SmileyCluster: Supporting Accessible Machine Learning in K-12 Scientific Discovery

Xiaoyu Wan[1], **Xiaofei Zhou**[1], **Zaiqiao Ye**[2], **Chase K. Mortensen**[3], **Zhen Bai**[1]
[1]University of Rochester, United States,
{xwan3@u.rochester.edu, xzhou50@ur.rochester.edu, zbai@cs.rochester.edu}
[2]Indiana University of Bloomington, United States, zaiqye@iu.edu
[3]Utah State University, United States, chase.mortensen@usu.edu

ABSTRACT

There is an increasing need to prepare young learners to be Artificial Intelligence (AI) capable for the future workforce and everyday life. Machine Learning (ML), as an integral subfield of AI, has become the new engine that revolutionizes practices of knowledge discovery. Making ML experience accessible to young learners, however, remains challenging due to its high demand for mathematical and computational skills. This research focuses on designing novel learning environments that help demystify ML technologies for K-12 students, and also investigating new opportunities for maximizing ML accessibility through integration with scientific discovery in STEM education. We developed *SmileyCluster* - a hands-on and collaborative learning environment that utilizes glyph-based data visualization and superposition comparative visualization to assist learning an entry-level ML technology, namely k-means clustering. Findings from an initial case study with high school students in a pre-college summer program show that SmileyCluster leads to positive change in learning ML concepts, methods and sense-making of patterns. Findings of this study also shed light on understanding ML as a data-enabled approach to support evidence-based scientific discovery in K-12 STEM education.

Author Keywords

Data visualization; hands-on learning; AI literacy; scientific discovery; STEM education

CCS Concepts

•**Human-centered computing** → **Human computer interaction (HCI);** *HCI design and evaluation methods;* User studies;

2020 2020 [IDC '20]Interaction Design and ChildrenJune 21–24, 2020London, United Kingdom Interaction Design and Children (IDC '20), June 21–24, 2020, London, United Kingdom 10.1145/3392063.3394440
978-1-4503-7981-6/20/06

Figure 1. Two students interacting with SmileyCluster system.

INTRODUCTION

Children growing up in the Artificial Intelligence (AI) era - the so-called "Generation AI" [16] - will be most impacted by the ever-growing advances in AI technologies. Machine Learning (ML) is an integral subfield of AI that helps computers identify patterns and make predictions from empirical data [65]. As the data revolution accelerates, ML plays an increasingly important role in everyday life such as education, healthcare, and transportation. It is critical, therefore, that young learners acquire sufficient ML literacy and become accustomed to data-enabled problem-solving in order to navigate in this increasingly technical and intelligent world.

AI education traditionally has a high demand for mathematical and computational skills. Emerging learning environments such as Cognimates [13] and eCraft2Learn [31] have shown promising learning experiences for young students to experiment with key AI technologies such as text understanding, image and speech recognition, robot control and model training [57]. Obstacles remain, however, to make AI education more accessible for K-12 students with diverse STEM education backgrounds. **First**, recent studies show that students with limited programming experience [31], especially those from low and medium socio-economic schools [13], struggle to advance their AI understanding. This is because most of existing K-12 AI learning environments require knowledge of block-based visual programming such as Scratch that not all students possess due to wide variation in K-12 CS education [61]. **Second**, while students can learn the basic training-testing pipeline of how machines learn from data, the underlying ML concepts (e.g. multi-dimension feature space, similarity comparison) and detailed ML methods (e.g. clustering, classification) remain inside the black-box. Understanding these concepts and methods will help learners not only make meaningful ML applications but also understand the basics of how ML works. The latter is critical to prepare young learners to be more sensible with respect to decision-making and ethical aspects of AI such as trust and fairness [24].

This study aims to maximize learning opportunities of ML for students with diverse STEM skills, and look inside the black-box to demystify ML. We propose to explore (1) design space of ***data visualization*** [11], ***hands-on exploration*** [36] and ***collaborative learning*** [8], which have shown promising benefits for deeper understanding of abstract and complex concepts via diagrammatic reasoning, conceptual metaphor, and offloading cognition for young students [23, 40, 41, 38, 30]; and (2) pedagogical opportunities of ML-empowered ***scientific discovery***. The natural connection between ML and ***practices of science*** such as "analyze and interpret data", "describe patterns and relationships", "make sense of phenomenon" [9] align well with the global urgency of cohesion of STEM education [32]. Meaningful STEM contexts, in turn, may motivate AI education through a sense of authenticity and real-world applicability [64]. This study focuses on two research questions:

(1) RQ1: How to design learning environments that support understanding the ML concepts and methods of k-means clustering for young students with diverse mathematical and programming skills?
(2) RQ2: How can the ML learning environment enhance scientific discovery in STEM contexts?

We developed ***SmileyCluster*** (Fig. 1) - a web-based collaborative learning environment that supports (1) learning basic ML concepts and methods involved in a particular ML technology - namely ***k-means clustering***, and (2) sense-making of scientific phenomena and scientific inquiry such as question asking, justification and argument [51]. The rationale for choosing k-means clustering as the initial study are three-fold. First, k-means clustering is one of the most common similarity-based ML methods and is central to basic ML concepts of multi-dimension feature space and similarity comparison. Understanding these concepts may lead to better understanding of other similarity-based ML methods such as k-nearest neighbor classification, information retrieval and anomaly detection. Second, k-means clustering is widely used in scientific knowledge discovery across STEM subjects such as biology [33] environmental science [1], and chemistry [45], which may lead to a tight connection with K-12 STEM learning. Third, due to the unsupervised nature of k-means clustering, it may promote data exploration and hypothesis generation of scientific phenomena [14], both of which are key scientific inquiry practices in K-12 STEM classes [9].

SmileyCluster introduces a novel conceptual metaphor called **face-overlay** that addresses the two main obstacles for K-12 students to learn k-means clustering: efficient similarity comparison (due to lack of knowledge of Euclidean distance on multi-dimension feature space [18]), and global understanding of data as a whole (as commonly struggled by young learners [3]). Face-overlay is inspired by ***face glyph data visualization*** [7, 17], which translates each attribute value of a data point to a visual element of a cartoon face such as position of eyes and shape of mouth [63], and also by ***superposition comparative visualization*** [23], which shows different visual objects in the same space through information overlay.

As an initial investigation, we conducted a case study with high school students in a data science class of an on-campus pre-college summer program. Findings of this study provide rich evidence that (1) the face-overlay metaphor effectively supports similarity comparison and global understanding of data for high school students; (2) the SmileyCluster system positively supports learning and interpreting k-means clustering. In addition, the participants also carried out meaningful scientific inquiry behaviors through interaction with the SmileyCluster system in pairs. This research contributes new knowledge in three-fold:

(1) Create a new learning environment to make k-means clustering accessible to young learners with limited mathematical and computational skills.
(2) Provide a novel conceptual metaphor to inform the design of future learning environments for similarity-based ML technologies.
(3) Gain insights about the role that ML can play in offering new learning opportunities of scientific discovery in K-12 STEM education.

RELATED WORK

K-12 AI education

Thanks to recent initiatives such as AI4K12 and AI4All, there are emerging learning environments that expose K-12 students to high-level AI concepts in areas such as text understanding, object recognition, speech recognition, and robot control through engaging activities such as gameplay, robot construction, drawing and interactive visualization [52, 4, 57]. Among the five big ideas of AI education proposed by the AI4K12 initiative, "computers can learn from data" has the most direct connection with knowledge discovery. Under this big idea, however, only a very narrow subfield of ML, namely image processing, is supported by existing AI learning technologies. This reveals a vast gap of effective learning scaffolding to help young learners obtain basic ML concepts and technologies that support a diverse range of pattern recognition and inference activities, which in turn lead to discovery and constructing new knowledge across different subject domains.

ML-Empowered K-12 Science Practice

In a recent study, Zimmermann-Niefield and colleagues [67] connected ML and youth science education through modeling, which is a key set of scientific and mathematical practices defined in the High School Common Core State Standards (CCSS) [18] and Next Generation Science Standards (NGSS) [9]. They demonstrated an approach that supports participants to build and evaluate ML models through sports play and scientific exploration. Besides modeling, more connections can be used for the design of AI-empowered learning environments for K-12 science practices, such as analyzing and interpreting data, asking questions, constructing explanations and engaging in argument [9, 2]. Our research proposes a design solution that stimulates more K-12 students' science practice behaviors described above by learning basic ML concepts and methods.

Data Visualization

Data visualization is an umbrella of technologies that facilitate knowledge discovery and sense-making through visual data exploration [11] and diagrammatic reasoning [19, 40]. In addition, data visualization is also commonly used to explain machine learnt patterns and models, such as clustering, decision trees, neural networks, and Bayes models [11, 15]. Glyph-based visualization is a special type of data visualization that maps individual data points to a familiar graphical representation such as a star, stick figure, leaf, bug or weathervane [63, 20]. Face glyph [7, 17] is a particular type of glyph-based data visualization that takes advantage of people's high processing capacity and sensitivity to human faces, and translates each attribute value of a data point to a visual element of a cartoon face such as position of eye, slant of eyebrow, shape of mouth, etc. Researchers have explored opportunities of using face glyphs to help people carry out visual reasoning in domains like ecology [34], sport [54], health assessment [26], marketing [5] and software security [60]. Children are naturally drawn to faces within the first few months after birth [35], and keep improving visual research ability of faces. While previous research shows that face glyph provides an alternative learning method for students in the science classroom to compare agricultural products in different regions [43], we are not aware of learning environments that apply face glyph to support learning of ML concepts and use of ML-empowered methods to support scientific discoveries.

Comparative visualization

Comparison is a common task in data visualization, with two well-established visualization methods - juxtaposition and superposition [23]. Juxtaposition design arranges individual objects for side-by-side comparison, while superposition design shows different objects in the same space through efficient information overlay. Compared to juxtaposition, which relies on frequently shifting attention between objects and is hard to scale to a large number of objects perceptually, superposition minimizes memory requirements and attention shifting by integrating multiple facets of information within one coherent visual space [58]. It is widely used in geographical information systems, which utilize the map-overlay metaphor to superimpose different thematic layers of information such as roads and hydrology [66]. While glyph-based visualization mainly relies on side-by-side comparison, we argue that superposition is particularly suitable for supporting similarity comparison as data represented in individual glyphs are "similar enough to one another that they can be viewed on the same plane to detect similarity and difference between objects" [23].

Hands-on and Collaborative Learning

Constructivism [46] and distributed cognition [25] theories emphasize the active process of meaning-making in learning, particularly through perception, manipulation, and social interaction with visual representations [50, 47].Manches and O'Malley (2012) argue that hands-on learning with digital manipulatives can promote knowledge acquisition through conceptual metaphor that bridges abstract concepts with perceptual (e.g. size, color) and manipulative properties (e.g. pick, contain) of objects that people are familiar with, therefore offloading cognitive tasks and enabling additional attention [28, 44, 37]. In addition, research shows that social interactions facilitate students' conceptual change in learning complex concepts [21] and construction of new scientific knowledge [22, 49, 29] through scientific argumentation and false inference evaluation [39, 48]. In this study, we explore to what extent direct manipulation of data points can facilitate learning of abstract ML concepts and support scientific discovery in collaborative learning.

DESIGN OF SMILEYCLUSTER

Learning Goals and Barriers for K-means Clustering

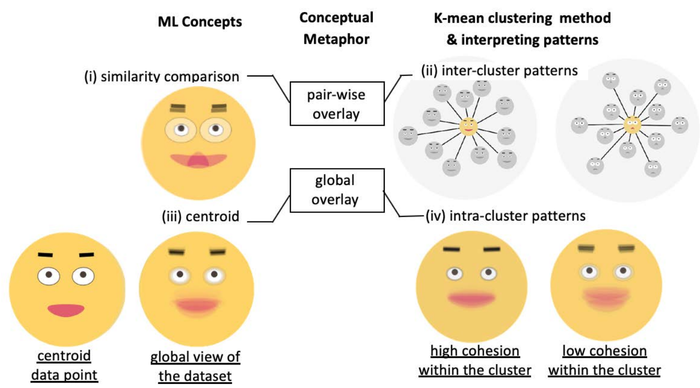
SmileyCluster aims to (1) support learning of entry-level ML concepts (multi-dimension feature space, similarity) and k-means clustering method (centroid computation, evaluate cluster number 'k' based on intra-cluster cohesion (similarity of data points within a cluster) and inter-cluster separation (difference of data points between clusters)) [56]; (2) support learning of pattern-interpretation (explain clusters, shared features within the same cluster, and differentiating features between clusters); and (3) enhance scientific inquiry learning behaviors such as question asking, explanation, and argument [9]. K-means clustering is a ML method by which each data point is assigned to the cluster with the closest centroid. We identify two main learning barriers for young learners to obtain these key ML concepts and related k-means method:

1. **Similarity comparison of multi-dimension feature space**: Real-world phenomena such as species and habitats require multi-dimensional features to describe their characteristics, but the common core high school mathematics curriculum only covers patterns of association in bivariate data [18]. Besides, Euclidean distance on multi-dimension feature space is a basic similarity comparison approach, but high school students are only introduced to two-dimensional (plane) Euclidean geometry [18].

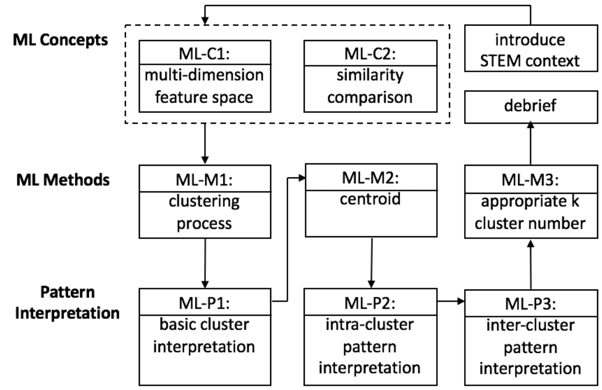
2. **Global understanding of a dataset**: It's often difficult for young students to develop a global understanding of a group of data as a whole. For example, research in math classrooms shows that middle school students are inclined to develop local comparisons of pointwise differences of data, while needing additional scaffolding to build a global view of the dataset [3].

Face-overlay Conceptual Metaphor

Inspired by the face glyph and superposition comparative visualization, we propose the face-overlay conceptual metaphor to provide vivid explanations of similarity comparisons of multidimensional features and to make patterns more readily interpreted. We designed face glyph in the form of popular emoji-style faces (Fig. 2(a)) that helps comprehension, memory, and communication due to people's familiarity with facial features, and the simultaneous representation of a global view of all the features of a single data point.



(a) Overview of face-overlay conceptual metaphor



(b) SmileyCluster learning flow

Figure 2. An illustration of face-overlay conceptual metaphor and corresponding SmileyCluster learning flow.

Face glyph and cluster design

Based on existing face glyph methods [7, 12, 53], we selected 16 facial features: eye (eccentricity, angle, separation, vertical position, size), pupil (size, position), brow (slant, size, vertical position, horizontal position, density), mouth (vertical position, curvature, width, opening size), while excluding controversial features such as nose width and length [17], as well as hair and face shape features for simplicity. Although real-world phenomena may require a higher dimension of features, we argue that the current face glyph representation is sufficient to help novice learners to obtain basic ML concepts and methods, and experience new approach for authentic scientific discovery by interpreting machine learnt patterns.

For cluster representation, we adapted the structure-driven placement [62] and used radial placement to arrange individual faces belonging to the same cluster around the center face, which is a synthesized face glyph that represents the average features of the cluster, at various distances corresponding to the similarity between the individual face and the center face).

Overlay design

Face-overlay supports both pair-wise comparisons between data points in pairs and groups by drag-and-dropping the face one by one on top of another (pair-wise), or a single click to make faces of the same cluster stack together. When they are stacked, every face is semi-transparent, with the sum of opacities equal to 100%.

1. Pair-wise overlay aggregates two faces to contribute to learning basic ML concepts of similarity comparison (Fig. 2(a)(i)), interpreting inter-cluster patterns (e.g. comparing centroid faces of different clusters), and identifying an appropriate k cluster number through inspecting inter-cluster separation (Fig. 2(a)(ii)).

2. Global overlay aggregates multiple faces from the same cluster (Fig. 2(a)(iii)), to address learning challenges for young students to develop a global understanding of a dataset as a whole, by revealing richer information about the mean and deviation of the dataset through the level of blurriness of each facial feature. Global overlay is designed to support learning of centroid computation (average representation of all the data points belonging to the same cluster) (Fig. 2(a)(iii)), interpreting intra-cluster patterns (e.g. shared features within the same cluster), and identifying appropriate k cluster number through inspecting intra-cluster cohesion (Fig. 2(a)(iv)).

SmileyCluster Design

Based on the face-overlay conceptual metaphor, we developed SmileyCluster, a web-based learning environment that supports hands-on, playful and collaborative ML learning activities. To fulfill the learning goals of SmileyCluster, which aims to support students to understand and apply k-means clustering in science contexts, while maintaining a gradual learning curve, we design the learning flow as shown in Fig. 2(b). Following the learning flow, we describe the user interface matched with each learning goal that aligns with key learning components in cluster analysis.

Introducing a STEM context

The dataset used for the k-means analysis is the seed data adapted from UCI Machine Learning Repository [6], which is related to STEM field and the clustering result is validated by scientists. We started with the introduction of the dataset and how the data was collected.

ML-C1: Multi-dimension feature space

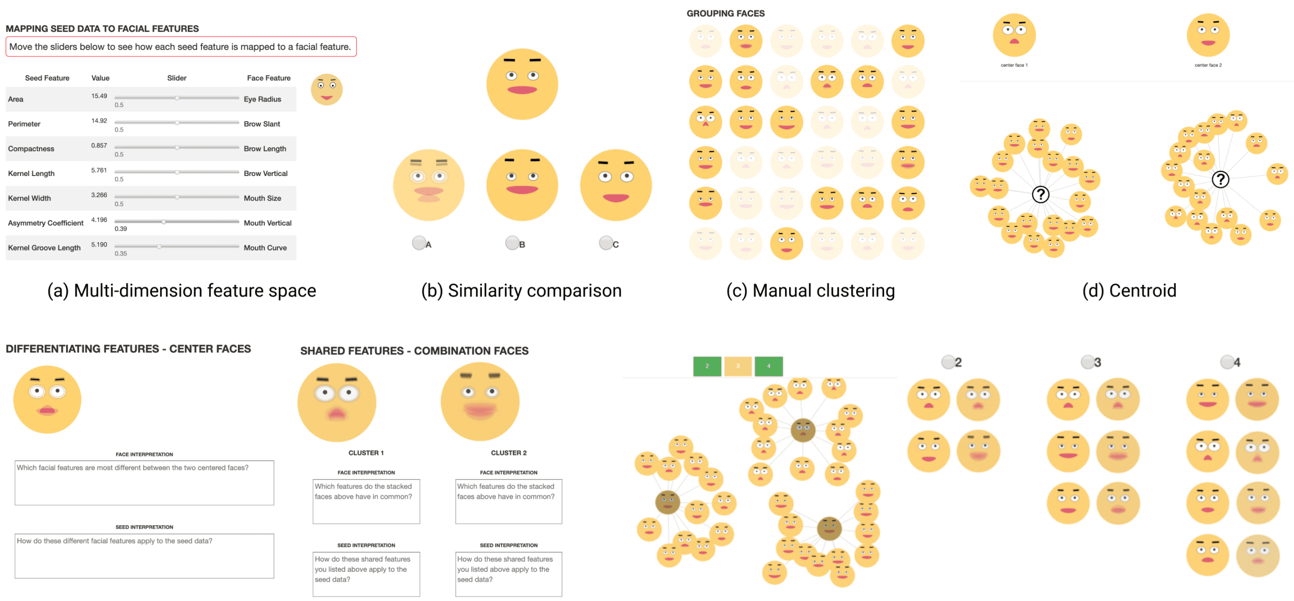
Fig. 3(a) shows two beginning steps while using the system. The interface design aims to achieve two goals: (1)introducing the dataset and the face mapping mechanism; (2) conveying the concept of multi-dimensional feature space.

ML-C2: Similarity comparison

The multiple-choice questions are designed to teach students about similarity comparison. While doing multiple choices, the face-overlay method is provided and serves as a way to help comparison. We provide a drag-and-drop function to put one face on the other (Fig. 3(b)).

ML-M1: Clustering process

The interface of grouping faces (Fig. 3(c)) is designed for students to explore the clustering process. By drag-and-drop, students can stack the faces that they think are similar together.



(a) Multi-dimension feature space (b) Similarity comparison (c) Manual clustering (d) Centroid

(e) Inter-cluster and intra-cluster pattern interpretation (f) Appropriate k cluster number

Figure 3. SmileyCluster Detailed Design.

It is a process for students to collaboratively discuss strategies to try grouping the faces and decide the k number.

ML-P1: Basic cluster interpretation

Following the face grouping activity, the system provides the clustering result generated by machine learning(k=2). We design the visualization specifically for the clusters of emojis. The cluster relationship is straightforward. The distance between one face and the center is generated according to the distance between data points. At this step, students are required to discuss in pairs their interpretation of the clustering results. The clusters are movable, in order to provide some freedom and fun while they discuss.

ML-M2: Centroid

Following the interpretation activity, we put a question mark on the center of each cluster, and raise a simple question for the students: Can you guess which center face belongs to which group? (Fig. 3(d)) By raising the question, we aim to bring up the centroid concept. After discussing, students can click on the question mark to reveal the answer.

ML-P2: Inter-cluster and intra-cluster pattern interpretation

We design two pages (Fig. 3(e)) to allow students to explore shared features within a cluster (left) and differentiating features between different groups (right). The students are asked to note down shared/differentiating facial features, as well as the corresponding seed dataset features to support sense-making of patterns in the actual science context.

ML-M3: Appropriate k cluster number

There are two pages (Fig. 3(f)) designed for students to decide the best number for k: (1) viewing different clustering by toggle the k number on top of the page (left); and (2) comparing the center face and combined faces of each cluster (right). The center faces are designed to help students be aware of the inter-cluster separation and the combined faces are designed to convey intra-cluster cohesion.

STUDY DESIGN

We conducted an initial investigation through a user case study with high school students to understand how effective SmileyCluster is in supporting (1) understanding the ML concepts and methods of k-means clustering,(2) sense-making with patterns, and (3) engaging students in scientific inquiry.

Participants

We recruited eight participants through the program coordinator of the on-campus pre-college summer program with a total of 12 students enrolled. There were 5 female and 3 male students, between 15-17 years old, with 5 domestic and 3 international students all with sufficient language skills for the class. The pre-study survey showed that 7 students have more than one year of programming experience at school or at home. Eight participants had a variety of AI knowledge or experience, through involvement in a Robotics club, some research experience in AI topics, or learning about the history of AI. We numbered our participants and will refer participants from P1-P8 in the following sections.

Procedure

The study took place in an on-campus computer lab and lasted about 2.5 hours, facilitated by one course instructor and four researchers. When the students showed up in the classroom, one researcher arranged for students who did not consent to be involved in the study to sit separately with participants of the study. Then the participants sat in pairs of their choice in front of a computer. The study procedure was (1) 25 minutes class instruction about AI and general difference between supervised and unsupervised learning, without revealing any content about clustering; (2) 15-min pre-study questionnaire about students' background in machine learning and cluster analysis; (3) 40-min interaction with the system; (4) 15-min post-study questionnaire and (5) 30-min focus group interview. All students in the class took part in the same learning activities, except we did not collect any observation data from students who did not take part in the study.

Data Collection

The research sites were set up by four researchers before the study. The web-based interface was set up on desktops in a lab that accommodated all 12 students in the summer camp, and each desktop was shared by a pair of participants for collaborative activities. We collected surveys, interviews, screen recordings, and video recordings of the system interactions from 8 participants. We set up two Canon Camcorders at the front and back of a desktop for each pair of participants to video record participants' interaction with each other, and interaction with the system accordingly.

Measures and Data Analysis

We measured the learning effects of SmileyCluster in supporting learning k-means clustering concepts and methods, sense-making of patterns, and scientific inquiry, through a combination of pre-post questionnaires for learning gains, observation of learning behaviors while participants interacted with the system and with each other, and semi-structured focus group interviews.

ML concepts, methods, and sense-making of patterns

Learning gains: We conducted pre-post questionnaires with written answers to questions relating to clustering, similarity comparison, center point, k-means clustering process and choosing an appropriate k number, and sense-making of patterns. Each answer was assigned 0 to 3 points according to the pre-designed rubric by researchers. Researchers graded the answers by matching participants' answers to the keywords listed in the rubric, with each match worth 1 point. Two researchers graded the questionnaires and reached inter-rater reliability with Cohen's kappa = 0.85 and 0.88 accordingly for pre- and post-study questionnaires. The final score of each question was computed by the means of two graders.

Learning behaviors: Learning behaviours are measured by two-fold: (1) we recorded participants' text-based answers to in-app questions as described in the section of "SmileyCluster Design" with ML concepts involved in [ML-C1, C2], methods in [ML-M1, M2, M3], and sense-making in [ML-P1, P2]; and (2) we recorded and jotted observation notes on participants' verbal behaviors on scientific inquiry episodes, including "question asking", "uncertainty", "argument", "justification", "suggestion", "sharing findings", "hypothesis generation", "evaluation", and "agreement" [51]. We adopted the qualitative case study method [10, 55], to conduct a holistic inquiry that investigates ML learning behaviours while using SmileyCluster. For in-app answers, we compared group's answers to multiple choice and open-ended questions, and with the suggested answers formulated by four researchers. For scientific inquiry learning behaviours, four researchers transcribed recordings into scripts, integrated observation notes with scripts to page-by-page analysis, and conducted open coding [10]. To establish a common set of codes and themes, we assigned two researchers to apply open and axial coding [10] to the same subset of transcripts independently, and then established a common understanding with a shared qualitative codebook. Each pair of researchers compared, discussed and stabilized the codes to reach 90% agreement. We then constructed cases [55, 10] for each research question in a second coding cycle. Each case was bounded by researchers' reflections on observational notes for each group, and by the page-by-page analysis for each learning component. Two researchers were assigned to synthesize the cases and generate the theme.

Semi-structured focus group interview: We obtained insights about participants' cluster learning experience ("When learning clustering analysis, what challenges you the most?"), glyph-based data visualization design ("Besides the face emoji, what other objects could you think of to map the multiple dimension data to?"), and the limitation of ML-based clustering ("Do you think that clustering analysis can always provide the best group result, or could the analysis have errors?"). Four researchers transcribed, coded the scripts, and established common set of codes and themes in a shared qualitative codebook. Each pair of researchers compared, discussed and stabilized the codes to reach 93% agreement. We analyzed interview scripts using the thematic analysis [59] approach to determine the high-level themes of participants' answers.

RESULTS

As shown in Table 1, results of the pre-post learning gain questionnaire indicate that the SmileyCluster system supported participants' learning of key ML knowledge components of k-means cluster, with a total score of the post-test increased by 7.13 points compared with the pre-test result.

Learning Outcomes for ML Concepts and Methods

Multi-Dimensional Feature Space

The face mapping mechanism appealed to and engaged participants in exploring and discussing the mapping relationship by moving the slider bar of the feature value to explore how each seed attribute is mapped to a facial feature (ML-C1 in SmileyCluster). For example, P7 and P8 moved each face feature between min and max values, and paused at the max values. P7 pointed, *"Look, I dragged it to extreme (values)"*. P8 laughed, and dragged another feature to extreme value.

We also found that participants were able to apply the feature mapping knowledge to other scenarios. When being asked for other objects they could map the features to during the interview, the participants provided answers such as "digital lego blocks", "Minecraft", "Cartoon characters", "trees", "stars", and "animal body or face".

Similarity Comparison

The pre-post learning gain questionnaire shows an increased understanding of the concept of similarity (Q1 to Q3). The correct rate for choosing the most similar face is 50% in the side-by-side comparison mode, and 100% in the overlay comparison mode with quicker completion. This shows that the face-overlay metaphor can efficiently support pair-wise visual comparison (Fig. 3(b)).

Questions (Scale 0-3)	$Mean_{Pre}(SD)$	$Mean_{Post}(SD)$	$Mean_{Diff}$
Q1. What does it mean to cluster a dataset?	1.06 ± 1.21	2.31 ± 0.88	1.25
Q2. What is the importance of similarity when clustering a dataset?	0.56 ± 0.90	1.75 ± 0.89	1.19
Q3. What makes two data points similar or dissimilar?	0.31 ± 0.88	1.75 ± 0.71	1.44
Q4. What is the center point of a group of data points?	0.63 ± 0.79	2 ± 0.93	1.37
Q5. Could you order the major steps for the k-means clustering algorithm?	0.88 ± 0.83	1.13 ± 0.99	0.25
Q6. Given two different numbers of groups for clustering the same dataset (e.g. one divides the dataset into 2 groups and one into 3 groups), how do you decide which number of groups gives a better result?	0.25 ± 0.46	1.88 ± 0.99	1.63

Table 1. Participants' learning gain scores in pre-test and post-test in 0-3 point scale.

We also found two different visual comparison strategies of pair-wise data points. With the global similarity comparison strategy, participants tend to compare the facial expression of the whole face. Group 2 used the global comparison strategy. P4: *"(The answer is) A, because (the face) is mostly going down."* P3: *"How about that one [points to C]."* P4: *"Hmm..it is definitely not B. I think C looks smiley instead of less worried than A and the [target face] also comes worriedly."*

With the local similarity comparison strategy, participants tend to compare by observing specific features, such as the eye size and mouth curve, in a sequential manner. Groups 1, 3 and 4 adopted this strategy. For example, in group 3, P6: *"How about this one? [points to A] The mouth is big and upside-down. And..this one? [points to C] How about this one? I feel like C. The eyebrows are the same size, the eyes are the same size. It is just..the mouth (is different)."*

Group 2 adopted the global strategy and selected the right answer, while the other three groups adopted the local comparison strategy with mixed results. However, we observed that groups adopting a local comparison strategy showed high uncertainty in comparing each feature and making decisions.

Centroid

The pre-post questionnaire shows an increased understanding of the concept of centroid (Q4). All four groups made the right choice of the center face (ML-M2 in SmileyCluster). We found that participants are able to identify the relationship between the center face and the cluster it belongs to. We observed some participants drag the cluster close to the average face for comparison. They also tended to discuss the center face by dragging two clusters closer to each other. P2: *"We'll put the cluster near the center face [drag the cluster to the center face]. Oh, it's kind of similar."* P1: *"yeah."* P2: *"The mouth of centroid is a little all over the place. Pretty much the same, just like cleaned up."* P1: *"Yeah. I'd say it's in the middle."* P2: *"Yeah, it's in the middle, so it is the average."*

We also observed that participants identified the outlier of the cluster by comparing between clusters. For example, P6: *"(The faces of this cluster) are kind of similar. I mean the outline of the shapes of all faces. But this one (point to a face that is dissimilar with the center face), this face's eye size changes so much."* In this case, participants dragged the outlier face when compared with the other cluster, and found the face looks more similar to the other cluster.

Clustering Process

The pre-post learning gain questionnaire shows a weak learning effect for the procedure of k-means clustering ($Mean_{diff}$ = 0.25). Participants show diverse grouping strategies when manually clustering the faces (ML-M1 in SmileyCluster, Fig. 3(a)). Group 1 chose to randomly pick a face to start. This approach ended up with many groups. Group 2 and 3 adopted the global comparison strategy as used in the pair-wise comparison task. For example, P3 stacked the faces together because *"those faces are scary and skewed"*, and when group 2 finished sorting some faces, they concluded with their observations, *"(Those faces) are getting progressively older and scarier"*. Group 4, on the other hand, adopted the local comparison strategy. They first identified an obvious feature, eye size, as P7 mentioned, *"Big eyes are easiest, and then small eyes."* And when they found the blurriness of two big-eyed faces increased, they started to compare the mouth curve. Overall, Group 2 and 3 performed the clustering task faster than Group 4, which indicates that the social familiarity of overall facial expressions can help students to obtain a global understanding of all the features of a data point simultaneously.

Appropriate k cluster number

The pre-post questionnaire shows the highest learning gain in understanding the method of choosing the appropriate k number for clustering ($Mean_{diff}$ = 1.63). When asked to choose the appropriate k number, all groups chose the right answer for both pages, except that Group 4 made the wrong choice in the first page. Although the correct rates are similar before and after seeing the center face and aggregated face, the evidence used by students to support their decision shows a big difference. Before learning about center faces and combination faces, all 4 groups decided based on intra-cluster similarity. For example, P3: *"I don't know. I wanna say three."* P4: *"Cause 4 has weird."* P3: *"No. The other thing is I don't know how much these are different than these. These are all very similar... I said these should be here. But a lot of these are a lot similar too. I think two might give a better average than three. I am gonna say three and not anything else."*

After learning about center faces and combination faces, all 4 groups took into consideration inter-cluster separation to support their decision-making. For example, P6: *"so these are the center faces. They are way too dissimilar. Wait, 'dissimilar' between groups."* ... P5: *"I want to choose 3."* P6: *"Yeah, me too. These are the center faces right? Yeah they are very different. (Point to K=4) This is the only very different one but the other three are similar. ... (Point to K=2) And they are not focused."*

Learning Effects for Sense-making of Patterns

Basic cluster interpretation

Participants are able to interpret the clustering result by looking at the machine clustered result (ML-P1). Interestingly, some students showed surprise about machine clustering results in comparison with the manual clustering the students performed. For example, P2 commented, *"Are you sure this is what we built?"*, and P3 echoed *"no"*. Then participants compared the features of two clusters and interpreted the results given by the feature mapping cheat sheet. P2 [pointing at the screen]: *"Oh, cool. [Points to cluster 1], these faces are all kind of smiley."* P1: *"And then these faces [in cluster 2] are all like spooky."* P2: *"Yikes."* P1: *"Faces in cluster 1 are all have big eyes, meaning the seeds have big area."* P2: *"Yes, and they all have about the same compactness."*

Intra-cluster pattern and Inter-cluster pattern interpretation

Analysis of the written answers shows that all groups explained shared features and differentiating features reasonably, with sufficient association between the facial and seed features. The example below shows a typical episode, reflecting how they went through the facial features, and reached an agreement through discussion. P1: *"Greater eyebrow slant."* P2: "Greater mouth curve." [P1 is typing] P2: *"Eyebrow length. Pretty high. . . ?, I don't know."* P1: *"Actually average eyebrow length. Eyebrow height is lower. Actually, is it lower?"* P2: *"Pretty high."* P1: *"It is a little lower (compared to the other face)."* P2 agreed and referred to the cheat sheet to interpret the result. P2: *"(the above features aligned with) large area, greater perimeter. and greater compactness of seed..."*

Group 1 reported the common facial features of the combination faces as: big eye radius, greater eyebrow slant, greater mouth curve, avg eyebrow length, lower eyebrow height, narrower mouth, avg mouth height. They translated these to the seed data respectively as large area, greater perimeter, avg compactness, shorter kernel, narrower kernel, avg asymmetry, longer kernel groove. The four groups gave relatively similar answers by comparing the cohesion and separation of features and interpreting the result when applied to seed data, while group 3 tried to interpret the results a step further: "We can see that they have similar area, perimeter, compactness, and kernel length. The kernel width, asymmetry coefficient, and kernel groove length are all different".

With the glyph and global overlay design, participants can compare the size of the blurred areas to make decisions on pattern interpretation. While interpreting shared patterns within a cluster, participants understood how the blurred areas represented variances of different features. For example, by comparing the width of blurred areas around eyes or mouths, they decided the most significantly different or consistent features to be counted into the pattern. It shows participants' ability to recognize parameters with relatively small variation range and debate on whether it is significant enough to be selected as a shared feature or a differentiating feature.

While translating the facial features to seed features, deeper discussion and reasoning behind different seed patterns were rare. One major reason could be that no authentic scientific questions about the dataset were presented in this study. In the current system design, the questions about pattern interpretation are very simple and straightforward for high school students, and we only asked them to check the cheat sheet and map facial features they found back to seed features.

Scientific Inquiry Related Learning Behaviors

We investigated the scientific inquiry related behaviours during the whole learning process and identified that their hands-on exploratory learning generated opportunities for students to engage in scientific discoveries, demonstrated by scientific behaviours such as suggestion, justification, question answering, and showing uncertainty. Such scientific behaviours are rich in human clustering, deciding the appropriate k cluster number and pattern interpretation activities.

Constructing explanations

During manual clustering, all four groups discussed the strategies to take and provide explanations for their solutions. For example, P7 and P8 hesitated as to which face to start with. P7: "You can group them based on their eyes first (suggestion)*, because big eyes are the easiest to identify (*justification*)."* When P8 struggled to find another one with big eyes, P7: *"Look at this one (points to the face), and this one (points to another face)."* (suggestion) P8: *"But I am not sure."* (uncertainty) P7: *"Why?"* (question asking) P8: *"These are relatively small eyes, and you group them together."* (Justification)

Obtaining, evaluating and communicating information

When selecting the best k number, there was uncertainty deciding between k=3 and k=4. In group 2, P3: *"I don't know. I wanna say three (groups)."* (uncertainty and suggestion) P4: *"Cause 4 clusters look weird?"* (question asking) P3: *"No. The thing is I don't know how much these (faces) are different than those (faces). These (faces) look all very similar.. I think 3 (groups) might give a better average than four. I am gonna say three and not anything else."* (justification)

Engaging in argument from evidence

When selecting the best k number from comparing the dissimilar features in face-overlay, participants engaged in scientific talk with argument. P8: *"I think it may be this one [points to the option k=3]."* (suggestion) P7: *"I don't think so. I think it is this one [points to the option k=4]."* (argument and suggestion) P8: *"Because you can see, the eyebrows."* (justification) P7: *"What?"* P8: *"The eyebrows [of three faces] are almost a line. And [points to 4 groups], there are hundreds of lines of eyebrows (meaning the blurriness is high when stacked).* (argument) *And for the mouth, you can see, [3 groups] are almost together"* (justification) P8: *"Right?"* (uncertainty)

Analyzing and interpreting data

Participants engaged in interpreting the shared features and differentiating features of faces and referring back to the original seed dataset using the multidimensional feature mapping relation table provided by the system. P1: *"Does the "area" mean [points to the stacked face and points to the cheat sheet], big area of eyes? I think the right column talks about the face and the left column talks about the seed. You know what I mean?"* (uncertainty and question asking) P2: [pointed to "area" on the cheat sheet] *"so (area of the seed) matches eye* *radius."* (sharing findings) P1: *"And Greater eyebrow slant.."* P2: *"Means pretty big kernel width of the seed."*

When discussing eyebrow length, P2: *"Eyebrow length, pretty...I don't know."* (uncertainty) P1: *"It should be average eyebrow length and lower eyebrow height. Actually is it lower?"* (justification and uncertainty) P2: *"pretty high."* P1: *"(The left face) is a little lower."* P2: *"(The above features aligned with) large area, greater perimeter and greater compactness of seed."*

DISCUSSION

Overall, the results of the study show promising effect that the face-overlay design metaphor is effective in supporting the understanding of the fundamental ML concept - similarity comparison on multi-dimension feature space. Building based on face-overlay, the SmileyCluster system shows positive effectiveness in enhancing the understanding of the k-means clustering ML method, sense-making of pattern in both the face and seed contexts. We also observe that while engaged in ML activities, participants carried out meaningful scientific inquiry behaviors through discussing in-app questions, expressing uncertainty, asking questions, offering suggestions and justifications, and engaging in argument.

SmileyCluster in Supporting ML Learning

Participants increased their understanding of ML concepts and methods after using SmileyCluster to learn cluster analysis, evidenced by the increase of learning gains to different extents in similarity computation, centroid, clustering, and selecting the appropriate k cluster number. As the first concept of clustering analysis, participants are able to explore the mapping relationship between facial features and data values and apply the concept to brainstorm other objects that could map the multi-dimensional feature space, such as the glyph shaped trees and block-based objects as lego blocks.

Face-overlay had the edge over side-by-side comparison in supporting the pairwise similarity comparison. When learning about the similarity comparison using both side-by-side comparison and overlay metaphor, four group pairs achieved a 50% and 100% correct rate accordingly. Particularly, groups that used the face-overlay version for pair-wise visual comparison made quicker decisions and were more accurate. This seems to confirm the findings of Tufte[58] that overlay helped participants with less attention shifting and lower memory requirements than comparing data points side-by-side.

Global overlay supported the global comparison between multi-dimension data points. When learning the centroid, participants in group 3 identified the outliers by dragging the cluster close to the center face and comparing the similarity. Face-overlay also extended participants' understanding of intra-cluster cohesion and inter-cluster separation to evaluate the best k cluster numbers both in pair-wise and global overlay. Before learning about the center and combination of faces using overlay, four groups decided the choice of selecting k based on intra-cluster cohesion, which compares the similarity between data points within a group. P3 in group 4 showed initially high uncertainty in selecting the k number. By learning about comparing the dissimilarity of center faces using overlay methods, participants are able to facilitate their decision making by observing the blurriness of stacked faces. that overlay helped participants with less attention shifting and lower memory requirements than comparing data points side-by-side [58]. Participants demonstrated clear scientific reasoning and adjusted their previous false inference evaluation [39, 48] for selecting the k number after learning center and combination of faces, which is achieved during scientific discussion with peers. Such process supported the benefit of collaborative scientific knowledge construction that new knowledge construction occurs when individuals realize contractions, inconsistencies and limitations of one's perception during interaction with peers [42, 27].

Participants were able to interpret patterns of the clustering result, the intra-cluster, and inter-cluster. With the glyph design and global overlay design, participants compared the size of the blurred areas of faces for intra-cluster similarity and inter-cluster dissimilarity. Using a cheat sheet helps them to map the facial features back to the seed data. However, such methods also limit the depth of the scientific discussion, given the design might be too easy for high school students. Group 2 and 3 paused their discussion after inputting the answers, and limited their ability to interpret the findings further without designing for authentic and systematic scientific questions about the dataset and data analysis.

Interestingly, we identified two typical strategies of similarity comparison, the global comparison and local comparison that are used frequently in similarity comparison and manual clustering activities. The global similarity comparison strategy focused on comparing the overall facial expressions while the local similarity comparison focused on comparing the individual features of the faces. Group 2 adopted the global similarity comparison throughout, and relative to the other groups, it appears that the global comparison strategy helps the participants to make quicker decisions, as group 2's process of grouping was much faster and smoother than group 4 who adopted the local similarity comparison strategy in clustering. It confirmed our hypothesis that participants' social familiarity with expressions helped them to focus on the global view of the face and save time by not needing to shift attention.

Early Indicators of Scientific Discovery

The SmileyCluster enabled the groups to engage in scientific inquiry, interpreting the clustering result with the seed data, and evaluating the k-number. Participants collaboratively worked on the tasks, took turns to elicit questions related to understanding the in-system questions, showed uncertainty when hesitating between answers, offered suggestions and justifications to reason about the strategies. All these behaviours are early signs of engaging in scientific inquiry of asking questions, analyzing and interpreting data, arguing evidence and evaluating results.

Design Considerations for Future System

First, from the recordings, we noticed most information was provided as text, and it took users a long time to read and understand. Therefore, we need to change the delivery of information to a more interactive and engaging approach, such as using animation and split the tutorials to enable the learning tasks more smoothly, while keeping information received at the pace of the students. Second, our analysis showed participants' limited engagement and lack of in-depth scientific discussions in pattern interpretation. This might also result from lacking the systematic scientific inquiry design that aligns with science curriculum, where it could elicit the contextualized and authentic scientific inquiries assisted with data analysis, modeling, and interpretation. In the future, we will consider adding more game-based design components to our system and co-design the system with school teachers for authentic scientific inquiries in order to increase the interactivity of the system and trigger participants' deeper interest and engagement. Lastly, we identified the need to better measure the learning outcome. As shown in the clustering process there is no exact measure for the clustering performance. For example, we could embed a measurement mechanism in our system to directly calculate the similarity, cohesion, and separation, and offer feedback for participants of each learning activity.

LIMITATIONS AND FUTURE WORK

Our work was conducted in an informal learning context with a prototype application, which had limited student samples in a one-time study. Given the novelty of the study, we did not formulate a solid rubric that aligned with any standard to measure participants' learning gains. Those factors limited the scope of our study. With the next iteration, we will expand our study design to align with the current STEM curriculum by running a co-design workshop with K-12 school teachers. Additionally, our current SmileyCluster system only targeted k-means clustering analysis and we aim to integrate other clustering analysis or other ML algorithms into our system, and enable the system to generate a user-provided dynamic dataset. Furthermore, the face glyph data visualization method has the innate limitation of the total number of features it can support. Although it is sufficient to support the learning of basic concepts and methods of ML and support using ML to make sense of a simple dataset, alternative data visualization approaches are yet to be explored in order to accommodate high-dimensionality datasets. Lastly, due to the nature of the pre-college summer school, most participants in this study have prior knowledge in programming and AI. Therefore, recruiting students with less programming and AI experience is needed in future studies.

CONCLUSION

The driving research questions of this study are two-fold: (1) *how to design learning environments that support understanding the ML concepts and methods for young students with diverse mathematical and programming skills;* and (2) *can ML learning environments support scientific discovery in STEM contexts?* To address the first research question, we proposed the **face-overlay** conceptual metaphor with the hypothesis that it can reduce cognitive barriers for ***multi-dimension feature space***, ***similarity comparison***, and ***global understanding of data***. Based on the face-overlay metaphor, we created the programming-free **SmileyCluster** system to support learning entry-level ML concepts and methods related to k-means clustering. In addition, we explored the design and pedagogical opportunities to provide exploratory and authentic learning experiences for sense-making of patterns in the context of real-world phenomenon, and engage students in scientific inquiry behaviors such as question asking, explanation and argument.

Findings of this study provide rich evidence that (1) the face-overlay metaphor can effectively support similarity comparison and global understanding for high school students; (2) the SmileyCluster system can positively support learning of k-means clustering, which is centered around similarity comparison and global understanding. In addition, the participants also carried out meaningful scientific inquiry behaviors while interacting with the SmileyCluster system in pairs. The study extends our knowledge of how glyph-based data visualization and superposition-based comparative visualization can jointly scaffold the understanding of similarity comparison and global patterns involving multiple features within an individual data point and multiple data points within a group, both of which are demanding for young learners with limited mathematical and statistical backgrounds. The face-overlay metaphor can inform the design of future technologies that support learning of similarity-based ML methods such as k-nearest neighbor classification, information retrieval and anomaly detection. Furthermore, the study deepens our understanding of how the exploratory nature of hands-on manipulation and unsupervised machine learning can elicit scientific inquiry learning behaviors. As scientific inquiry plays a critical role in cultivating a scientific mindset for young learners, this study opens new opportunities for integration of ML literacy with existing K-12 STEM education frameworks.

Overall, this study sheds light on making ML literacy more accessible to young learners. This is achieved through a novel design approach to demystify ML out of its black-box, and initial exploration of ML as a crosscutting learning component to elicit data-driven scientific discovery and productive scientific inquiry in STEM education.

SELECTION AND PARTICIPATION OF CHILDREN

We recruited high school students through the pre-college summer program from the university in person. We sent the consent form to fully inform parents and youth prior to signing up. We explicitly explained the forms to participants for the study in person and answered their questions. The consent forms for participation and video recording were signed and obtained from youth prior to the study. All youth who volunteered were selected to participate in our study. We followed our approved Institutional Review Board protocol for this research study.

ACKNOWLEDGEMENT

This work was supported in part by the National Science Foundation(REU grant 1659250). We would like to thank our collaborator Dr. Michael Daley, Dr. Raffaella Borasi, and Dr. Dave Miller from the Warner School of Education, the director and program coordinator of the University of Rochester pre-college summer program, Zhaoxiong Ding for his contribution in the early project brainstorming, and young students who chose to participate in our study.

REFERENCES

[1] Murat Ay and Ozgur Kisi. 2014. Modelling of chemical oxygen demand by using ANNs, ANFIS and k-means clustering techniques. *Journal of Hydrology* 511 (2014), 279–289.

[2] Philip Bell, Leah Bricker, Carrie Tzou, Tiffany Lee, and Katie Van Horne. 2012. Exploring the science framework. *Science and Children* 50, 3 (2012), 11.

[3] Dani Ben-Zvi and Abraham Arcavi. 2001. Junior high school students' construction of global views of data and data representations. *Educational studies in mathematics* 45, 1-3 (2001), 35–65.

[4] Harald Burgsteiner, Martin Kandlhofer, and Gerald Steinbauer. 2016. Irobot: Teaching the basics of artificial intelligence in high schools. In *Thirtieth AAAI Conference on Artificial Intelligence*.

[5] Anthony Chan, Leyland F Pitt, and Deon Nel. 2014. Let's face it: Using Chernoff faces to Portray Social Media Brand Image. *Corporate Ownership & Control* 609 (2014).

[6] Niewczas Jerzy Kulczyck Piotr A. Kowalski Piotr Lukasik Szymon Charytanowicz, Małgorzata and Slawomir Zak. 2012. (2012). https://archive.ics.uci.edu/ml/datasets/seeds

[7] Herman Chernoff. 1973. The use of faces to represent points in k-dimensional space graphically. *Journal of the American statistical Association* 68, 342 (1973), 361–368.

[8] Michelene TH Chi and Ruth Wylie. 2014. The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational psychologist* 49, 4 (2014), 219–243.

[9] National Research Council. 2013. APPENDIX F: Science and Engineering Practices in the Next Generation Science Standards. *Next Generation Science Standards: For States, By States* (2013).

[10] John W Creswell and Cheryl N Poth. 2016. *Qualitative inquiry and research design: Choosing among five approaches*. Sage publications.

[11] MC Ferreira De Oliveira and Haim Levkowitz. 2003. From visual data exploration to visual data mining: A survey. *IEEE Transactions on Visualization and Computer Graphics* 9, 3 (2003), 378–394.

[12] Geen De Soete and Wilfried Do Corte. 1985. On the perceptual salience of features of Chernoff faces for representing multivariate data. *Applied psychological measurement* 9, 3 (1985), 275–280.

[13] Stefania Druga, Sarah T Vu, Eesh Likhith, and Tammy Qiu. 2019. Inclusive AI literacy for kids around the world. In *Proceedings of FabLearn 2019*. 104–111.

[14] Richard Dubes and Anil K. Jain. 1980. Clustering methodologies in exploratory data analysis. In *Advances in computers*. Vol. 19. Elsevier, 113–228.

[15] Wei Fan and Albert Bifet. 2013. Mining big data: current status, and forecast to the future. *ACM sIGKDD Explorations Newsletter* 14, 2 (2013), 1–5.

[16] Kay Firth-Butterfield. 22 May 2018. Generation AI: What happens when your child's friend is an AI toy that talks back? (22 May 2018). Retrieved 2019-03-19.

[17] Bernhard Flury and Hans Riedwyl. 1981. Graphical representation of multivariate data by means of asymmetrical faces. *J. Amer. Statist. Assoc.* 76, 376 (1981), 757–765.

[18] National Governors Association Center for Best Practices. 2010. C. o. C. S. S. O. Common Core Science Standards. National Governors Association Center for Best Practices, Council of Chief State School. (2010).

[19] Michael Friendly. 2008. A brief history of data visualization. In *Handbook of data visualization*. Springer, 15–56.

[20] Johannes Fuchs, Niklas Weiler, and Tobias Schreck. 2015. Leaf glyph visualizing multi-dimensional data with environmental cues. (2015).

[21] Barbara Gabbert, David W Johnson, and Roger T Johnson. 1986. Cooperative learning, group-to-individual transfer, process gain, and the acquisition of cognitive reasoning strategies. *The Journal of Psychology* 120, 3 (1986), 265–278.

[22] Hannie Gijlers and Ton de Jong. 2013. Using concept maps to facilitate collaborative simulation-based inquiry learning. *Journal of the learning sciences* 22, 3 (2013), 340–374.

[23] Michael Gleicher, Danielle Albers, Rick Walker, Ilir Jusufi, Charles D Hansen, and Jonathan C Roberts. 2011. Visual comparison for information visualization. *Information Visualization* 10, 4 (2011), 289–309.

[24] Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Franco Turini, Fosca Giannotti, and Dino Pedreschi. 2018. A survey of methods for explaining black box models. *ACM computing surveys (CSUR)* 51, 5 (2018), 1–42.

[25] James Hollan, Edwin Hutchins, and David Kirsh. 2000. Distributed cognition: toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction (TOCHI)* 7, 2 (2000), 174–196.

[26] Wang Hongwei and Liu Hui. 2013. Quantitatively plotting the human face for multivariate data visualisation illustrated by health assessments using laboratory parameters. *Computational and mathematical methods in medicine* 2013 (2013).

[27] Edwin Hutchins. 1991. The social organization of distributed cognition. (1991).

[28] Edwin Hutchins. 2006. The distributed cognition perspective on human interaction. *Roots of human sociality: Culture, cognition and interaction* 1 (2006), 375.

[29] Jamie Lee Jensen and Anton Lawson. 2011. Effects of collaborative group composition and inquiry instruction on reasoning gains and achievement in undergraduate biology. *CBE—Life Sciences Education* 10, 1 (2011), 64–73.

[30] David W Johnson and Roger T Johnson. 2009. Energizing learning: The instructional power of conflict. *Educational Researcher* 38, 1 (2009), 37–51.

[31] KM Kahn, Rani Megasari, Erna Piantari, and Enjun Junaeti. 2018. AI programming by children using snap! block programming in a developing country. (2018).

[32] Todd R Kelley and J Geoff Knowles. 2016. A conceptual framework for integrated STEM education. *International Journal of STEM Education* 3, 1 (2016), 11.

[33] John A Kupfer, Peng Gao, and Diansheng Guo. 2012. Regionalization of forest pattern metrics for the continental United States using contiguity constrained clustering and partitioning. *Ecological Informatics* 9 (2012), 11–18.

[34] Michael D Lee, Rachel E Reilly, and Marcus E Butavicius. 2003. An empirical evaluation of Chernoff faces, star glyphs, and spatial visualizations for binary data. In *Proceedings of the Asia-Pacific symposium on Information visualisation-Volume 24*. Australian Computer Society, Inc., 1–10.

[35] Klaus Libertus, Rebecca J Landa, and Joshua L Haworth. 2017. Development of attention to faces during the first 3 years: Influences of stimulus type. *Frontiers in psychology* 8 (2017), 1976.

[36] Andrew Manches and Claire O'malley. 2012. Tangibles for learning: a representational analysis of physical manipulation. *Personal and Ubiquitous Computing* 16, 4 (2012), 405–419.

[37] Andrew Manches, Claire O'Malley, and Steve Benford. 2010. The role of physical representations in solving number problems: A comparison of young children's use of physical and virtual materials. *Computers & Education* 54, 3 (2010), 622–640.

[38] Paul Marshall. 2007. Do tangible interfaces enhance learning?. In *Proceedings of the 1st international conference on Tangible and embedded interaction*. 163–170.

[39] Camillia Matuk and Marcia C Linn. 2018. Why and how do middle school students exchange ideas during science inquiry? *International Journal of Computer-Supported Collaborative Learning* 13, 3 (2018), 263–299.

[40] Anthony McCosker and Rowan Wilken. 2014. Rethinking 'big data'as visual knowledge: the sublime and the diagrammatic in data visualisation. *Visual Studies* 29, 2 (2014), 155–164.

[41] Kelly S Mix. 2010. Spatial tools for mathematical thought. *Space and language* (2010), 41–66.

[42] Naomi Miyake. 1986. Constructive interaction and the iterative process of understanding. *Cognitive science* 10, 2 (1986), 151–177.

[43] José Jesús Reyes Nuñez, Anita Rohonczi, Cristina E Juliarena de Moretti, Ana María Garra, Carmen Alicia Rey, María V Alves de Castro, Anabella S Dibiase, Teresa A Saint Pierre, and Mariana A Campos. 2011. Updating research on Chernoff faces for school cartography. In *Advances in Cartography and GIScience. Volume 2*. Springer, 3–20.

[44] James Patten and Hiroshi Ishii. 2000. A comparison of spatial organization strategies in graphical and tangible user interfaces. In *Proceedings of DARE 2000 on Designing augmented reality environments*. 41–50.

[45] Federico Perini. 2013. High-dimensional, unsupervised cell clustering for computationally efficient engine simulations with detailed combustion chemistry. *Fuel* 106 (2013), 344–356.

[46] Jean. Piaget. 1936. *Origins of intelligence in the child*. London: Routledge Kegan Paul.

[47] Sara Price, Jennifer Sheridan, Taciana Pontual Falcao, and George Roussos. 2008. Towards a framework for investigating tangible environments for learning. *International Journal of Arts and Technology* 1, 3/4 (2008), 351–368.

[48] Rod D Roscoe and Michelene TH Chi. 2008. Tutor learning: The role of explaining and responding to questions. *Instructional Science* 36, 4 (2008), 321–350.

[49] Victor Sampson and Douglas Clark. 2009. The impact of collaboration on the outcomes of scientific argumentation. *Science education* 93, 3 (2009), 448–484.

[50] Mike Scaife and Yvonne Rogers. 1996. External cognition: how do graphical representations work? *International journal of human-computer studies* 45, 2 (1996), 185–213.

[51] Tanmay Sinha, Zhen Bai, and Justine Cassell. 2017. Curious minds wonder alike: studying multimodal behavioral dynamics to design social scaffolding of curiosity. In *European conference on technology enhanced learning*. Springer, 270–285.

[52] Nicole Sintov, Debarun Kar, Thanh Nguyen, Fei Fang, Kevin Hoffman, Arnaud Lyet, and Milind Tambe. 2016. From the lab to the classroom and beyond: extending a game-based research platform for teaching AI to diverse audiences. In *Thirtieth AAAI Conference on Artificial Intelligence*.

[53] Ruixia Song, Zhaoxia Zhao, and Meifang Ou. 2009. A novel clustering method for chernoff faces based on V-system. In *2009 International Conference on Information and Automation*. IEEE, 1556–1561.

[54] Ruixia Song, Zhaoxia Zhao, and Xiaochun Wang. 2010. An application of the V-system to the clustering of Chernoff faces. *Computers & Graphics* 34, 5 (2010), 529–536.

[55] Robert E Stake. 1995. *The art of case study research*. sage.

[56] Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. 2013. Data mining cluster analysis: basic concepts and algorithms. *Introduction to data mining* (2013), 487–533.

[57] David Touretzky, Christina Gardner-McCune, Fred Martin, and Deborah Seehorn. 2019. Envisioning AI for K-12: What Should Every Child Know about AI?. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 9795–9799.

[58] Edward. R. Tufte. 2006. *Beautiful evidence (Vol. 1)*. CT: Graphics Press, Cheshire.

[59] Mojtaba Vaismoradi, Jacqueline Jones, Hannele Turunen, and Sherrill Snelgrove. 2016. Theme development in qualitative content analysis and thematic analysis. (2016).

[60] Nathan VanHoudnos, William Casey, E Wright, D French, S Moon, B Lindauer, E Kanal, P Jansen, and J Carbonell. 2017. This malware looks familiar: Laymen identify malware run-time similarity with chernoff faces and stick figures. In *proceedings of the 10th EAI InternationalConference on Bio-inspired Information and Communications Technologies*.

[61] Jennifer Wang, Hai Hong, Jason Ravitz, and Sepehr Hejazi Moghadam. 2016. Landscape of K-12 computer science education in the US: Perceptions, access, and barriers. In *Proceedings of the 47th ACM Technical Symposium on Computing Science Education*. 645–650.

[62] Matthew O Ward. 2002. A taxonomy of glyph placement strategies for multidimensional data visualization. *Information Visualization* 1, 3-4 (2002), 194–210.

[63] Matthew O Ward, Georges Grinstein, and Daniel Keim. 2015. *Interactive data visualization: foundations, techniques, and applications*. AK Peters/CRC Press.

[64] David Weintrop, Elham Beheshti, Michael Horn, Kai Orton, Kemi Jona, Laura Trouille, and Uri Wilensky. 2016. Defining computational thinking for mathematics and science classrooms. *Journal of Science Education and Technology* 25, 1 (2016), 127–147.

[65] Ian H Witten and Eibe Frank. 2002. Data mining: practical machine learning tools and techniques with Java implementations. *Acm Sigmod Record* 31, 1 (2002), 76–77.

[66] Jo Wood, Jason Dykes, Aidan Slingsby, and Keith Clarke. 2007. Interactive visual exploration of a large spatio-temporal dataset: Reflections on a geovisualization mashup. *IEEE transactions on visualization and computer graphics* 13, 6 (2007), 1176–1183.

[67] Abigail Zimmermann-Niefield, Makenna Turner, Bridget Murphy, Shaun K Kane, and R Benjamin Shapiro. 2019. Youth Learning Machine Learning through Building Models of Athletic Moves. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children*. 121–132.