Brock, G., Nyre, L., Tonheim, A. Nygård., et al. (2018). Making the news: Digital creativity support for journalists. In *Proceedings of the 2018 CHI conference on human factors in computing systems CHI '18* (pp. 1–11). New York, USA: ACM. https://doi.org/10.1145/3173574.3174049.

Mangano, N., Baker, A., Dempsey, M., Navarro, E., & van der Hoek, A. (2010). Software design sketching with calico. In *Proceedings of the IEEE/ACM international conference on automated software engineering ASE '10* (pp. 23–32). New York, USA: ACM. https://doi.org/10.1145/1858996.1859003.

Martin, C., Gardner, H., Swift, B., & Martin, M. (2016). Intelligent agents and networked buttons improve free-improvised ensemble music-making on touch-screens. In *Proceedings of the 2016 CHI conference on human factors in computing systems CHI '16* (pp. 2295–2306). New York, USA: ACM. https://doi.org/10.1145/2858036.2858269.

McDaniel, R., Fanfarelli, J. R., & Lindgren, R. (2017). Creative content management: Importance, novelty, and affect as design heuristics for learning management systems. *IEEE Transactions on Professional Communication, 60*, 183–200. https://doi.org/10.1109/TPC.2017.2656578

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

McKay, K. M., Alexander, S., & Kaufman, J. C. (2017). Measuring the muses: Validating the kaufman domains of creativity scale (k-DOCS). *Psychology of Aesthetics, Creativity, the Arts, 11*, 216–230. https://doi.org/10.1037/aca0000074

Mednick, M. T., Mednick, S. A., & Mednick, E. V. (1964). Incubation of creative performance and specific associative priming. *Journal of Abnormal and Social Psychology, 69*, 84–88. https://doi.org/10.1037/h0045994

Mohamed, F., Noor, N. N. M., Farhana, R., & Khairuddin, R. (2020). Simple ideas creativity of pre-service teachers in constructing game materials for learning plant diversity of Malaysia. *Journal of Critical Reviews, 7*, 511–515. https://doi.org/10.31838/jcr.07.06.94

Moran, J. D., III, Milgram, R. M., Sawyers, J. K., & Fu, V. R. (1983). Original thinking in preschool children. *Child Development, 54*, 921–926. https://doi.org/10.2307/1129896

Myers, B. A., Lai, A., Le, T. M., Yoon, Y., Faulring, A., & Brandt, J. (2015). Selective undo support for painting applications. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems CHI '15* (pp. 4227–4236). New York, USA: ACM. https://doi.org/10.1145/2702123.2702543.

Nakakoji, K. (2005). Seven issues for creativity support tool researchers. In M. Resnick (Ed.), *National science foundation workshop report on creativity support tools* (pp. 67–70). URL: http://www.cs.umd.edu/hcil/CST (chapter 7).

Nakazato, N., Yoshida, S., Sakurai, S., Narumi, T., Tanikawa, T., & Hirose, M. (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 CSCW '14* (pp. 75–83). New York, USA: ACM. https://doi.org/10.1145/2531602.2531637.

Nelson, C., Brummel, B. J., Grove, F., Jorgenson, N., Sen, S., & Gamble, R. (2010). Measuring creativity in software development. In *Proceedings of the international conference on computational creativity ICCC '10* (pp. 205–214). Lisbon, Portugal: Association for Computational Creativity.

Nouwena, M., Schepers, S., Mouws, K., Slegers, K., Kosten, N., & Duysburgh, P. (2016). Designing an educational music game: What if children were calling the tune? *International Journal of Child-Computer Interaction, 9–10*, 20–32. https://doi.org/10.1016/j.ijcci.2016.10.001

Oh, H., Harriman, A., Narula, A., Gross, M. D., Eisenberg, M., & Hsi, S. (2016). Crafting mechatronic percussion with everyday materials. In *Proceedings of the tenth international conference on tangible, embedded, and embodied interaction TEI '16* (pp. 340–348). New York, USA: ACM. https://doi.org/10.1145/2839462.2839474.

Oxford Dictionary. (2021). *Oxford learner's dictionary*. URL: https://www.oxfordlearnersdictionaries.com/definition/english (Visited May 25, 2021).

Papavlasopoulou, S., Giannakos, M. N., & Jaccheri, L. (2017). Empirical studies on the maker movement, a promising approach to learning: A literature review. *Entertainment Computing, 18*, 57–78. https://doi.org/10.1016/j.entcom.2016.09.002

Papert, S. (1980). *Mindstorms: children, computers, and powerful ideas*. New York, USA: Basic Books, Inc.

Piotrowski, J. T., & Meester, L. (2018). Can apps support creativity in middle childhood? *Computers in Human Behavior, 85*, 23–33. https://doi.org/10.1016/j.chb.2018.03.030

Quayyum, F., Cruzes, D. S., & Jaccheri, L. (2021). Cybersecurity awareness for children: A systematic literature review. *International Journal of Child-Computer Interaction, 30*, Article 100343. https://doi.org/10.1016/j.ijcci.2021.100343

Read, J. C. (2008). Validating the fun toolkit: an instrument for measuring children's opinions of technology. *Cognition Technology and Work, 10*, 119–128. https://doi.org/10.1007/s10111-007-0069-9

Read, J. C., Fitton, D., & Horton, M. (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 IDC '14* (pp. 105–114). New York, USA: ACM. https://doi.org/10.1145/2593968.2593986.

Read, J. C., & MacFarlane, S. (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 IDC '06* (pp. 81–88). New York, USA: ACM Press. https://doi.org/10.1145/1139073.1139096.

Read, J. C., MacFarlane, S., & Casey, C. (2002). Endurability, engagement and expectations: Measuring children's fun. In *Proceedings of interaction design and children 2 IDC'02* (pp. 1–23). New York, USA: ACM. URL: https://www.researchgate.net/publication/228870976_Endurability_Engagement_and_Expectations_Measuring_Childrenaposs_Fun.

Read, J. C., & Markopoulos, P. P. (2013). Child–computer interaction. *International Journal of Child-Computer Interaction, 1*, 2–6. https://doi.org/10.1016/j.ijcci.2012.09.001

Remy, C., MacDonald Vermeulen, L., Frich, J., Biskjaer, M. M., & Dalsgaard, P. (2020). Evaluating creativity support tools in HCI research. In *Proceedings of the 2020 ACM designing interactive systems conference DIS '20* (pp. 457–476). New York, USA: ACM. https://doi.org/10.1145/3357236.3395474.

Ritter, S. M., & Rietzschel, E. F. (2004). Lay theories of creativity. In C. M. Zedelius, B. C. N. Muller, & J. W. Schooler (Eds.), *The Science of Lay Theories* (pp. 95–126). Cham, Germany: Springer. https://doi.org/10.1007/978-3-319-57306-9_5.

Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity Research Journal, 24*, 92–96. https://doi.org/10.1080/10400419.2012.650092

Sapounidis, T., Demetriadis, S., Papadopoulos, P. M., & Stamovlasis, D. (2019). Tangible and graphical programming with experienced children: A mixed methods analysis. *International Journal of Child-Computer Interaction, 19*, 67–78. https://doi.org/10.1016/j.ijcci.2018.12.001

Sarkar, P., & Chakrabarti, A. (2008). Studying engineering design creativity developing a common definition and associated measures. In J. Gero (Ed.), *Invited paper in the proceedings of the nsf workshop on studying design creativity*. Alexandria, USA: National Science Foundation.. URL: https://cpdm.iisc.ac.in/cpdm/ideaslab/paper_scans/UID_17.pdf.

Sarkar, P., & Chakrabarti, A. (2011). Assessing design creativity. *Design Studies, 32*, 348–383. https://doi.org/10.1016/j.destud.2011.01.002

Sarooghi, H., Libaers, D., & Burkemper, A. (2015). Examining the relationship between creativity and innovation: A meta-analysis of organizational, cultural, and

environmental factors. *Journal of Business Venturing, 30*, 714–731. https://doi.org/10.1016/j.jbusvent.2014.12.003

Savransky, S. (2000). *Engineering of creativity: introduction to triz methodology of inventive problem solving*. Boca Raton, USA: CRC Press.

Schmitt, L., Buisine, S., Chaboissier, J., Aoussat, A., & Vernier. (2012). Dynamic tabletop interfaces for increasing creativity. *Computers in Human Behavior, 28*, 1892–1901. https://doi.org/10.1016/j.chb.2012.05.007

Sharma, K., & Giannakos, M. (2021). Sensing technologies and child-computer child–computer interaction: Opportunities, challenges and ethical considerations. *International Journal of Child-Computer Interaction, 30*, Article 100331. https://doi.org/10.1016/j.ijcci.2021.100331

Shugrina, M., Lu, J., & Diverdi, S. (2017). Playful palette: An interactive parametric color mixer for artists. *ACM Transactions on Graphics*, 36. https://doi.org/10.1145/3072959.3073690

Siangliulue, P., Chan, J., Gajos, K. Z., & Dow, S. P. (2015). Providing timely examples improves the quantity and quality of generated ideas. In *Proceedings of the 2015 ACM SIGCHI conference on creativity & cognition C & C '15* (pp. 83–92). New York, USA: ACM. https://doi.org/10.1145/2757226.2757230.

Singh, Y. K. (2006). *Fundamental of research methodology and statistics*. New Delhi, India: Newage International (P) Ltd. Publisher.

Star, K., Paraskevopoulos, F., Taramigkou, M., Apostolou, D., Schot, M., & Mentzas, G. N. (2015). A playful affinity space for creative research. In *Proceedings of the 2015 ACM SIGCHI conference on creativity & cognition C & C '15* (pp. 107–110). New York, USA: ACM. https://doi.org/10.1145/2757226.2757244.

Sternberg, R. J., & Lubart, T. I. (1999). The concept of creativity: Prospects and paradigms. In R. J. Sternberg (Ed.), *Handbook of creativity* (pp. 3–15). Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511807916.003 (chapter 1).

Sylla, C., Coutinhoa, C., Branco, P., & Muller, W. (2015). Investigating the use of digital manipulatives for storytelling in pre-school. *International Journal of Child-Computer Interaction, 6*, 39–48. https://doi.org/10.1016/j.ijcci.2015.10.001

Sylla, C., Pereira, I. S. P., Brooks, E., & Zagalo, N. (2018). T-books: A block interface for young children's narrative construction. *International Journal of Child-Computer Interaction, 18*, 59–66. https://doi.org/10.1016/j.ijcci.2018.07.002

Theodoropoulos, A., & Lepouras, G. (2021). Augmented reality and programming education: A systematic review. *International Journal of Child-Computer Interaction, 30*, Article 100335. https://doi.org/10.1016/j.ijcci.2021.100335

Thew, S. L., Sutcliffe, A. G., De Bruijn, O., McNaught, J., Procter, R. N., Jarvis, P., et al. (2011). Supporting creativity and appreciation of uncertainty in exploring geo-coded public health data. *Methods of Information in Medicine, 50*, 158–165. https://doi.org/10.3414/ME09-01-0070

Torrance, E. P. (1966). *Torrance tests of creative thinking: norms technical manual* (Research ed.). Princeton, USA: Personnel Press.

Torrance, E. P. (1988). The nature of creativity as manifest in its testing. In R. J. Sternberg (Ed.), *The nature of creativity: contemporary psychological perspectives* (pp. 43–75). New York, USA: Cambridge University Press (chapter 2).

Torrance, E. P. (1998). *The torrance tests of creative thinking: directions manual figural (streamlined) forms A & B*. Bensenville, USA: Scholastic Testing Service.

Torres, C., Li, W., & Paulos, E. (2016). ProxyPrint: Supporting crafting practice through physical computational proxies. In *Proceedings of the 2016 ACM conference on designing interactive systems DIS '16* (pp. 158–169). New York, USA: ACM. https://doi.org/10.1145/2901790.2901828.

Torres, C., & Paulos, E. (2015). MetaMorphe: Designing expressive 3D models for digital fabrication. In *Proceedings of the 2015 ACM SIGCHI conference on creativity & cognition C & C '15* (pp. 73–82). New York, USA: ACM. https://doi.org/10.1145/2757226.2757235.

Tsvyatkova, D., & Storni, C. (2019). A review of selected methods, techniques and tools in child–computer interaction (CCI) developed/adapted to support children's involvement in technology development. *International Journal of Child-Computer Interaction, 22*, Article 100148. https://doi.org/10.1016/j.ijcci.2019.100148

United Nations. (1990). *Convention on the rights of the child*. URL: https://www.ohchr.org/en/professionalinterest/pages/crc.aspx (Visited September 23, 2020).

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

Verbruggen, S., Depaepe, F., & Torbeyns, J. (2021). Effectiveness of educational technology in early mathematics education: A systematic literature review. *International Journal of Child-Computer Interaction, 27*, Article 100220. https://doi.org/10.1016/j.ijcci.2020.100220

Vidal, R. (2012). To be human is to be creative. *AI & Society*, 28. https://doi.org/10.1007/s00146-012-0415-1

Vissers, J., De Bot, L., & Zaman, B. (2013). Memoline: evaluating long-term UX with children. In *Proceedings of the 12th international conference on interaction design and children IDC '13* (pp. 285–288). New York, USA: ACM. https://doi.org/10.1145/2485760.2485836.

Voigt, M., Plattfaut, R., Ortbach, K., Malsbender, A., Niehaves, B., & Becker, J. (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 PACIS 13'*. Atlanta, USA: URL:.

Voiskounsky, A. E., Yermolova, T. D., Yagolkovskiy, S. R., & Khromova, V. M. (2017). Creativity in online gaming: Individual and dyadic performance in minecraft. *Psychology in Russia: State of the Art, 10*, 144–161. https://doi.org/10.11621/pir.2017.0413

Vygotsky, L. S. (2004). Imagination and creativity in childhood. *Journal of Russian and East European Psychology, 42*, 7–97. https://doi.org/10.1080/10610405.2004.11059210

Walsh, G., Druin, A., Guha, M. L., Bonsignore, E., Foss, E., Yip, J. C., et al. (2012). Disco: A co-design online tool for asynchronous distributed child and adult design partners. In *Proceedings of the 11th international conference on interaction design and children IDC '12* (pp. 11–19). New York, USA: Association for Computing Machinery. https://doi.org/10.1145/2307096.2307099.

Wang, H.-C., Cosley, D., & Fussell, S. R. (2010). Idea expander: Supporting group brainstorming with conversationally triggered visual thinking stimuli. In *Proceedings of the 2010 ACM conference on computer supported cooperative work CSCW '10* (pp. 103–106). New York, USA: ACM. https://doi.org/10.1145/1718918.1718938.

Wang, Y., Gao, Y., & Lian, Z. (2020). Attribute2Font: Creating fonts you want from attributes. *ACM Transactions on Graphics, 39*. https://doi.org/10.1145/3386569.3392456

Wang, T., Huo, K., Chawla, P., Chen, G., Banerjee, S., & Ramani, K. (2018). Plain2Fun: Augmenting ordinary objects with interactive functions by AutoFabricating surface painted circuits. In *Proceedings of the 2018 designing interactive systems conference DIS '18* (pp. 1095–1106). New York, USA: ACM. https://doi.org/10.1145/3196709.3196791.

Weisberg, R. W. (2015). On the usefulness of value in the definition of creativity. *Creativity Research Journal, 27*, 111–124. https://doi.org/10.1080/10400419.2015.1030320

Willis, K. D., Lin, J., Mitani, J., & Igarashi, T. (2010). Spatial sketch: Bridging between movement & fabrication. In *Proceedings of the fourth international conference on tangible, embedded, and embodied interaction TEI '10* (pp. 5–12). New York, USA: ACM. https://doi.org/10.1145/1709886.1709890.

Witt, M., & Robra-Bissantz, S. (2012). Sparking motivation and creativity with online ideation games. In S. Goltz, M. Magnor, H.-J. Appelrath, H. K. Matthies, W.-T. Balke, & L. Wolf (Eds.), *lecture notes in informaticsINFORMATIK 2012* (pp. 1006–1023). Bonn, Germany: Gesellschaft für Informatik e.V.. URL: https://subs.emis.de/LNI/Proceedings/Proceedings208/1006.pdf.

Wojtczuk, A., & Bonnardel, N. (2011). Designing and assessing everyday objects: Impact of externalisation tools and judges' backgrounds. *Interactive Computing, 23*, 337–345. https://doi.org/10.1016/j.intcom.2011.05.004

Yang, X., Lin, L., Cheng, P.-Y., Yang, X., & Ren, Y. (2019). Which EEG feedback works better for creativity performance in immersive virtual reality: The reminder or encouraging feedback? *Computers in Human Behavior, 99*, 345–351. https://doi.org/10.1016/j.chb.2019.06.002

Yarosh, S., Radu, I., Hunter, S., & Rosenbaum, E. (2011). Examining values: an analysis of nine years of IDC research. In *Proceedings of the 10th international conference on interaction design and children IDC '11* (pp. 136–144). New York, USA: ACM. https://doi.org/10.1145/1999030.1999046.

Yoon, S. H., Verma, A., Peppler, K., & Ramani, K. (2015). Handimate: Exploring a modular robotics kit for animating crafted toys. In *Proceedings of the 14th international conference on interaction design and children IDC '15* (pp. 11–20). New York, USA: ACM. https://doi.org/10.1145/2771839.2771841.

Yoshida, N., Fukushima, S., Aida, D., & Naemura, T. (2016). Practical study of positive-feedback button for brainstorming with interjection sound effects. In *Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems CHI EA '16* (pp. 1322–1328). New York, USA: ACM. https://doi.org/10.1145/2851581.2892418.

Yu, L., & Nickerson, J. V. (2011). Cooks or cobblers? Crowd creativity through combination. In *Proceedings of the SIGCHI conference on human factors in computing systems CHI '11* (pp. 1393–1402). New York, USA: ACM. https://doi.org/10.1145/1978942.1979147.

Zaman, B. (2009). Introducing a pairwise comparison scale for UX evaluations with preschoolers. In T. Gross, J. Gulliksen, P. Kotzé, L. Oestreicher, P. Palanque, R. O. Prates, & M. Winckler (Eds.), *Lecture notes in computer science: Vol. 5727. Human-computer interaction – INTERACT 2009. INTERACT 2009* (pp. 634–637). Berlin, Germany: Springer. https://doi.org/10.1007/978-3-642-03658-3_68.

Zaman, B., Vanden Abeele, V., & De Grooff, D. (2013). Measuring product liking in preschool children: An evaluation of the smileyometer and this or that methods. *International Journal of Child-Computer Interaction, 1*, 61–70. https://doi.org/10.1016/j.ijcci.2012.12.001

Zhang, Y., Han, T., Ren, Z., Umetani, N., Tong, X., Liu, Y., et al. (2013). BodyAvatar: Creating freeform 3D avatars using FirstPerson body gestures. In *Proceedings of the 26th annual ACM symposium on user interface software and technology UIST '13* (pp. 387–396). New York, USA: ACM. https://doi.org/10.1145/2501988.2502015.

Zheng, C., Do, E. Y.-L., & Budd, J. (2017). Joinery: Parametric joint generation for laser cut assemblies. In *Proceedings of the 2017 ACM SIGCHI conference on creativity & cognition C & C '17* (pp. 63–74). New York, USA: ACM. https://doi.org/10.1145/3059454.3059459.