Consensual Assessment in the New Domain of E-Textiles: Comparing Insights from Expert, Quasi-Expert, and Novice Judges

Anthony Phonethibsavads, Ph.D., University of California, Irvine, aphoneth@uci.edu

Maggie Dahn, Ph.D., University of California, Irvine, dahnm@uci.edu

Kylie Peppler, Ph.D., University of California, Irvine, kpeppler@uci.edu

Deborah Fields, Ph.D., Utah State University, deborah.fields@usu.edu

Yasmin Kafai, Ed.D., University of Pennsylvania, kafai@upenn.edu

**Consensual Assessment in the New Domain of E-Textiles: Comparing Insights
from Expert, Quasi-Expert, and Novice Judges**

Creativity as a construct continues to capture the collective imagination of the
educational and social sciences research communities, yet critical questions remain about how to
best define and measure creativity (Barbot & Reiter-Palman, 2019; Plucker et al., 2004). The
Consensual Assessment Technique (CAT; Amabile, 1982), which has been widely regarded as a
gold standard for assessing creativity (e.g., Baer & McKool, 2009; Carson, 2006; Kaufman &
Baer, 2008), relies on creativity ratings of expert judges who use their subjective viewpoints to
rate creative products (Amabile, 1982) without use of rubrics or operational definitions of
creativity (Mayer, 1999). Although the psychometric validity of CAT has recently been called
into question (Myszkowski & Storme, 2019), CAT remains a core method shared among
creativity researchers who have shaped its implementation over the years (Cseh & Jeffries,
2019).

As CAT continues to be a central method for assessing creativity, a persistent question
across CAT research is how the panel of judges is assembled (e.g., Hickey, 2001, Kaufman et al.,
2008). Because domain experts (e.g., critics, curators) are the arbiters who decide which products
should be preserved, these experts simultaneously determine what characteristics and values
constitute creativity (Csikszenetmihalyi, 1988; Kaufman & Baer, 2012). Additionally, because
the CAT is predicated on consensus among judges, the technique only works if judges are
assessing the same construct (e.g., implicit criteria for creativity), because assessments based on
incomplete representations would impede decision-making (Plucker & Makel, 2010). Therefore,
valid inferences from the CAT are predicated on the idea that inter-rater agreement represents the
judgments and values of the domain gatekeepers, which spotlights the importance of defining the
expertise levels of raters (Hickey, 2001; Kaufman & Baer, 2012; Phonethibsavads et al., 2019).
Although domain experts are usually thought to be the most reliable judges for creativity, some
emergent domains may be lacking in history and precedent, so persons with expert-level
qualifications may be scarce (Csikszentmihalyi, 1988; Glăveanu, 2010). Additionally, because
the scarcity of experts in emergent domains can make creativity research more difficult, we
explore two possible alternatives: First, we explore the extent that quasi-experts may be suitable
substitutes for experts; and second, we examine the efficacy of novice judges, who might yield
valid inferences about general audience opinions of creativity, despite lacking domain
knowledge.

Our purpose in this paper is to suggest how the field can effectively use CAT in judging
creative products as it is applied in an emergent domain in which creativity is valued—in this
case specifically, electronic textiles (hereafter: e-textiles). Importantly for our purposes here, in
their review of CAT studies, Kaufman and Baer (2012) suggest that in addition to experts,
novices and quasi-experts might also be employed in rating creative products using CAT when
reliability and validity can be shown to exist among raters. Typically, showing agreement within
groups of expert raters is a core, defining feature of CAT, as inter-rater reliability is commonly
reported in tandem with results of those ratings. Therefore, valid inferences across varying levels
of expertise would require showing agreement across groups. Baer and Kaufman explain, "The
process would be comparable as that to certify novices; one would need to show that the
creativity ratings of the quasi-experts were sufficiently similar to the ratings of experts in the
domain" (p. 89). In this paper, we pursue questions of reliability and validity across groups of
expert, quasi-expert, and novice assessments of creativity within the new domain of e-textiles

and thus spark conversation and debate as to how rater expertise is relevant to CAT in an emergent domain such as e-textiles. Our analyses show high inter-rater agreement within and across groups in independent ratings of e-textile products. These findings have implications for how raters with varying expertise might be used in future studies of creativity in emerging domains.

Background

The New Domain of E-Textiles

This study strives toward a better understanding of judges' criteria for creativity in a new type of tangible, programmable media called e-textiles, and furthermore, how their criteria may vary across expertise. E-textiles include designs of programmable garments, accessories, and costumes, which incorporate elements of circuitry, embedded computing, and novel materials such as conductive fibers, Velcro, sensors for light, touch, or sound, and actuators such as LEDs or speakers (Buechley, 2006). Though electronics and textiles have long been integrated in various forms through innovations in materials such as armor and metal-wrapped threads (Peppler & Glosson, 2013b), e-textiles as we know them today began to take hold in the mid 1990s when researchers from MIT developed wearable computers, which allowed for hardware to be more closely integrated with the human body (Starner, 2002). Naturally, because clothes are universally worn on human bodies, the integration of electronics with textiles afforded a more seamless synthesis between computers and bodies (etechtex, 2017). Over time, e-textiles garnered attention within education research and practice through the invention of the LilyPad Arduino, an electronic toolkit microcontroller that made the making of e-textiles more accessible to a general audience by supporting the creation of e-textile artifacts by novices and experts alike (Buechley et al., 2008). Engagement with e-textiles can connect to a variety of learning outcomes, including conceptual understanding of basic circuitry concepts such as current flow, polarity, and connections (Peppler & Glosson, 2013a), programming concepts (Keune & Peppler, 2019) as well as artistic outcomes of expression and design.

As a new emerging domain, e-textiles contribute to the advancement of a new digital media e-textiles field while also enriching disciplinary foundations in crafts, circuitry, and code. Csikszentmihalyi (2014) explains that an "environment has two salient aspects: A cultural, or symbolic aspect which here is called the *domain;* and a social aspect called the *field*. Creativity is a process that can be observed only at the intersection where individuals, domains, and fields interact" (p. 103). Essentially, there is a mutual influence between creators and audiences in which audiences recognize and evaluate novel products, and creators work within those standards while stretching boundaries. This field component implies that colleagues and domain norms are essential to the realization of individual creativity. Such a view removes the aura of mystery around creativity and instead, emphasizes the importance of sustained discussion with peers and the need for an appreciation of the constraints that one is augmenting or violating while producing a creative contribution.

While prior creativity research has focused on eminent individuals who revolutionized their respective domains (e.g., Csikszentmihalyi, 1996; Simonton, 1994) or mental processes in problem-solving or artistic creation (Guilford, 1967; Sawyer, 2006b; Torrance, 1972; Wallas, 1926), we have chosen to situate our work in a sociocultural framework (Csikszentmihalyi, 1988; Glăveanu, 2010) where individuals build upon culturally valued practices, meanings, and designs to produce new variations of the domain, which, if deemed valuable by the field, become part of the canon in the domain's evolving standards and conventions. This paper describes our efforts to understand the recognition and evaluation of creativity within e-textile productions and

how e-textile experts and those with limited or no experience evaluated creativity of individual e-textile artifacts displayed in an online gallery by employing the Consensual Assessment Technique (Amabile, 1982). We were interested in how consistent (or divergent) experts, quasi-experts, and novices were in their ratings of creativity and the rationales they employed in judging the e-textile products.

Applying CAT in New Domains

Our work is particularly important as new domains emerge and evolve and creativity moves into online communities, where novices and experts assess creative works based on their previous experiences and preferences (Shirky, 2008). By contrast, in art museums for example, expert curators choose the works of art that are to be included in the museum exhibits and, by consequence, shape the standards and values of the domain by critics and the public. While traditionally creative domains have been guarded by a select few (i.e., experts in the field), the emergence of Web 2.0 has in many ways democratized the curatorial aspects of creative domains by enabling anyone to be active participants and assess the quality of online contributions through comment postings, rating systems, and other forms of monitoring online activity, thus shaping the very media people consume (Jenkins, 2007). In their study of a social media component of a virtual gaming environment, Kafai and Peppler (2011) found that ideas that were "spreadable" on social media dynamically evolved along with the community's interests and through conversation among community members. As social media and online communities change the way ideas are taken up and spread across a particular community, the question of how expertise and creativity are defined in these communities becomes a valuable point of consideration for understanding new and emergent domains. (Peppler & Solomou, 2011; Peppler & Glosson, 2013b).

To better understand how creativity is defined across domains (especially as they evolve), researchers of creativity have argued for applying techniques that fit well with Csikszentmihalyi's sociocultural paradigm for creativity, including the Consensual Assessment Technique (CAT), which was pioneered by Amabile (1982) and has been tested by various psychologists interested in creativity (e.g., Kaufman et al. 2008). This technique allows for the assessment of products' creativity by domain experts because knowledgeable persons in a field are the arbiters who determine what constitutes creativity (Kaufman & Baer, 2012). For example, Nobel Prizes, Pulitzers, and Oscars are decided upon by expert committees in their respective creative domains. Similarly, the CAT differs from psychometric batteries because, while testmakers often construct operational definitions to streamline how constructs are defined, with CAT, the panel is instructed to use their own subjective definition of creativity as they rate artifacts. Important to this technique, experts typically rate creativity relative to the other projects in the sample rather than against some ultimate norm. Historically, the CAT instrument has been used by experts to assess parallel products (artifacts produced via a standardized prompt with similar materials). However, here we build on prior research that established the validity of using the CAT in rating non-parallel creative products (Baer et al. 2004).

In the original paper outlining CAT, Amabile (1982) wrote, "it may be difficult to apply this assessment technique to products that are at the frontiers of a particular domain" (p. 1010), suggesting that there may not be enough shared expertise to constitute what is truly novel and useful within domains on the fringe or cutting-edge of a field. Here we test Amabile's suggestion by enlisting domain pioneers in e-textiles as judges. Still, it is worth asking, whose opinions count for creativity (Kaufman & Baer, 2012), especially in an emergent domain? In the domain of e-textiles, there is much we have to learn about how creativity is defined and manifests in

creative products through a sociocultural perspective.

Few studies have focused on assessing creativity in the context of new or emergent domains, especially in digital or media art. Exceptions include Peppler and colleagues' previous initial explorations of applying creativity assessments within the domain of e-textiles (Peppler & Solomou, 2011; Tan & Peppler, 2015). An additional example of assessing creativity in a less established domain includes Phonethibsavads's (2020) work on creativity in improvisational theater. Though the popular brand of improv theater as it is known today has existed for over 40 years, it remains a niche interest, and therefore, its inner workings are novel and unfamiliar to the majority of the general public. In this study, Phonethibsavads noted that there were key differences in the ways that expert and novice judges appraised creativity: experts scored on the basis performance technique (e.g., decisive choices, commitment to one's character, emergent funny patterns), novices scored on the basis of scene content (e.g., story, character, premise) and the justification of gimmicks (e.g., audience participants filling in lines of dialogue for the performers). Though experts favored performances that demonstrated strong fundamentals, novices favored performances that included gimmicks because they appeared more "impressive" to them, so they gave considerably higher scores to gimmick-driven scenes. Perhaps, the reason for this difference is fact that novices had not had previous exposure to gimmicks— whereas experts may have had significant prior exposure through their early improv classes—so if creativity is perceived as being a blend of novelty and value relative to a social context, then by default, the perceived novelty alone may have inflated novice scores relative to those of experts.

However, it is also possible that in a better established domain, such as film, the opinions of experts and novices may overlap considerably. Unlike e-textiles or present-day improv practice, film has existed for well over 100 years, and its history is well-documented and ingrained in mainstream daily life (though film is still relatively new in the course of human history). Interestingly, film is so integral to the modern human experience that there are accredited degree programs at prestigious universities, and various outlets in the Web 2.0 era that allow laypersons to evaluate films as professionals do (e.g., YouTube, Rotten Tomatoes, IMDB). These various outlets afford non-professional reviewers with an unprecedented degree of visibility so their opinions can tangibly influence the opinions and consumption behaviors of other viewers. Creativity researchers hypothesized that if film has been ingrained in society's experience, perhaps the inherent familiarity and pervasiveness of film would lead to more similarity between expert and novice opinions. In a study by Plucker and colleagues (2009), they found in film that the clarity between the three levels of expertise: expert, quasi-expert, and novice show overlap and are not as categorically different when using a 10-point rating scale similar to those used by popular media outlets (e.g., IMDB). Although this was not a true application of Amabile's (1982) CAT, it yields valid inferences regarding expert-novice differences because judges were evaluating films in a naturalistic setting and providing ordinal scores based on their own subjective opinions without consulting rubrics or other judges.

In contrast to this example from film, an emergent domain has limited history and pathways for participation, and by default, a limited population of experts (or even those with any experience at all) to rate the creativity of such art productions. However, Kaufman & Baer (2012) argue that the expertise requirement depends on the question being asked, so judgments from non-expert judges can potentially yield valid inferences if their opinions and judgments were consistent with experts. Therefore, quasi-experts, those with some training in the domain, are possible candidates for rating creativity across varying domains specific to their training and experience.

Expertise in CAT

If experts generally arbitrate what is creative in a domain, then it would be necessary for experts to demonstrate some degree of agreement, because if they disagreed, then other persons in the field (e.g., students, consumers) may not know whom to trust (Kaufman & Baer, 2012). However, expertise may not be necessary for a panel of judges to achieve a sufficient level of inter-rater agreement in CAT because novices and quasi-experts might still have judgments that reflect the opinions of experts, and if non-expert judges can still yield valid inferences, then it would make creativity research much easier in a practical sense for the recruitment of judges (Cseh & Jeffries, 2019).

Few studies have clearly shown that non-expert judges can be used as replacements when necessary and those that have attempted to explore the significance of expertise in this way have had mixed results. Cseh and Jeffries (2019) describe how CAT researchers have explored the possibilities of using novices or quasi-experts in addition to or instead of expert judges across a variety of domains. For example, Hickey (2001) compared the consistency of ratings provided by different types of expertise (e.g., music educators, eminent composers) and found that music educators demonstrated higher consistency in scores because they were accustomed to appraising emergent talent, while eminent composers tended to demonstrate more variation in scores because they based their ratings on their own stylistic preferences rather than shared standards. Additionally, Kaufman et al. (2005) found that the scores from gifted non-expert writers were highly correlated with those from experts when appraising creativity of poems and short stories. Although society usually values the opinions of experts more highly than those of non-experts, it may be difficult knowing whose opinions should count for judging creativity when people with considerable expertise demonstrate disagreement.

In summary, Cseh and Jeffries (2019) explain that "expertise in a field only sometimes increases consensus, and it is not clear when or why judge expertise is vital" (p. 9). They elaborate on other important considerations for evaluating consensus among raters, including variation in duration and nature of particular expertise in relation to the types of creative products judged and how tasks are constructed as judges are asked to rate products. Discussion of previous studies using CAT suggests that there is a need to better understand how expertise is defined and how that expertise (or lack thereof) interacts with the domain and creative products in question. For our purposes, we are interested in understanding how notions around what it means to be expert (or less expert) in the domain of e-textiles impacts how creative products are judged to be novel and useful. Building from general findings that factors such as the domain, task structure, and how expertise is defined all impact the extent to which raters across expertise levels agree in their CAT ratings, here we aim to explore how expertise plays out in the emergent field of e-textiles.

The current study employs a variation of the CAT to stimulate a conversation among groups with various expertise about creativity in the emergent domain of e-textiles e-textiles (Buechley et al., 2013). Today a range of construction kits are available that support embedding electronic hardware into textiles (e.g., Adafruit, Gemma, FLORA, myDazzu, etc.), though participants in this study used the LilyPad Arduino, one of the first of its kind (Buechley & Eisenberg, 2008). This study turns to an archive of images created by LilyPad Arduino designers (i.e., participants in workshops as well as the general public) that was hosted on the first-ever e-textile web community, LilyPond (active from 2009 to 2012). The web community allowed designers to share e-textile artifacts and socialize with others in the growing e-textile community (Lowell & Buechley, 2011). The LilyPond web community was prominently used by a

distributed group of Do-It-Yourself (DIY) e-textile designers as well as participants in organized e-textile workshops and afterschool spaces across the United States (Gauntlett, 2011). Early e-textile researchers often used the LilyPond website as a way to document artifacts produced during e-textile workshops and so the majority of artifacts on the site came from these experiences (Lowell & Buechley, 2011). On the LilyPond site, users could share their own work, browse artifacts that were uploaded to the site, and comment on or "love" others' projects. Lowell and Buechley (2011) found that people "used the site in three primary ways: to document projects, find information and inspiration, and make community connections" (p. 365). Participants' artifacts from the LilyPond website were the data sources used to prompt creativity ratings by groups of judges with varying levels of expertise for this study. We describe these workshops and data sources in greater detail in our methods below.

We were guided by the following research questions: (1) To what extent does domain expertise (i.e., by experts, quasi-experts, and novices) influence each group's assessments of e-textile creativity? (2) What is the degree of inter-judge reliability among and between novices, quasi-experts, and experts in this domain? And, (3) How does each expertise group conceive of creativity in this domain and where are points of convergence?

Methods

Contextual Factors Surrounding the Creation of Artifacts

Most e-textile artifacts were created in workshops where materials such as LilyPad LEDs, LilyPad button boards, coin cell batteries and holders, conductive thread, batteries, LEDs, microcontrollers, and various fabrics were provided. Attendees were also encouraged to bring their own garments to customize. Some participants worked with LilyPad Arduino units in the development of their projects. The LilyPad Arduino is a programmable microcontroller that relays information via conductive thread to the other components on the circuit (see Figure 1). The workshops took place in informal learning environments with loose facilitation and limited instruction, and where participation was voluntary; they occurred on weekend mornings or in afterschool programs with moderate-sized groups of about 25 people. Workshops were marketed as educational events for families so there was a mix of children and adults working alongside one another on individual projects. Most participants were new to e-textiles or had limited exposure, though some had participated in previous workshops. Not one participant had extensive experience with e-textiles. In e-textile workshops, participants were often encouraged to search the LilyPond web community for inspiration for their own projects.

Figure 1

LilyPad materials
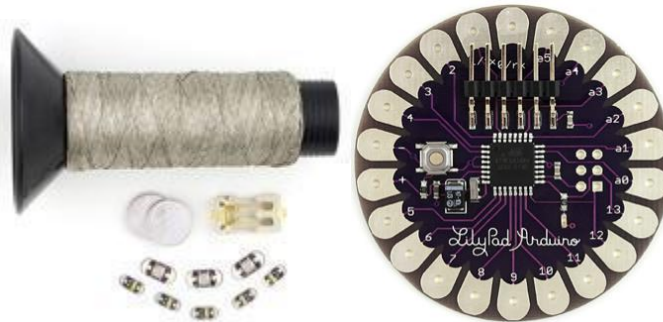
Note: (left) LilyPad basic e-sewing kit: Conductive thread, battery holder, 3V coin cell battery, push button switch, light emitting diode (LED); (right) LilyPad Arduino microcontroller

Although originally the CAT was developed to assess the creativity of artifacts that were made under experimental and tightly constrained conditions (Amabile et al., 1986), some researchers have found that the CAT can also yield valid inferences when assessing the creativity of products created under naturalistic conditions (e.g., Baer et al., 2004), and our workshop format fits into this latter category. Most artifacts were uploaded to the LilyPond web community so others learning about e-textiles could view and interact with created artifacts. The overall goal of the e-textile workshops was for participants to create personally meaningful artifacts by learning how to integrate circuits with fabrics.

Procedures and Data Sources

All research was conducted with institutional review board (IRB) approval. We limited our analysis to e-textile projects displayed within the LilyPond web community (N=166), most of which were created in settings similar to the e-textile workshops described above. However, some projects were created in more formal e-textile courses or by individual e-textile enthusiasts, who voluntarily uploaded their projects from home or other places where they engaged in making e-textiles. Individual judges in expert, quasi-expert, and novice groups independently rated a random selection of about 25% of the artifacts displayed on the site (N=40) following procedures outlined in prior work on the CAT (Amabile, 1982; Amabile, 1996; Baer & McKool, 2009; Kaufman et al., 2008). Additionally, following the precedent of how the CAT has been used across creativity research, raters scored for overlapping constructs, which helped avoid conflating creativity with other constructs. Expert and quasi-expert judges additionally scored for *technical goodness*, and *aesthetic appeal*, though novice judges only scored for *creativity*, rationale for which we explain below in our description of judges. As is traditional in CAT research, by asking expert and quasi-expert judges to score for separate but overlapping constructs, we had an opportunity to explore how those other constructs may be connected to creativity. All artifacts were scored in relation to the three constructs (i.e., creativity, technical goodness, aesthetic appeal) using a 5-point Likert scale (e.g., 1=poor; 3=average; 5=excellent). All judges responded via an online survey and were given links to the specific projects they were asked to rate from their subjective viewpoints on the categories. Instructions at the top of the survey prompted raters to score independently (i.e., not to consult others), and there were no time limits. Artifacts were displayed in random order as judges made their assessments.

Raters viewed photographs of each artifact, and these photographs were supplemented with brief written descriptions about the artist's vision, similar to exhibits at art museums (see Figure 2). As previously noted, it is methodologically important to explain that because we pulled projects from the online community, the projects were non-parallel in nature, meaning the projects were not necessarily created in response to the same prompts or using the same materials (Baer et al., 2004), and projects came from a diverse collection of of designers (varying in ages, skill, geographic region, gender, etc.) who had access to a wide variety of materials and tools (e.g., LEDs, LilyPad Arduinos, unique t-shirts, etc.) either on their own or through workshops like the one described above. Given that not all e-textile artifacts were created under the exact same conditions, the LilyPond environment presented each artifact in a more naturalistic setting, which allowed us to make inferences more valid to creativity in the real world, rather than in a tightly controlled experiment (Kaufman & Baer, 2008). Displayed within LilyPond, each project had at least one associated image, a short descriptive passage about the project and the process of creation, a few associated tags to index the project, and a list of the electronic materials used (see Figure 2).

Figure 2

Sample e-textile project and information displayed within the LilyPond community

·My project altered a t-shirt I have that is a map of the state of Florida with tracks of hurricanes. I used the LED to highlight my hometown when I push the button, stored in my sleeve. I wanted to make a personally meaningful project about my home.

In line with the CAT methods, judges from all three expertise groups (i.e., novice, quasi-expert, expert; described below) rated the e-textile projects independently and did not meet or deliberate their ratings with one another or with the principal investigators until after all the judges' ratings had been submitted. Each expertise group met with researchers after completing their surveys for an informal, video recorded, semi-structured focus group. Focus groups were video recorded and transcribed following the meetings. Each expertise group was asked the same questions, which inspired a conversation around their opinions of general creativity, creativity in e-textiles, the projects online, and the website used as a venue for viewing and judging the e-textile artifacts. We asked:

- How would you define "creativity" in general?
- What does it mean to be creative in e-textiles work?
- What are some of the projects from the survey/LilyPond website that seemed to demonstrate creativity?
- Which projects did you think were most creative and why? Which project was the most creative? Why?
- Which project was the least creative? Why?

These questions are aligned with other qualitative work that aims to dig deeper into judges' subjective viewpoints and how they come to score creative products differently (e.g., Phonethibsavads et al., 2019).

Description of Judges

Three different groups of judges were asked to rate the participants' works on the LilyPond website. The groups varied in expertise levels and included: (a) *expert* artists (N=3) who were pioneers in the domain of e-textiles; (b) *quasi-expert* classroom teachers (N=4) with some previous e-textile experience; and (c) *novice* undergraduate students with no previous e-textile experience (N=19). In order to ensure even comparison across groups, we chose a random selection of the novice undergraduate students (N=4) to run our analyses.

The expert group consisted of artists and computer scientists (N=3) who were chosen due

to their experiences creating groundbreaking materials in the field of e-textiles and extensive contributions to the art, design, and education fields as artists and educators. Expert judges each had more than 10 years of experience working with the materials and are recognized for their expertise in the field of e-textiles. All had taught higher education courses related to e-textiles in art and design. A few additional details on their expertise are worth noting. One of the expert artists helped to define the field of e-textiles through her work at MIT and created some of the very first modern-day e-textile projects. The second expert is also a pioneer in the field of wearable electronics and her work is displayed in the permanent collection at the Museum of Modern Art. The third expert has similar expertise in e-textiles, has displayed her work nationally and internationally, and runs a university e-textiles-focused lab.

The quasi-expert group consisted of professional teachers (N=4) who had rigorous e-textile training via professional development and experience teaching (about 100 hours) using e-textile curriculum and materials. The quasi-expert group received professional development during a summer workshop. Prior to their professional development, none of the teachers had experience using or teaching with e-textiles. Each teacher was chosen due to their experience in education and regarded by their supervisors as exceptional teachers in either an elementary or secondary classroom. We might expect that this group would have the highest internal consistency in rating because they had similar backgrounds as school teachers and their hours spent training with e-textiles may have calibrated them to look for similar criteria across artifacts.

Finally, the novice group consisted of undergraduate pre-service teachers (N=19; N=4 randomly selected for analysis here) who had no previous experience with e-textiles, which we confirmed in an online survey prior to inviting undergraduate judges to be part of the novice group. All novice judges were in a teacher education program and enrolled in the same educational psychology course at the same university. There was no training or discussion regarding the field of e-textiles prior to administering the surveys, and for practical purposes, novice judges were only asked to rate artifacts based on creativity and not other overlapping constructs (e.g., technical goodness, aesthetic appeal). This was for practical reasons since we were asking undergraduates to volunteer to participate in the study and wanted to keep their time commitment short. Though this approach presents a potential methodological limitation since novices did not score for overlapping constructs like experts and quasi-experts did, we reason that it would be less necessary to follow this standard administration of CAT since novices lacked technical knowledge of e-textiles. We also conjecture that inclusion of overlapping constructs may have prompted novices to "overthink" their scores rather than trust their gut. However, we acknowledge that perhaps the inclusion or exclusion of overlapping constructs may have influenced creativity ratings. Whether or not our approach influenced our results is unknown, but overall, for our purposes, asking novices to only rate for creativity was simpler for data collection and more authentic to real-life creativity appraisal.

Data Analysis Methods

The data were analyzed in four steps. First, we calculated the inter-rater agreement of each expertise group (i.e., novice, quasi-expert, expert) and compared internal consistency based on the intraclass correlation coefficient (ICC) because it treats each judge's score as a test-retest for creativity; therefore, a higher ICC would suggest that different judges within that same group are more likely to give similar scores to similarly creative products (Bartko, 1966; Cseh & Jeffries, 2019). Although prior research has utilized Cronbach's alpha to measure internal consistency, we decided to report the ICC instead because the values were identical, and the use

of the ICC as inter-rater agreement is more consistent with the underlying assumptions of inter-judge agreement than Cronbach's alpha (Cseh & Jeffries, 2019). In this case, our goal was primarily consistency of subjective ratings, not absolute agreement. By calculating the ICC, we could compare expertise groups to each other on the basis of their consistency.

Second, we calculated the mean creativity scores and standardized deviations for all three groups of raters. To further determine if there were any significant differences between groups, we used a general linear model repeated measures ANOVA with Bonferroni post-hoc analyses. Third, a correlation matrix for the three sets of ratings was created and examined to highlight the results within and across groups. Lastly, focus group conversations were videotaped and then later transcribed for further analyses. We used a constant comparative method (Glaser, 1965) to analyze focus group transcripts and mine for themes to reveal similarities and differences among groups so we could better understand how they came to understand creativity in this newer domain.

Findings
Quantitative Results: Comparing Ratings Within Groups

We utilized Amabile's (1982) Consensual Assessment Technique (CAT), by prompting each judge to assign ratings of 1 to 5 for creativity, to each of the 40 randomly-selected artifacts, 5 being high. To statistically evaluate the degree of similarities in conceptions of creativity within the three groups, we calculated the Intraclass Correlation Coefficient (ICC). The ICC inter-judge agreement was 0.678[1] for novices, 0.766 for quasi-experts, and 0.829 for experts (see Table 1). The novice group yielded moderate consistency, which is reasonable for a broad construct, and the results for the quasi-expert and expert groups are quite high but also consistent with prior ratings of creative products particularly by experts and quasi-experts, which were found to be similar to the levels found in almost all the research studies that have employed the Consensual Assessment Technique (Baer et al., 2004). The lower ICC suggests that there was a smaller proportion of reliable variance, which may suggest that novices were less internally consistent in giving their scores, and this lower internal consistency may be because novices were only prompted to score for "creativity" while quasi-experts and experts were asked to score additionally for overlapping constructs (e.g., beauty, aesthetic goodness, etc.).

Table 1
Intraclass Correlation Coefficient for Creativity Ratings by Group

Group	**N**	**Mean**	**SD**	**Intraclass Correlation Coefficient (ICC)**
Novices	4	3.27	1.19	0.678
Quasi-Experts	4	3.27	1.18	0.766
Experts	3	2.58	1.20	0.829

Note: E-textile projects rated on a scale of 1 to 5, 5 being highest.
*Acceptable levels of inter-judge reliability using the ICC are generally .70-.90 (Bartko, 1967)

Although, the novice scores may be less consistent than the other groups, they may still be representative of creativity as perceived by general audiences (e.g., laypersons, consumers),

---

[1] Because there were 19 novices, we took 10 random samples of four judges and calculated the ICC for each group and then a mean ICC for all 10 groups (see Appendix A for novice ICC calculations). This mean came out to be .679, and we report one combination that was closest to this mean estimate (.678).

and because the ICC typically increases with the number of judges, it was necessary to analyze the ICC of the novice group with varying numbers of judges. If we treat novice scores as being representative of general public opinion, then it is necessary to examine their scores in larger sample sizes because general audience members are more numerous than experts and quasi-experts (Kaufman & Baer, 2012). Bartko (1966) asserts that the ICC increases rapidly as sample size (i.e., N) approaches 10, then remains steady as it approaches 20, so we compared novice scores at these thresholds and found very strong inter-judge agreement. Despite the moderate ICC (.678) at n=4, we found very strong agreement (.865) with a random sample of n=10 and near-identical agreement (.919) at N=19 (see Table 2). We take this to mean that consistency in ratings may be found within a larger group of novices (n=10+).

Table 2

Intraclass Correlation Coefficient for Novice Creativity Ratings by N

Sample Size	N	Mean	SD	Intraclass Correlation Coefficient (ICC)
20% of Novices	4	3.27	1.19	0.678
50% of Novices	10	3.24	1.23	0.865
100% of Novices	19	3.28	1.14	0.919

Quantitative Results: Comparing Ratings Between Groups

Results from the ICC calculations suggest that there was moderate agreement among novices and good agreement among quasi-experts and experts. We were most interested in understanding whether novices, quasi-experts, and experts had high agreement between all groups, which would help draw conclusions about how to support CAT with domains where experts may be scarce. Therefore, it was necessary to examine the variances between groups because the variances indicate the extent to which judges within a group are giving consistent scores to similarly creative products (see Table 3).

Table 3

ANOVA between expertise groups

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1.072	2	0.536	4.843	0.042*
Within Groups	0.885	8	0.111		
Total	1.957	10			

*significant at .05

The ANOVA suggests that only 4.2% of the variance between groups is overlapping, and because the p-value is lower than .05, we infer that the difference is statistically significant between groups, and the difference co-occurs with rater expertise. However, the ANOVA does not reveal precisely where the differences are, so we ran a Bonferroni post-hoc test to more

closely examine differences between groups (see Table 4).

Table 4
Bonferroni post-hoc analysis of creativity scores between expertise groups

Expertise	Expertise	Mean Difference	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	*Upper Bound*
Novice	**Quasi-Expert**	-0.00376	0.23522	1	-0.7131	0.7056
	Expert	0.69902	0.25406	0.075	-0.0672	1.4652
Quasi-Expert	**Novice**	0.00376	0.23522	1	-0.7056	0.7131
	Expert	0.70278	0.25406	0.073	-0.0634	1.469
Expert	**Novice**	-0.69902	0.25406	0.075	-1.4652	0.0672
	Quasi-Expert	-0.70278	0.25406	0.073	-1.469	0.0634

* The mean difference is significant at the 0.05 level.

The post-hoc test reveals a statistically significant difference between novice and expert judges (p<.05) but no significant difference between novices and quasi-experts, or quasi-experts and experts. The significant difference between novices and experts supports findings in previous literature (e.g., Hickey, 2001; Kaufman et al., 2008), though the lack of significant differences between quasi-experts and the other two groups may certainly be due to a lack of power with CAT.

If we want to argue that quasi-experts can serve as suitable substitutes for experts, we need to examine all possible combinations of experts and quasi-experts in the same panel. To further explore the conclusion that quasi-experts and experts achieved good agreement, we calculated the ICC for all 18 possible combinations of two experts and two quasi-experts. On average, the ICC was .794, which still demonstrates a strong degree of agreement among raters (see Appendix B for all possible combinations of 2+2 Experts and Quasi-Experts). Naturally, the ICC varies depending on the combinations of judges, and because only 3 of the 18 observed cases where the ICC was lower than .7, we are comfortable suggesting that experts and quasi-experts demonstrate good inter-rater agreement for the large majority of cases.

Still, it is important to consider the effect of sample size on the ICC of novices because, in large numbers, they may be more reliable and more accessible than experts, especially in an emerging domain. If we interpret novice scores as being representative of general audiences, then, arguably, the opinions of large numbers of novices may be more valuable than small numbers of experts because new domains must attract new participants in order to sustain their practices, and novice opinions are representative of persons who may eventually join the field or adopt domain practices into their own. Because the ICC becomes increasingly more consistent with larger sample sizes, we propose that larger numbers of novice judges may be equally reliable as small panels of judges. However, it is important to acknowledge that, despite similarly high agreement, the opinions of novice judges are representative of an untrained demographic. Thus, it is necessary to compare qualitative opinions between different expertise groups because there is an apparent measurable difference between the opinions of novices and experts, and

despite both groups being similarly reliable, it is important to unpack the justification behind their scores so that individuals (e.g., artists, students) can better discern whose opinions to trust (Cseh & Jeffries, 2019; Kaufman & Baer, 2012).

Qualitative Analysis of Focus Group Discussions of Creativity Dimensions in E-Textile Artifacts

We present qualitative analysis of focus group discussions about creativity dimensions in e-textiles to highlight how each expertise group conceptualized creativity in their subjective ratings. Across all three groups, we identified sub-scales of creativity described as being: (1) novel and (2) valuable. In brief, we found that all three groups identified products as creative with respect to the "novel and valuable" criteria explored in prior research (e.g., Amabile, 1996; Csikszentmihalyi, 1988; Plucker et al. 2004). Here we unpack how groups specified their criteria for "novel" and "valuable." Points of agreement across these two overarching themes highlight the potential for future CAT research to enlist those with expertise or sufficiently train judges to rate creative e-textile products in reliable and valid ways. We illustrate more fully the depth of explanations of creativity demonstrated by the expert group to show how pioneers in the domain of e-textiles conceive of creativity in its richest sense. Although the smaller sample of the novice group did not show sufficient agreement with the expert group to serve as reliable raters for creativity, we also include analysis from their focus group here because we hypothesize that perhaps we could train novices to a point to put them on par with quasi-experts in future studies or as our findings show, perhaps a larger group of novices could be used for judgments of creative products.

Novel. Perhaps unsurprisingly, for all three groups, the extent to which artifacts were novel or original weighed heavily in perceptions of creativity. Participants explained that creative products were "original," "novel," "rare," "different," or "adventurous." The experts agreed that how creative something is depends on the context and the approach taken by the artist to position the project in relation to existing work. In particular, the distinction was made between new innovations (e.g., the first turn-signal illuminated bike jacket, now an iconic artifact in the field) and their subsequent variations that served similar purposes—the latter being deemed as less creative. Experts had a more complex view of what made artifacts novel or original. In their own words, experts explained:

> This comes up in conversation a lot in my courses with my students. They get upset when they realize that someone else has done their idea before. They thought they were being super unique…so a lot of people have made expandable dresses that deal with personal space issues. But what is *your* dress? And how does it work? So within the context of my teaching I encourage this hybrid approach, be aware of your context but don't let it stop you.

In this way, the artists were able to move beyond their desire to be "the first" and situate work within larger historical trends when working with materials or concepts. Work that is creative may be novel within certain contexts and historical breaks. These experts understand creativity as emerging from prior success and building upon that to redefine what an artifact is in a particular context, so they appeared to value novelty as an emergent property of previously existing and successful artifacts.

The teacher quasi-experts also understood creativity to be about repurposing materials or using materials in ways that "no one else has thought of," but they only seemed to value novelty if there was some utility behind it. They juxtapose this creativity with the idea of being a

"copycat" in the context of the turn-signal bike jacket, explaining that if there were "100 turn signal jackets" that were a "basic copycat" then that would not qualify as creative. However, if the products were "really improved upon" such as a shirt that measures movement, then it could once again be considered creative. Additionally, quasi-experts explained that creativity was about repurposing materials or "reusing something in a different way" or "creating something brand new" or "adding a whole new concept." Although the teacher quasi-experts did not explicitly draw on the value of historical context in their explanations for why particular products were more creative than others, their understanding of how to repurpose or iterate with materials to create novel products was well aligned with the experts' conceptions of creativity.

The novice judges (i.e., undergraduate teacher education students) also aligned with the basic understanding that for something to be creative, it has to include some novel feature. They explained that "reproducing" was not creative behavior but "remixing" is because it "makes something different." In one judge's own words, "So, to me if you are being creative you are being original by creating something rather than like reproducing something that has already been created." Across the three groups, creativity was not about copying but could include repurposing or remixing others ideas' and works. In the case of e-textiles, remixing is a particularly interesting practice that involves the remixing of ideas but also of physical materials, oftentimes recycling older materials to make new e-textile projects. In all, each group agreed that originality or the extent to which artifacts were novel in that they introduced new ideas or built upon ideas or products in existence affected how they rated creativity.

Valuable. When we use the term "valuable," we are not describing monetary value, but rather, cultural, functional, or personal value that the artifacts have for the audience or creator. In addition to the novelty criterion, all three expertise groups commented that, to an extent, that utility and/or personal meaningfulness factored into their ratings. Experts said that it was important to understand the function and intention of a project, which was often conveyed through written text or multiple images on the LilyPond website, because it influenced their decision to rank some projects highly and, conversely, the lack of such understanding influenced their decision to rank a project lower. This was especially true for projects that were unfinished work or whose functions were difficult to discern from the pictures alone. In these instances, the text conveyed some of the work and decision-making that artists put into their creations. This trend was reiterated in later discussions of which projects were least creative, in which experts explained that supportive text and clear images played a role in their rating of the projects. Experts also focused on how creativity depends on the context and whether or not the artist went beyond the scope of what the pieces are typically used for in comments like the following: "thinking about not just a switch as a switch but as interaction." Novice judges referenced the purpose more generally, as one judge explained, "If I could see a definite purpose, then I scored it for creative."

For all groups, the extent to which a project was valuable was also often connected to functionality or personal meaning. For instance, experts picked out one project where the creator had put a single LED on a t-shirt with a picture of the state of Florida, positioning the LED at the location of his hometown. Because the creator explained the reason for putting a light at his hometown on the map, judges tended to rank this as more creative than other projects of comparable technological simplicity. A designer's choice of materials also factored into both panels' assessment of the degree to which a project was "inspired." Experts also spoke about creativity as "a personal thing," and elaborated about how individual artists work to develop particular creative themes in their work, such as themes of subverting the use of the materials

(i.e., having an LED on the inside of clothing as opposed to the outside) as a form of creative expression.

      Additional Connections. There were additional connections across the focus groups that may support our emergent understanding of how experts, quasi-experts, and novices aligned on their conceptions of creativity within the domain of e-textiles. Though the focus of CAT is on judging creative products, all three groups recognized creativity as part of a process. Novice judges mentioned that having sketches (e.g., drafts) or circumstantial knowledge about the process (e.g., why the artifact was unfinished) would help them better determine what was creative. Novices also said that videos of artists making items would make it easier to rate the creativity of artifacts. All three groups noted that the text on the LilyPond site explaining the projects helped them make decisions about creativity ratings. Creativity was furthermore seen as an inspired act, and some judges commented that the LilyPond site helped artists brainstorm and get more ideas for their e-textile projects.

      Additionally, the expert and quasi-expert groups both said that age affected the way they rated their projects. For example, if an artist was younger, judges tended to rate creativity higher and judge more leniently because they would not expect a young person to do such complex work, a finding that aligns with findings on using the CAT in the context of drama performance (Phonethibsavads et al., 2019). Judges across groups also aligned with general conceptions of what made artifacts less creative. The least creative products were unfinished work or were difficult to discern the function from the picture alone. All judges rated these items as less creative because they were considered "sloppy," "plain," "too simple," "copycat" work, or work that "did not make sense." Finally, both the novice and quasi-expert groups commented that it is difficult to judge the creativity of something that they did not understand or know much about since the field was new to them. They admitted that they did not know what they were judging at times, but teacher quasi-experts explained that as they gained experience and understanding of the field as a teacher, their confidence grew.

Discussion

      Our findings have implications for future studies of creativity using the CAT with emerging domains, such as e-textiles. First, we found that both experts and quasi-experts were reliable judges for assessing creativity and therefore, concluded that both experts and quasi-experts could serve as suitable judges in future studies assessing creativity in e-textiles. Additionally, because emerging domains may still be forming their disciplinary identities, they would depend on attracting outsiders to grow their communities and institutionalize conventions and practices, so it may be appropriate to utilize larger panels of novice judges because, arguably, general consumers (i.e., novices) are also gatekeepers in the emerging domain landscape (Csikszentmihalyi, 1988). Although larger panels of novices may be equally reliable as small panels of experts, it is important to note that they are different kinds of gatekeepers because novices represent the opinions of general audiences (e.g., consumers) while experts and quasi-experts represent the opinions of authority figures (e.g., teachers, critics). Perhaps, this democratization of who decides what is creative may lay the groundwork for future studies about the ways that students and artists receive feedback, and if we see how students and artists listen and respond to novice versus expert feedback, it may inform how novice and expert gatekeepers influence the development of the field.

      Although novices were not found to be reliable judges at the n=4 level, we found reliability within the novice group at n=10 and also considerable overlap in how all three

expertise groups conceptualized creativity through qualitative analysis of focus group interviews. These findings suggest that perhaps even novice judges could be adequately trained or a sufficient number could be recruited to be reliable judges for e-textile creativity studies. Future studies could further explore the level of experience and understanding of e-textiles that would be necessary to create a group of reliable raters without needing to recruit judges with considerable expertise.

Limitations

There are several limitations of the present study worth noting and expanding upon in future work. To begin, as we reflected on earlier, novices only rated for creativity and not other constructs. Although we explain this choice above as due to novices lacking domain-specific knowledge of e-textiles, novices' creativity ratings may have actually been conflated with other constructs (e.g., beauty, aesthetic qualities); however, we argue that their scores are still representative of novice judgments in the field because judges gave scores on the basis of their own subjective opinions and preferences, not rubrics or operational definitions. As another limitation, the small number of judges for each group (three experts, four quasi-experts, and three novices) was constrained by the number of true experts we were able to identify and recruit. In general, as the number of judges approaches 10, the ICC increases by effect of sample size (Bartko, 1966). However, since we found a good ICC with small numbers, then we expect to find a similarly high ICC with larger numbers of judges.

As an additional limitation, perhaps there were other factors in play that made these groups rate creativity the way they did outside of their expertise grouping. For example, perhaps because all quasi-experts were teachers and all novices were undergraduate pre-service teachers, their creativity ratings could have been more so a reflection of their judgments from a teacher-perspective, not necessarily domain-expertise; however, we argue that this perspective may have allowed for fairer comparisons between judges because the teacher-perspective was likely to be consistent across novice and quasi-expert groups. In future studies we could increase the number of judges, but only slightly, so as to not calculate an inflated ICC due to a high N-value. Additionally, we could recruit novice and quasi-expert judges who are not teachers or pre-service teachers.

As a further limitation, the field of e-textiles is still in the early stages of its evolution as an emergent domain, so it may be difficult to know if these results will remain relevant over the course of different generations. To be clear, macro-level trends and conventions in domains typically affect the ways that individuals contribute to their fields because each domain has gatekeepers who determine what products are novel and valuable (Csikszentmihalyi, 1988). Additionally, at the individual-level, Guilford (1967) characterizes creative thinking as cycles between divergent thinking (e.g., expanding, increasing variety) and convergent thinking (e.g., eliminating irrelevant options, making unrelated pieces fit together), and, contrary to the societal myth of the lone genius, much real-world creativity occurs in social environments (Amabile, 1996; Plucker et al. 2004). Because we saw creativity as largely social, we can extend Guilford's divergent-thinking and convergent-thinking models to the domain-level, and, perhaps, argue that each domain also goes through cycles of divergence and convergence because groups of individuals are simultaneously co-creating a new artistic domain while differentiating their own styles and voices (Simonton, 2010). Because all expert judges were pioneers in the field, we believe that they had high agreement because the domain was still finding its identity, meaning that it was likely in a convergent stage. Perhaps, the experts had similar judgments because they

were working to synthesize a new domain that took elements from engineering, fashion, and art, and this emergent domain needed to establish consistent standards. Conversely, if a domain has been around for a few generations, then different artists may begin to realize that they are independently deriving the same ideas (e.g., 100 turn-signal jackets), and they might work to differentiate themselves and their products from their peers, thus prompting a pattern of divergence.

New Directions

Furthermore, as discussed previously, as social media changes the landscape of how ideas stretch and redefine a domain, what constitutes creativity is also in flux. For instance, Peppler & Solomou (2011) elaborate on crowdsourcing as an approach that distributes expertise among members of a social media site, thus "fundamentally changing the nature of how we view and assess 'creativity,' calling into question of who constitutes the 'field,' and expanding the methodologies that we can use to investigate creativity" (p. 14).

Creating social environments that promote creativity is needed to cultivate practices that youth initiate in these new domains, and further exploration of the impact creativity has on youth learning in new domains is needed to continue this conversation. Particularly within the educational and social sciences, recognizing creativity within disciplines is a central area of study. An emphasis on creativity causes us to value creative processes within disciplines rather than simply demonstrating mastery. Likewise, future research on creativity within social sciences domains can expand both research and practice in the respective fields. It is new territory for creativity research that in the domain of e-textiles, rooted in digital media and learning and is heavily represented in online communities through the affordance of Web 2.0 capabilities, would have such agreements between the various participants involved. Specifically, future research is needed to better understand why there is consensus between quasi-experts and experts within the domain of e-textiles as well as considerable overlap between experts, quasi-experts, and novices in their conceptualizations of creativity.

Acknowledgements

References

Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology*, *43*(5), 997–1013. https://doi.org/10.1037/0022-3514.43.5.997

Amabile, T. M., Hennessey, B. A., & Grossman, B. S. (1986). Social influences on creativity: The effects of contracted-for reward. *Journal of Personality and Social Psychology*, *50*(1), 14. https://doi.org/10.1037/0022-3514.50.1.14

Amabile, T.M. (1996). *Creativity in context: Update to "The social psychology of creativity."* Westview Press.

Baer, J., Kaufman, J.C., & Gentile, C. (2004). Extension of the consensual assessment technique to nonparallel creative products. *Creativity Research Journal*, *16* (1), 113-117. https://doi.org/10.1207/s15326934crj1601_11

Baer, J., Kaufman, J.C., & Riggs, M. (2009). Brief report : Rater-domain interactions in the

Consensual Assessment Technique. *The International Journal of Creativity & Problem Solving, 19*(2), 87-92.

Baer, J., & McKool, S. (2009). Assessing creativity using the Consensual Assessment Technique. In C. Schreiner, Ed., *Handbook of assessment technologies, methods, and applications in higher education* (pp. 65-77). IGI Global. https://doi.org/10.4018/978-1-60566-667-9.ch004

Barbot, B., & Reiter-Palmon, R. (2019). Creativity assessment: Pitfalls, solutions, and standards. *Psychology of Aesthetics, Creativity, and the Arts*, *13*(2), 131–132. https://doi.org/10.1037/aca0000251

Bartko, J. J. (1966). The intraclass correlation coefficient as a measure of reliability. *Psychological Report*s, *19*(1), 3-11. https://doi.org/10.2466/pr0.1966.19.1.3

Buechley, L. (2006). *A construction kit for electronic textiles*. In Proceedings of IEEE International Symposium on Wearable Computers (ISWC) (pp. 83-92). Montreux, Switzerland. https://doi.org/10.1109/ISWC.2006.286348

Buechley, L., & Eisenberg, M. (2008). The LilyPad Arduino: Toward wearable engineering for everyone. *IEEE Pervasive*, *7*(2), 12-15. https://doi.org/10.1109/MPRV.2008.38

Buechley, L., Eisenberg, M., Catchen, J., & Crockett, A. (2008). The LilyPad Arduino: Using computational textiles to investigate engagement, aesthetics, and diversity in computer science education. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 423-432). ACM. https://doi.org/10.1145/1357054.1357123

Carson, S. (2006, April 19). Creativity and mental illness. [Invitational panel discussion] Yale Mind Matters Consortium, New Haven, CT, United States.

Cseh, G. M., & Jeffries, K. K. (2019). A scattered CAT: A critical evaluation of the consensual assessment technique for creativity research. *Psychology of Aesthetics, Creativity, and the Arts*, *13*(2), 159–166. https://doi.org/10.1037/aca0000220

Csikszentmihalyi, M. (1988). Society, culture, and person: A systems view of creativity. In *The systems model of creativity* (pp. 47-61). Springer, Dordrecht. https://doi.org/10.1007/978-94-017-9085-7_4

Csikszentmihalyi, M. (1996), C*reativity: Flow and the psychology of discovery and invention*. HarperCollins.
Etechtex (2017, May). History of e-textile. E-textile: Technology of textile. https://etechtex.wordpress.com/2017/05/28/history-of-e-textile/

Csikszentmihalyi, M. (2014). Creativity and genius: A systems perspective. In M. Csikszentmihalyi (Ed.), *The systems model of creativity: The collected works of Mihaly Csikszentmihalyi* (pp. 99-125). Springer Netherlands. https://doi.org/10.1007/978-94-017-9085-7_10

Gauntlett, D. (2011). *Making is connecting*. Polity Press.

Glăveanu, V. P. (2010). Paradigms in the study of creativity: Introducing the perspective of cultural psychology. *New Ideas in Psychology*, *28*(1), 79-93. https://doi.org/10.1016/j.newideapsych.2009.07.007

Glaser, B. G. (1965). The constant comparative method of qualitative analysis. *Social Problems*, *12*(4), 436–445. https://doi.org/10.1525/sp.1965.12.4.03a00070
Guilford, J.P. (1967). *The nature of human intelligence.* McGraw-Hill.

Hennessey, B.A. (1994). The Consensual Assessment Technique: An examination of the relationship between ratings of process and product creativity. *Creativity Research Journal,* 7, 193-208. https://doi.org/10.1080/10400419409534524

Hickey, M. (2001). An application of Amabile's consensual assessment technique for rating the creativity of children's musical compositions. *Journal of Research in Music Education*, *49*(3), 234-244. https://doi.org/10.2307/3345709

Jenkins, H. (2007). Afterword: The future of fandom. In J. Gray , C. Sandvoss , and C. L. Harrington (Eds.), *Fandom: Identities and communities in a mediated world*, (pp. 357-64). NYU Press.

Kafai, Y., & Peppler, K. (2011). Beyond small groups: New opportunities for research in computer-supported collective learning. In H. Spada, G. Stahl, N. Miyake, & N. Law, (Eds.) *Proceedings of CSCL 2020: Connecting research to policy and practice* (pp. 17-24). International Society of the Learning Sciences.

Kaufman, J. C., & Baer, J. (2012). Beyond new and appropriate: Who decides what is creative? *Creativity Research Journal*, *24*(1), 83–91. https://doi.org/10.1080/10400419.2012.649237

Kaufman, J.C., Baer, J., & Cole, J.C. (2009). Expertise, domains, and the Consensual Assessment Technique. *Journal of Creative Behavior. 43*(4), 223-233. https://doi.org/10.1002/j.2162-6057.2009.tb01316.x

Kaufman, J. C., Baer, J., Cole, J. C., & Sexton, J. D. (2008). A comparison of expert and nonexpert raters using the Consensual Assessment Technique. *Creativity Research Journal*, *20*(2), 171-178. https://doi.org/10.1080/10400410802059929

Kaufman, J. C., Baer, J., & Gentile, C. A. (2004). Differences in gender and ethnicity as measured by ratings of three writing tasks. *Journal of Creative Behavior*, *38*(1), 56-69. https://doi.org/10.1002/j.2162-6057.2004.tb01231.

Kaufman, J. C., Gentile, C. A., & Baer, J. (2005). Do gifted student writers and creative writing experts rate creativity the same way? *Gifted Child Quarterly*, *49*(3), 260-265. https://doi.org/10.1177/001698620504900307

Kaufman, J.C., Plucker, J., & Baer, J. (2008). *Essentials of creativity assessment,* John Wiley & Sons.

Keune, A., & Peppler, K. (2019). Child-material computing: Material collaboration in fiber crafts. In K. Lund, G. P. Niccolai, E. Lavoué, C. Hmelo-Silver, G. Gweon, & M. Baker, (Eds.), A wide lens: Combining embodied, enactive, extended, and embedded learning in collaborative settings, 13th International Conference on Computer Supported Collaborative Learning (CSCL) 2019, Volume 2 (pp. 913-914). International Society of the Learning Sciences

King's College London. (2017). Statistics advisory service FAQ - Questionnaire development. Retrieved from https://www.kcl.ac.uk/ioppn/depts/BiostatisticsHealthInformatics/SAS/faqs9.aspx#a9_3

Lowell, E., & Buechley, L. (2011). LilyPond: An online community for sharing e-textile projects. In *Proceedings of the ACM Cognition & Creativity Conference*, (pp. 365-366). ACM. https://doi.org/10.1145/2069618.2069698

Mayer, R. E. (1999). Fifty years of creativity research. In R. J. Sternberg (Ed.), *Handbook of human creativity* (pp. 449–460). Cambridge University Press. https://doi.org/10.1017/CBO9780511807916.024

Myszkowski, N., & Storme, M. (2019). Judge response theory? A call to upgrade our psychometrical account of creativity judgments. *Psychology of Aesthetics, Creativity, and the Arts*, *13*(2), 167–175. https://doi.org/10.1037/aca0000225

Peppler, K., & Glosson, D. (2013a). Learning about circuitry with e-textiles in after-school

settings. In L. Buechley, K. Peppler, M. Eisenberg, & Y. Kafai (Eds.), *Textile messages: dispatches from the world of E-textiles and education*. Peter Lang Publishing.

Peppler, K., & Glosson, D. (2013b). Stitching circuits: Learning about circuitry through e-textile materials. *Journal of Science Education and Technology*, *22*(5), 751-763.

Peppler, K. & Solomou, M. (2011). Building creativity: Collaborative learning and creativity in social media environments. *On the Horizon*, *19*(1), 13-23. https://doi.org/10.1108/10748121111107672

Phonethibsavads, A. (2020). *Learning and creativity through consensus-building and rules-deviation in improvisational theater* [Doctoral dissertation, Indiana University].

Phonethibsavads, A., Peppler, K. A., & Bender, S. M. (2019). Utilizing the consensual assessment technique to compare creativity in drama spaces, *Creativity – Theories, Research, Applications*. *6*(1)*, 4-19.

Plucker, J. A., Beghetto, R. A., & Dow, G. T. (2004). Why isn't creativity more important to educational psychologists? Potentials, pitfalls, and future directions in creativity research. *Educational Psychologist*, *39*(2), 83-96. https://doi.org/10.1207/s15326985ep3902_1

Plucker, J. A., Kaufman, J. C., Temple, J. S., & Qian, M. (2009). Do experts and novices evaluate movies the same way?. *Psychology & Marketing*, *26*(5), 470-478. https://doi.org/10.1002/mar.20283

Plucker, J. A., & Makel, M. C. (2010). Assessment of creativity. In J. C. Kaufman & R. J. Sternberg (Eds.) *The Cambridge handbook of creativity*, (pp 48-73). Cambridge University Press. https://doi.org/10.1017/CBO9780511763205.005

Sawyer, K. (Ed.) (2006a). *Handbook of the Learning Sciences*. Cambridge University Press. https://doi.org/10.1017/CBO9780511816833

Sawyer, K. (2006b). *Explaining creativity*. Oxford University Press.

Shirky, R. (2008). *Here comes everybody*: *The power of organizing without organizations.* Penguin Press.

Simonton, D. K. (1994). *Greatness: Who makes history and why*. Guilford Press.

Simonton, D. K. (2010). Creativity in highly eminent individuals. In J. C. Kaufman & R. J. Sternberg (Eds.) *The Cambridge handbook of creativity*, (pp 174-188). Cambridge University Press. https://doi.org/10.1017/CBO9780511763205.012

Starner, T. (2002). Wearable computers: No longer science fiction. *IEEE Pervasive Computing / IEEE Computer Society [and] IEEE Communications Society*, *1*(1), 86–88. doi:10.1109/MPRV.2002.993148

Tan, V., & Peppler, K. (2015, June). Creative design process in making electronic textiles. In *Proceedings of the 14th International Conference on Interaction Design and Children* (pp. 327-330). ACM.

Torrance, E. P. (1972). Can we teach children to think creatively?. *The Journal of Creative Behavior*, *6*(2), 114-143. https://doi.org/10.1002/j.2162-6057.1972.tb00923.x

Wallas, G. (1926). *The art of thought*. J. Cape: London.

Appendix A

ICC for 10 Unique Combinations of Novice Judges

Combination	1st Novice	2nd Novice	3rd Novice	4th Novice	ICC
1	4	10	7	11	0.573
2	18	15	2	12	0.766
3	14	19	17	7	0.677
4	7	4	5	14	0.706
5	8	10	14	7	0.703
6	19	5	3	6	0.701
7	1	10	2	14	0.605
8	2	17	4	16	0.576
9	7	9	8	3	0.813
10	13	2	9	15	0.678

Appendix B

ICC for All Possible Combinations of Two Experts and Two Quasi-Experts

Combination	Experts		Quasi -Experts		ICC
1	1	2	1	2	0.719
2	1	2	1	3	0.697
3	1	2	1	4	0.738
4	1	2	2	3	0.843
5	1	2	2	4	0.843
6	1	2	3	4	0.858
7	1	3	1	2	0.774
8	1	3	1	3	0.746
9	1	3	1	4	0.796
10	1	3	2	3	0.869
11	1	3	2	4	0.884
12	1	3	3	4	0.894
13	2	3	1	2	0.691
14	2	3	1	3	0.692
15	2	3	1	4	0.727
16	2	3	2	3	0.828
17	2	3	2	4	0.830
18	2	3	3	4	0.858