

Generative Inquiry: Creativity Assessments of Youth Epistemic Agency and in a Computational Data Science Workshop

Justice T. Walker, University of Maryland, College Park, justicew@alumni.upenn.edu

Seyedahmad Rahimi, University of Florida, srahimi@ufl.edu

Amanda Barany, University of Pennsylvania, amanda.barany@gmail.com

Lin Lin-Lipsmeyer, Southern Methodist University, llipsmeyer@smu.edu

Abstract: Computational data science (CDS) is relevant in areas of artificial intelligence (AI) and machine learning (ML) for reasons related to field innovation and public literacy. Much is known about pre-college learning with guided CDS in traditional subjects, and comparatively less about how to design and identify generative moments in open inquiry, which are inherent and valued in the field. We examine two illustrative cases of high school-aged youth's final projects to assess strategies used when enacting a self-directed inquiry project. We report how creativity indicators show up in activities emphasizing learner agency. Our findings represent an initial effort to use creativity to assess process outcomes in the context of open inquiry. We recommend creativity-based assessments of inquiry-driven by generative acts to promote knowledge construction and learning. Insights are discussed in terms of how they extend ideas about assessing constructs such as epistemic agency in open-ended learning.

Introduction

Computational data science (CDS) forms the bedrock for advancement in emerging computing areas such as artificial intelligence (AI) and machine learning (ML), where innovation can be field-defining (Education & Human Resources, 2020). This recognition motivates a growing body of relevant literature about how CDS might support technical and public literacies (Tang & Williams, 2019). Most existing efforts focus on using CDS to support traditional academic goals, such as achieving learning objectives in mathematics, life sciences, social sciences, and humanities. In contrast, fewer studies adopt more generative approaches that iteratively examine and emphasize open inquiry activities to foster conceptual understanding, practical skills, sociopolitical awareness, and innovation in these rapidly evolving fields (Jiang, Lee, & Rosenberg, 2022). This is relevant in middle and high school education (grades 6-12) because youth in these age groups use learning experiences to inform their future career interests and pursuits. Thus, activities that offer opportunities to participate authentically and generatively represent a valuable path toward realizing priorities for the field and the public.

Addressing these issues is elusive in pre-college learning environments for two reasons. First, authentic open inquiry in CDS requires an open library or repository of materially flexible data to develop and pursue novel research questions and pursuits. Second, few existing educational frameworks account for learning processes and outcomes in CDS contexts where open inquiry practices vary by learners and their diverse epistemological styles (Turkle & Papert, 1990). Social media platforms such as X/Twitter and YouTube, with their publicly accessible Application Programming Interfaces (APIs), have long been used by industry and researchers to carry out sophisticated inquiries (Bazzaz et al., 2021). Efforts to leverage these access points for pre-college learning have shown early promise (Barany et al., 2023; Walker et al., 2023) in part because social media is a relevant space for youth (Acquah et al., 2024; Barany et al., 2024; Walker, Slater & Kafai, 2019) who tend to use such platforms regularly. Prior works use a sandbox metaphor to describe CDS as an open inquiry, where learners have agency to define research questions and make decisions about data and strategies for collection, analysis, and representation—all of which position learners as epistemic agents, cognitive authorities, and constructors of knowledge. These developments create opportunities to begin assessing agency's generative nature in the CDS context. Starting here may inform the development of education frameworks to support epistemic agency (in terms of process and products) in open inquiry learning designs and practices.

This initial study examines projects from seven high school-aged youth who participated in a ten-day online CDS learning workshop using a sandbox approach. The curriculum was designed to encourage learner agency (by giving youth access to social media-based data repositories) and ownership in both the process and final products achieved (by giving youth authority over epistemic decision-making). Using deductive qualitative analysis of video recording transcripts of final project presentations and presentation slides (to triangulate and confirm), we address the following research question: How did learner project narratives and presentations reflect enacted creativity in their epistemic practice? Transcripts were a vital data source because they represent first-hand narratives of participant rationales for strategies taken to carry out an inquiry. We reasoned that while many choices participants made could be construed as agentic, only a few had generative and original qualities that align

with creativity, perhaps more consistent with what might support learner interest and sustained engagement. Our observations in two illustrative cases presented here suggest that the extent to which learners could enact agency that is generative could be understood and operationalized in terms of creativity. These two cases provide initial insights sufficient to illustrate and describe how creativity shows up in the context of CDS. We discuss these findings in relation to the literature and our working hypothesis that learning in open inquiry environments that encourage agency can be understood in terms of how much creativity a learner can enact.

Background

Computing education, epistemic agency, and creativity

Computing education has made significant inroads in supporting authentic and expressive pre-college learning activities, wherein learners use computational tools and resources to inquire and construct knowledge (Odden, Silva & Malthe-Sorensen, 2021). This can be attributed, in part, to the development of data mining approaches and related technologies that reduce the entry threshold needed for meaningful participation that is both flexible and sophisticated (Resnick & Silverman, 2005). Data mining is the process of discovering patterns, trends, and useful information from large datasets using statistical analysis, machine learning, and computational techniques to analyze data and extract insights that may not be immediately apparent (Romero & Ventura, 2013). As the field expands, learning to mine data can empower pre-college learners to engage with the skills and concepts driving innovation (King & Satyanarayana, 2013), including in AI and ML, where these skills are essential. Data mining provides a concrete avenue for exploring open inquiry, where learners can autonomously identify patterns, generate insights, and make decisions using real-world data. This aligns with the growing importance of equipping youth with computational thinking and data literacy for academic and societal needs. Tools used in data mining include a diversity of programming languages (e.g., from block-based to text-based), digital tools (e.g., AppInventor, NetLogo, etc.), and physical computing electronics (e.g., Microbits, LilyPad Arduinos). These have not only made computational inquiry more accessible but also made it possible for learners to pursue personally relevant and complex inquiries. While this progress has provided foundational examples for designing and supporting learning experiences suited for the 21st century, we know comparatively less about data science-adjacent fields like CDS, where open inquiry is closely tied to epistemic practice. We also know little about identifying when learners are novelly enacting agency in learning, and there are currently no unifying frameworks in computing for operationalizing these moments.

Epistemic agency in education can be understood as the myriad ways learners form and enact knowledge and practices in a given community (Scardamalia & Bereiter, 1991). This includes the ability to define or direct epistemological goals and make decisions about the knowledge of processes taken up to carry out or engage with those priorities (Damşa et al., 2014; Rosenberg & Jones, 2022). From an equity perspective, epistemic agency broadens pluralistic conceptualizations of authentic participation in knowledge construction (see Turkle & Papert, 1990) to include learners as legitimate actors in the process of mobilizing resources and configuring systems that contribute to knowledge production (Miller et al., 2018; Varelas, Tucker-Raymond & Richards, 2015). A growing body of literature examines how epistemic agency is negotiated between learners and teachers (Ko and Krist, 2019) about education priorities (Miller et al., 2018). This represents a shift away from science education instantiations that privilege science as an accumulation of knowledge, theories, or logic and toward conceptualizations emphasizing knowledge as a set of generative practices shaped by social and cultural contexts (Stroupe, Moon & Michaels, 2019). In academic disciplines (Odden et al., 2021), learning is situated (Brown, Collins & Duguid, 1989), and activities support enculturation into practice communities, with their value determined by their alignment with authentic science inquiries. Odden and colleagues (2021) helped to operationalize and assess epistemic agency in computational inquiry by examining how college student ownership manifested at different inquiry stages. Despite this progress, further research is needed to understand how epistemic agency can be generative during investigations. In the next section, we describe the literature on creativity, how this theoretical lens provides relevant vantage points for understanding epistemic agency, and how it can be enacted novelly in computing education and adjacent fields.

Creativity is the capacity to generate novel, valuable, and appropriate products (e.g., ideas, solutions, patterns) beyond conventionality (Sternberg & Kaufman, 2010). Creativity researchers have used proxies such as divergent thinking to operationalize creative thinking. Guilford (1956) operationalized creativity as divergent thinking with four sub-facets: *fluency* (the ability to produce many relevant ideas), *flexibility* (the ability to come up with relevant ideas from different categories or themes), *originality* (the ability to produce statistically rare ideas), and *elaboration* (the ability to implement an idea in detail and high quality). Such multi-faceted models and theories can help researchers develop accurate creativity *process* assessments and target creativity sub-facets to enhance creativity (e.g., Rahimi, 2020; Rahimi et al., 2023; Rahimi & Shute, 2021). In this study, we also

leverage the creativity assessment of a *product* that emerged from Rahimi et al., 2023 in their evaluation of a computationally rich learning environment to include *surprise*, *complexity*, and *effort* as three additional elements of creativity. Moreover, Rhodes (1961) introduced a four-P model of creativity, suggesting that creativity can be understood relative to the (1) person (e.g., personality, intellect, habits, and traits), (2) process (e.g., perception, learning, thinking, and communication), (3) press (i.e., the environment that one lives in or interacts with), and (4) product (e.g., a communicated thought to others in words, paint, clay, metal, stone, fabric, or other materials). In this study, we focus on the process and the product as represented in first-hand participant accounts and their final projects.

Scholars have also considered the potential relationship between agency and creativity—including the possibility that agency might facilitate or even encourage learner creativity or vice versa (Karwowski, Zielińska, & Jankowska, 2022). This can be understood through Self-Determination Theory (SDT; Ryan & Deci, 2002), which accounts for how much autonomy can lead to enhanced internal motivation (i.e., doing something for the sake of the task rather than for external reasons), creativity, and engagement to facilitate *flow* (Csikszentmihalyi, 1997). During flow, learners may lose sense of time, resulting in an enjoyable experience that can enhance learning. In this way, creativity and epistemic agency may positively influence each other.

In computing education, creativity is often used to examine how learning designs might empower learners to personalize their products rather than stances that examine how creativity might be a generative means of enacting agency in the learning process (Buechley et al., 2008). We take up creativity to also represent an epistemic characteristic (or virtue as described by Kieran, 2018) that can define the extent to which scientific practice can produce knowledge in ways that are generative in meaning, utility, or context (Sanchez-Dorado, 2020). In other words, creativity is a lens that helps understand how epistemic computing processes might be functional and generative. This vantage point allows us to discern the potential for creativity might drive agency in a particular learning environment design (e.g., open inquiry or sandbox learning environments). This conceptualization is the starting point in our effort to assess what it looks like to be agentic in CDS and how learning designs might support enactments considered generative. Thus, our examination of two vignettes seeks to illustrate the potential relation between learner agency in a CDS activity and how learners can be creative in learning and while enacting inquiry (i.e., the ways epistemic agency might be understood as a creative endeavor).

Methods

Context and participants

This study took place during a 2-week (Monday - Friday), 4-hour-a-day, online workshop using Zoom and Google Colaboratory. The workshop used a modified 17-module “Coding Like a Data Miner” pilot curriculum (Barany et al., 2023; Walker et al., 2023a; 2023b). Participants learned about data science, statistics, and computational data mining during the first week. Inquiries were scaffolded, e.g., they could modify existing inquiry examples and codes provided by instructors. During the second week, the inquiry was fully open. This open inquiry framing encouraged participants to take ownership and agency in their projects. This agency was epistemic because it was couched in computational data mining phases that facilitate inquiry. To set this up, participants were given instructions to identify their research questions, devise their own data collection and analysis strategies, and generate visualizations using productive means. On the last day, the participants were asked to deliver a 10–20-minute final presentation about their project, including details about the processes and strategies they chose to carry out their work. Participants were instructed to introduce and justify the topic they were exploring and then detail their processes for data collection, processing, statistical analysis, and visualizations of findings. Participants were then instructed to, when possible, annotate their methods and strategies in the slide portion of their presentation. This included annotations of code and process drafts. Eight youths, ages 11–17, registered and subsequently participated in the publicly available workshop. Participants identified as three Asian descent, three African American/Black, one Latinx, and one Caucasian/White. Four participants self-identified as male, and four as female. Participants had varying experiences with programming and data science, from very experienced to no experience in coding or carrying out statistics. The data included in our analysis comprised seven participants because one did not participate in the final presentation due to technical issues.

Data sources and analysis

Data sources included transcripts from video recordings of participants' final presentation narratives and slides (Google or PowerPoint) of their final projects (used to triangulate with narratives). Transcripts of youth presentations gave insights into aspects of their self-described creative processes, while slides were assessed as youths' summative creative products to confirm and validate what the youth explained. While student preparation and work time recordings were available on Zoom, these moments primarily consisted of quiet, off-screen focused

work and were not included in the analysis. The video recordings ranged from 9:13 to 19:18 minutes. Video recordings were transcribed verbatim using Otter.ai and were edited as needed when transcription errors arose. Pseudonyms were assigned to all participants in the workshop. Transcriptions and final presentations were organized in a table for qualitative coding—analysis of transcripts involved primarily deductive approaches (Ravitch & Carl, 2019) wherein authors applied creativity codes derived from published creativity assessment in Rahimi et al., 2023. Specifically, author one coded transcripts using creativity facets: complexity, effort, elaboration, flexibility, fluency, originality, and surprise. Author one also noted the features of transcript text that warranted each code application (e.g., aspects of participant explanations that could be construed as complex in the context of CDS). Author one applied codes independently and then met to discuss observations with authors two, three, and four. When interpretive consensus could not immediately be established, all authors deliberated until a consensus (i.e., 100% interrater reliability) was reached (Ravitch & Carl, 2019). Two participants were purposefully sampled and presented in the findings to, consistent with the goal of this study, illustrate initially the quality and scope of how we observed creativity in the context of open inquiry-based CDS.

Findings

Acts of creativity signal epistemic agency in computational data science

Our analysis suggests that acts of creativity served as reasonable indicators of epistemic agency in CDS because they signal instances of when and how learners were generative rather than replicate a strategy developed by others, which could also be construed as agentic (but not generative). This included indicators such as when participants stated they deviated from guided examples provided in the first week and how these deviations could be characterized by features of the creativity construct we operationalized. We observed seven creativity facets across projects. The facets observed included creativity features like complexity (the various ways they interpreted data or offered nuanced interpretations), effort (instances when participants stated spending considerable time on an approach), elaboration (when a participant's stated data processing strategies required consideration of data structure), and flexibility (when participants identified multiple strategies for a strategy). We also observed instances of fluency (when participants provided multiple rationales or strategies for a pursuit), originality (when a participant stated having to devise an inquiry or plan different than what had been taught in the workshop), and surprise (moments when the participants revealed something about their project or finding that is unexpected). More than one creativity facet could occur at each data mining stage. For illustrative purposes in this exploration, we present vignettes of two cases where creativity unfolded for Andrew and Janet.

Andrew was a 15-year-old who self-identified as a Chinese-American male with significant coding experience. He completed a CDS project that assessed public sentiments about programming languages. Based on what he reported about enacting the project, we observed evidence of four creativity facets (Table 1).

Table 1
Creativity Facets Evident in Andrew's Data Mining Presentation

Creativity Facets Illustrated with Andrew:	Complexity	Elaboration	Fluency	Originality
	✓	✓	✓	✓

Fluency (the number and variety of approaches or perspectives used to engage or carry out a process) was the most prominent in Andrew's presentation. He argued that programming is vital in several industries spanning medicine and astronomy, and yet there are misconceptions about the merits of each, as he explained:

The problem is that each programming language has its own strengths, but misconceptions ... may exist, and many may use un-ideal languages, or they might just not hear about tons of programming languages. So by analyzing the general sentiment around programming languages...I can...maybe reveal possible misconception[s]. (Andrew, 8/11/2023)

Andrew's presentation ideas and articulation of multiple justifications for why perspectival evidence could be useful (to resolve misconceptions and raise awareness) indicate his fluency on the subject. His topic was also original. When presenting his data processing strategy, Andrew described his challenge as the "hashtag assembly" issue (Andrew, 8/11/2023) and his steps to resolve it, which included using code to identify and remove retweets and duplicates. Then, he explained how he used code to make word clouds of the remaining output, as he noted:

So to better analyze the problems [with assembly] I generated word clouds, and these help to show that spam had not been fully removed and where to look for spam...Like this Java one

which has [words like] ‘painkillers,’ ‘pill,’ ‘tablets,’ (shown in figure 1a) and many other things unrelated to the Java programming language. [So I] found that the spam that escaped the first spam filter was because it contained links. And every time the spammer sent a link, the link would defer each tweet. So then that would interfere with the code [I used], which only removed absolutely, completely identical tweets. So I just removed all the links from the tweets. And then I removed the links, and then I removed the duplicates, which will [remove the] spam. And after removing some very specific few users who escaped [my] filters, the hashtag Java word cloud looked like that (shown in figure 1b), which is a lot more accurate to what people are posting. (Andrew, 8/11/2023)

Andrew described an original approach for processing data because he explained using the word clouds to assess data structure rather than to visualize text after analysis as was taught in our workshop (see Figures 1a and 1b for before and after results). Andrew’s originality is also illustrated in his computational elaboration and strategy, wherein he described a detailed plan he devised to eliminate duplicate tweets, which he described as “spam,” or tweets where platform users found ways to over-represent ideas or bypass the filtering strategies. Andrew’s agency is reflected in his enactment of creativity because his explanations revealed that he leveraged what he learned as a tool to carry out a novel data processing sequence aligned with and suited to his epistemic priorities.

Figure 1

Andrew’s Originality in Data Assessment/Validation Using Visualizations (Word Clouds) to Identify and Remove Off-Topic Tweets.



The word clouds visualize his data before (a) and after (b) erroneous tweets are removed.

Andrew also shared his strategies for preparing data for analysis and his steps to restructure data frames conducive to the analyses he wished to enact. Specifically, Andrew sought to automate sentiment assessments. To accomplish this, he noted several challenges he’d need to resolve, as he explained:

I also wanted to do ... some basic automation to try and detect sentiment...the code proved to be quite difficult...initially, I use that code [but it] wouldn't work as intended because if a tweet contain multiple keywords...then the program would think that there was one tweet [that contained the first and another that contained the second]. Another issue inherent to... using automation, like this way, is that it's going to be not as accurate as if I did it manually. But I had over 900 tweets, so it will be impractical to go through all of them manually, at least as a single individual. I then adopted this code. So I created data frames that contained [only one key word,] and created pivot tables based on those data frames, which allowed me to successfully do what I wanted. (Andrew, 8/11/2023)

Andrew’s fluency is evident in multiple ways in which he conceptualized challenges with using code to automate analysis of a large set of Tweets as he explained his understanding of how the code handles the data set (i.e., treating multiple keywords in a Tweet as multiple Tweets) and how that handling could introduce output inaccuracies (i.e., overestimating or overrepresenting a sentiment). His fluency is also indicated in how he expresses this consideration against the related issue of practicality—because manually analyzing is not feasible for the 900 tweets he wished to examine for the workshop. He described resolving this tension by devising a strategy that uniquely pieced together various ideas from the course—using code and search terms to isolate data—to devise multiple data frames that circumvented the challenge identified. Andrew’s explanations can be understood here as the extent to which he could enact an epistemic process creatively. Along with originality and elaboration, we observed many instances of the fluency facet of creativity in his descriptions. We noted this indicator at several data mining phases, including how Andrew explained his research inquiry and its relevance.

It is also reflected in the strategies he stated using to process and disentangle complex data sets to carry out analyses that were sometimes beyond what was possible in what we provided as starting points in the workshop.

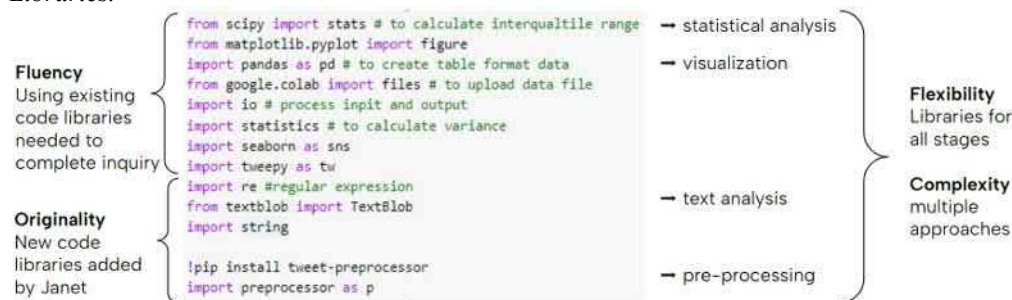
Our second case involved Janet, a 16-year-old skilled and experienced programmer who self-identified as an African-American female. Her presentation showed complexity, flexibility, fluency, originality, and surprise (Table 2). Like all participants, Janet pursued an original line of inquiry (to assess public and educator opinions about using artificial intelligence (AI) technologies in education). Like the prior case with Andrew, the approaches Janet described reflected her fluency in computing and the process of data mining for a personally relevant inquiry.

Table 2
Creativity facets evident in Janet's Data Mining Presentation.

Creativity Facets Illustrated with Janet:	Complexity	Flexibility	Fluency	Originality	Surprise
	✓	✓	✓	✓	✓

We observed fluency and originality in the steps she stated taking to prepare for the data mining process. She explained how she called a series of relevant and unique code libraries as a starting step in a plan to explore her ideas. Janet's explanations also showed flexibility (because she identified and called code libraries that enabled her to carry out tasks along the entire inquiry arc) and complexity (because she stated using a variety of approaches to achieve single or emergent outcome research questions aligned strategies). These creativity indicators are annotated in Janet's annotated slide shown in Figure 2.

Figure 2
Janet's Originality, Fluency, Flexibility, and Complexity in Her Use of New and Existing Code Libraries.



Janet explained that after she called the code libraries needed to conduct the research inquiry, she scraped and processed the X/Twitter data. In doing this, Janet shared about her decisions to reduce or “clean” the data set to be sure viewpoints were fairly represented before analysis as she explained, “so then, I decided to remove tweets with a massive amount of retweets as it skewed the data...I didn't want to have an opinion that was so popular that it kind of overshadowed the rest of them because I wanted to make sure that everything was represented fairly” (Janet, 8/11/2023). Here, she explained the concept of skew and how outliers can create skew. Janet learned this topic in the workshop but in the context of quantitative statistics. This indicates originality because her explanation showed that she was using the understanding in a novel way. When Janet described the visualizations produced in the analysis, we observed fluency because she used several (e.g., word clouds, pie charts, and histograms) to communicate a cohesive narrative about her observations during the analysis. Janet explained how she first used a word cloud to explore expressed sentiments and identify potential irregularities in her collected data, such as erroneous opinions or ideas unrelated to her research question. This new use was also original. Following this, she explained how she leveraged a pie chart to illustrate the heterogeneity of opinions for and against AI use in education and representative proportions of those viewpoints in her sample. In this finding, Janet made an unexpected observation as she explained, “Interestingly, negative opinions, such as cheating, were held not really, as much as I would expect. Yes, it was a little higher [than] non-educational users, but generally, it was lower than I thought.” (Janet, 8/11/2023). Janet explained that as she explored a related question about why the viewpoints in her sample opposed the use of AI, she found a surprising insight—that academic integrity is not a common concern represented in the sample she collected. This observation, characterization, and use as an “interesting” insight indicate the surprise creativity facet because Janet emphasizes an apparent paradox (uncommon concerns about ChatGPT cheating).

The approaches Andrew and Janet explained indicated how creativity manifested within each data mining stage and across the trajectory of their inquiries. Because the workshop was designed to support open inquiry, participants could enact epistemic strategies consistent with their epistemological styles and priorities. These acts could be identified and understood by participants' creativity in this process. In the next section, we discuss our observations in these cases in relation to the current literature.

Epistemic agency, creativity, and the relationship between the two

This research is motivated by a need to design and identify educational experiences that situate learners as knowledge producers (Damşa et al., 2014) rather than participants in the productions or inquiries of others. We address this need in computing education, where achieving public literacy (Tang & Williams, 2019) and participation equity (Stroupe, Moon & Michaels, 2019) are persistently elusive. While there have been steady inroads to address these issues by expanding access to fields such as CDS (Jiang, Lee, Rosenberg, 2022), there is a dearth in the literature on how to enact these areas in ways that support and sustain generative moments in epistemic agency (Odden, Silva & Malthe-Sorensen, 2023)—which in turn can encourage learner engagement, interest, and motivation. Consistent with the central aim of this exploratory study, we address these issues by identifying where we observed youth describe actions that can be understood as acts of creativity (Sternberg & Kaufman, 2010) when participating in an open inquiry CDS workshop. We highlight two illustrative case examples of Alex and Janet because the explanations provided about their inquiry processes revealed creativity facets (Guilford, 1956). This is remarkable because it suggests creativity is a plausible vantage point for identifying when learners engage with agentic and generative processes. We observed this specifically in how participants described enacting their projects, which indicated their originality and fluency in flexible, complex, elaborate, and sometimes surprising ways. In other words, we found that creative acts in an epistemic inquiry can indicate generative forms of agency since learners can make decisions consistent with their epistemological styles (Resnick & Silverman, 2005), priorities, and needs. We contribute to the literature on epistemic agency and creativity in two ways. First, we provide starter insights about epistemic practices that can serve as creativity indicators at various stages of open inquiry in CDS. This builds on Odden, Silva & Malthe-Sorensen's (2023) work to include computational data mining and operationalizing creativity to nuance how we understand the merits of inquiry practice that is generative. Secondly, we provide initial insights into how we might broaden conceptualizations of what is considered mastery in an academic discipline or community of practice to include indicators beyond instantiations that emphasize creativity in what a learner might produce and include practices or behaviors a learner might enact.

Considering the relationship between epistemic agency and creativity, our findings showed a close connection based on participants' presentations of their data mining processes as reflected in their final product. Epistemic agency, the capacity for individuals to act independently and make choices and decisions (Odden, Silva & Malthe-Sorensen, 2023; Rosenberg & Jones, 2022), plays a pivotal role in fostering creativity (Rhodes, 1961; Sternberg & Kaufman, 2010). When participants feel a sense of agency, they are empowered to explore new ideas, take risks, and experiment with different approaches (Ryan & Deci, 2002). This autonomy allows for a more fluid and open creative process, as rigid structures or external limitations do not constrain participants. Moreover, agency provides the freedom to engage with one's unique perspective, interest, and experiences which may lead to novel and imaginative solutions (Damşa et al., 2014; Kieran, 2018). Creativity, in turn, can enhance a sense of agency by reinforcing the belief that one's ideas and actions have the potential to make a meaningful impact (Sternberg & Kaufman, 2010). This reciprocal relationship between agency and creativity highlights the importance of fostering environments that encourage individual autonomy (Ryan & Deci, 2002), as it lays the groundwork for the emergence of innovative and original ideas. The nature of the relationship between epistemic agency and creativity needs to be further examined. However, based on these initial findings in our study, we postulate that epistemic agency may lead to creativity and, *vice versa*, creativity can be a driving force for enacting agency. To make creative products, one has to go through a process (Rhodes, 1961) that is usually agentic. We do not think these two constructs are isolated, one-time episodes during a learning experience. Instead, as mentioned earlier, creativity and agency might have a cyclical relationship wherein one amplifies or invites the other, leading to an ongoing engagement (flow; Csikszentmihalyi, 1997) with internal motivation (SDT; Ryan & Deci, 2002). More work is needed to explore the relationship between agency and creativity, examining moment-to-moment processes. For instance, we could gather data from learners' interactions with the learning environment to reveal the process they went through from a more systematic point of view. This may include direct observation across the workshop experience to assess moments of agency and creativity iterations that may not have appeared in final presentations. Future work along these initial insights might also include studies to validate the relationship between the epistemic agency and creativity and evaluate the efficacy of CDS epistemic practices to shape formal learning, social, and affective outcomes.

References

- Acquah, A., Barany, A., Scarola, A., Johnson, M. A., Reza, S. M., Rivera, C., & Walker, J. T. (2024, October). Cultural Relevance for Epistemic Practice in High School Computational Data Mining. In *2024 IEEE Frontiers in Education Conference (FIE)* (pp. 1-9). IEEE. <https://doi.org/10.1109/FIE61694.2024.10893416>.
- Barany, A., Reza, S., Johnson, M., Barrera, A., Badreddin, O., Fuentes, C., & Walker, J. T. (2023). Towards the design of a culturally relevant curriculum for equitable, data mining-based CS education. In *Blikstein, P., Van Aalst, J., Kizito, R., & Brennan, K. (Eds.), Proceedings of the 17th International Conference of the Learning Sciences - ICLS 2023* (pp. 1498-1501). International Society of the Learning Sciences. <https://doi.org/10.22318/icls2023.614567>.
- Barany, A., Scarola, A.D., Acquah, A., Reza, S.M., Johnson, M.A. and Walker, J. (2024), Learning Designs that Empower: Navigating Sandbox Data Science at the Intersection of Computing, Big Data, and Social Media. *Information and Learning Sciences*, Vol. 125 No. 10, pp. 794-812. <https://doi.org/10.1108/ILS-12-2023-0211>.
- Bazzaz Abkenar S., Haghi Kashani M., Mahdipour E., Jameii S.M. (2021). Big Data Analytics Meets Social Media: A Systematic Review of Techniques, Open Issues, and Future Directions. *Telematics and Informatics*, Volume 57. <https://doi.org/10.1016/j.tele.2020.101517>.
- Brown, J. S., Collins, A., & Duguid, P. (1989). Situated Cognition and the Culture of Learning. *Educational Researcher*, 18(1), 32-42. <https://doi.org/10.3102/0013189X018001032>.
- Buechley, L., Eisenberg, M., Catchen, J., & Crockett, A. (2008, April). 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). <https://doi.org/10.1145/1357054.1357123>.
- Csikszentmihalyi, M. (1997). Flow and the Psychology of Discovery and Invention. *HarperPerennial*, New York, 39, 1-16. <https://awspntest.apa.org/record/1996-97915-000>.
- Damşa, C. (2014). Shared Epistemic Agency and Agency of Individuals, Collaborative Groups, and Research Communities. In *Joseph L. Polman, Eleni A. Kyza, D. Kevin O'Neill, Iris Tabak, William R. Penuel, A. Susan Jurow, Kevin O'Connor, Tiffany Lee, and Laura D'Amico (Eds.), Learning and Becoming in Practice: The International Conference of the Learning Sciences (ICLS) 2014*. Volume 1. Colorado, CO: International Society of the Learning Sciences, pp. 440-447. <https://doi.org/10.22318/icls2014.440>.
- Education & Human Resources. (2020). STEM Education for the Future: A Visioning Report. National Science Foundation. <https://www.nsf.gov/edu/Materials/STEM%20Education%20for%20the%20Future%20-%202020%20Visioning%20Report.pdf>.
- Guilford, J. P. (1956). The Structure of Intellect. *Psychological Bulletin*, 53(4), 267. <https://psycnet.apa.org/doi/10.1037/h0040755>.
- Jiang, S., Lee, V. R., & Rosenberg, J. M. (2022). Data Science Education Across the Disciplines: Underexamined Opportunities for K-12 Innovation. *British Journal of Educational Technology*, 53(5), 1073-1079. <https://doi.org/10.1111/bjet.13258>.
- Karwowski, M., Zielińska, A., & Jankowska, D. M. (2022). Democratizing Creativity by Enhancing Imagery and Agency: A Review and Meta-Analysis. *Review of Research in Education*, 46(1), 229-263. <https://doi.org/10.3102/0091732X221084337>.
- Kieran, M. (2018). Creativity as an Epistemic Virtue. *Routledge Handbook of Virtue Epistemology*, 167-177.
- King, B. R., & Satyanarayana, A. (2013, June). Teaching Data Mining in the Era of Big Data. In *2013 ASEE Annual Conference & Exposition* (pp. 23-1140). <https://doi.org/10.18260/1-2--22525>.
- Ko, M. L. M., & Krist, C. (2019). Opening up Curricula to Redistribute Epistemic Agency: A Framework for Supporting Science Teaching. *Science Education*, 103(4), 979-1010. <https://doi.org/10.1002/scs.21511>.
- Miller, E., Manz, E., Russ, R., Stroupe, D., & Berland, L. (2018). Addressing the Epistemic Elephant in the Room: Epistemic Agency and the Next Generation Science Standards. *Journal of Research in Science Teaching*, 55(7), 1053-1075. <https://doi.org/10.1002/tea.21459>.
- Odden, T. O. B., Silvia, D. W., & Malthe-Sørenssen, A. (2023). Using Computational Essays to Foster Disciplinary Epistemic Agency in Undergraduate Science. *Journal of Research in Science Teaching*, 60(5), 937-977. <https://doi.org/10.1002/tea.21821>.
- Rahimi, S. (2020). Inspire, Instruct, or Both? Game-Based Assessment and Support of Creativity (Order No. 28153346). Available from ProQuest Dissertations & Theses Global. (2490711365). <https://utep.idm.oclc.org/login?url=https://www.proquest.com/dissertations-theses/inspire-instruct-both-game-based-assessment/docview/2490711365/se-2>.
- Rahimi, S., & Shute, V. J. (2021). First Inspire, Then Instruct to Improve Students' Creativity. *Computers & Education*, 174, 104312. <https://doi.org/10.1016/j.compedu.2021.104312>.

- Rahimi, S., Walker, J. T., Lin-Lipsmeyer, L., & Shin, J. (2023). Toward Defining and Assessing Creativity in Sandbox Games. *Creativity Research Journal*, 0(0), 1–19. <https://doi.org/10.1080/10400419.2022.2156477>.
- Ravitch, S. M., & Carl, N. M. (2019). *Qualitative Research: Bridging the Conceptual, Theoretical, and Methodological*. Sage Publications.
- Resnick, M., & Silverman, B. (2005, June). Some Reflections on Designing Construction Kits for Kids. In *Proceedings of the 2005 Conference on Interaction Design and Children* (pp. 117-122). <https://doi.org/10.1145/1109540.1109556>.
- Rhodes, Mel. (1961). An Analysis of Creativity. *The Phi Delta Kappan*, vol. 42, no. 7, pp. 305–10. JSTOR, <http://www.jstor.org/stable/20342603>.
- Romero, C., & Ventura, S. (2013). Data Mining in Education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 3(1), 12-27. <https://doi.org/10.1002/widm.1075>.
- Rosenberg, J. M., & Jones, R. S. A (2022) Secret Agent? K-12 Data Science Learning Through the Lens of Agency. <https://doi.org/10.35542/osf.io/eyzkv>.
- Ryan, R. M., & Deci, E. L. (2002). Overview of self-determination theory: An organismic-dialectical perspective. In E. L. Deci & R. M. Ryan (Eds.), *Handbook of self-determination research* (pp. 3–33). University of Rochester Press. <https://psycnet.apa.org/record/2002-01702-001>.
- Sánchez-Dorado, Julia (2020). Novel & worthy: creativity as a thick epistemic concept. *European Journal for Philosophy of Science* 10 (3):1-23. <https://doi.org/10.1007/s13194-020-00303-y>.
- Scardamalia, M., & Bereiter, C. (1991). Higher Levels of Agency for Children in Knowledge Building: A Challenge for the Design of New Knowledge Media. *Journal of the Learning Sciences*, 1(1), 37–68. https://doi.org/10.1207/s15327809jls0101_3.
- Sternberg R.J., Kaufman J.C. (2010). Constraints on Creativity: Obvious and Not So Obvious. In *Kaufman JC, Sternberg RJ, eds. The Cambridge Handbook of Creativity*. Cambridge Handbooks in Psychology. Cambridge University Press; 2010:467-482. <https://doi.org/10.1017/CBO9780511763205.029>.
- Stroupe, D., Moon, J., & Michaels, S. (2019). Introduction to Special Issue: Epistemic Tools in Science Education. *Science Education*, 103(4), 948-951. <https://doi.org/10.1002/sce.21512>.
- Tang, K. S., & Williams, P. J. (2019). STEM Literacy or Literacies? Examining the Empirical Basis of these Constructs. *Review of Education*, 7(3), 675-697. <https://doi.org/10.1002/rev3.3162>.
- Torrance, E. P. (1972). Can We Teach Children to Think Creatively? *Journal of Creative Behaviour*. 6 (2), 114-143. <https://doi.org/10.1002/j.2162-6057.1972.tb00923.x>.
- Turkle, S., & Papert, S. (1990). Epistemological Pluralism: Styles and Voices within the Computer Culture. *Signs: Journal of Women in Culture and Society*, 16(1), 128-157. <https://doi.org/10.1086/494648>.
- Varelas, M., Tucker-Raymond, E., & Richards, K. (2015). A Structure-agency Perspective on Young Children's Engagement in School Science: Carlos's Performance and Narrative. *Journal of Research in Science Teaching*, 52(4), 516-529. <http://dx.doi.org/10.1002/tea.21211>.
- Walker, J., Slater, S., & Kafai, Y. (2019). A Scaled Analysis of How Minecraft Gamers Leverage YouTube Comment Boxes to Participate and Collaborate. In *Lund, K., Niccolai, G. P., Lavoué, E., Hmelo-Silver, C., Gweon, G., & Baker, M. (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 1* (pp. 440-447). Lyon, France: International Society of the Learning Sciences. <https://doi.org/10.22318/cscl2019.440>.
- Walker, J.T., Barany, A., Acquah, A., Barrera, A., Reza, S., Del Rio Guzman, K., O. & Johnson, M. (2023a). Coding Like a Data Miner: A Sandbox Approach To Computing-Based Data Science for High School Student Learning. *EngageCSEdu*. ACM. <https://doi.org/10.1145/3631986>.
- Walker, J.T., Barany, A., Barrera, A., Reza, S., Guzman Del Rio, K., Acquah, A., Badreddin, O., Johnson, M. (2023b). Coding Like a Data Miner: A Sandbox Approach to Computing-Based Data Science for High School Student Learning. In *2023 IEEE Frontiers in Education Conference (FIE)*, College Station, TX, USA, 2023, pp. 1-5, <https://doi.org/10.1109/FIE58773.2023.10343283>.

Acknowledgments

This work was partly supported by a grant from the National Science Foundation (#2137708). Any opinions, findings, conclusions and/or recommendations expressed herein are those of the authors and do not necessarily reflect the views of the National Science Foundation, University of Maryland, University of Florida, University of Pennsylvania, or Southern Methodist University.