2023 Annual Conference & Exposition
Baltimore Convention Center, MD | June 25 - 28, 2023
The Harbor of Engineering Education for 130 Years
ASEE
Paper ID #40242

Data Science (Dataying) for Early Childhood

Dr. Safia A. Malallah, Kansas State University

Safia Malallah is a postdoc in the computer science department at Kansas State University working with Vision and Data science projects. She has ten years of experience as a computer analyst and graphic designer. Besides, she's passionate about developing curriculums for teaching coding, data science, AI, and engineering to young children by modeling playground environments. She tries to expand her experience by facilitating and volunteering for many STEM workshops.

Lior Shamir, Kansas State University

Associate professor of computer science at Kansas State University.

Dr. William Henry Hsu, Kansas State University

William H. Hsu is an associate professor of Computing and Information Sciences at Kansas State University. He received a B.S. in Mathematical Sciences and Computer Science and an M.S.Eng. in Computer Science from Johns Hopkins University in 1993, and a Ph

Joshua Levi Weese, Kansas State University

Dr. Josh Weese is a Teaching Assistant Professor at Kansas State University in the department of Computer Science. Dr. Weese joined K-State as faculty in the Fall of 2017. He has expertise in data science, software engineering, web technologies, computer science education research, and primary and secondary outreach programs. Dr. Weese has been a highly active member in advocating for computer science education in Kansas including PK-12 model standards in 2019 with an implementation guide the following year. Work on CS teacher endorsement standards are also being developed. Dr. Weese has developed, organized and led activities for several outreach programs for K-12 impacting well more than 4,000 students.

Mr. Salah Alfailakawi, Kansas State University

Salah Alfailakawi is a PhD student in Educational Technology (ET) Graduate Programs at Kansas State University's College of Education. His areas of interest include social/cultural issues in ET, the impact of ET on learners and teachers, as well as pract

Data Science (Datayng) for Early Childhood

Safia A. Malallah, Kansas State University, safia@ksu.edu
Lior Shamir, Kansas State University, lshamir@ksu.edu
William Henry Hsu, Kansas State University, bhsu@ksu.edu
Joshua Levi Weese, Kansas State University, weeser@ksu.edu
Salah Alfailakawi, Kansas State University, Bosloh@ksu.edu

Abstract: Although evidence suggests that children as young as four years old can develop coding and engineering projects based on data science concepts, data science is often overlooked in early childhood research, and limited resources existed slow its inclusion into this field of study. This paper proposes the Datayng framework to teach data science concepts to young children ages 4–7 years old. The framework development included identifying K–12 data science elements and then validating element suitability for young students. Six cycled steps were identified: identifying a problem, questioning, imagining and planning, collecting, analyzing, and story sharing. This paper also presents examples of data decision problems and demonstrates use of a proposed Insight-Detective method with a plan worksheet for Datayng.

Introduction

The expected growth of data science careers worldwide over the next ten years means that students of all ages, including children in early childhood education, must be consistently exposed to data science concepts to meet future industry requirements [1, 2]. Students who learn data science at a young age are better equipped to implement the concepts at later stages where they will have more chances to practice and develop their skills [3]. However, current data science research for early childhood is very limited, and although previous data science frameworks for K–12 education have claimed that the content is suitable for kindergarteners, application has proven that, in reality, the content is more appropriate for students in grade 4 and beyond [4].

Therefore, this paper proposes a data science framework suitable for the developmental stages of young children ages 4–7 years old. The Datayng framework, an original term intended to create a buzzword for early childhood data science, utilizes nine stages by extracting and validating the framework elements and element relationships. This framework also uses six stages to introduce the foundation of Datayng thinking for children who have never been exposed to data-driven solutions and familiarize children with the creation of basic solutions from numbers and insights. Children can then use the framework on an intermediate level to escalate analysis, identify decision-making problems, and create intermediate solutions. Eventually, children can advance to using the K–12 data science frameworks, which will also expose them to future careers in data. In this work, Datayng is defined as the process of solving a simple data decision problem by collecting a small data sample and then analyzing it for insights to make a decision in a way that is suitable for early childhood education.

Background

Early Childhood Development Milestones for Datayng

Piaget's theory of cognitive development asserts that everyone advances through four main stages of development: (1) sensorimotor, (2) preoperational, (3) concrete operational, and (4) formal operations. Children ages 0–7 years old are in the sensorimotor and preoperational stages, meaning their logical abilities are limited. Therefore, the introduction of data science is most beneficial for preschool children ages 3–5 years old who are in the preoperational stage of cognitive development. At this stage, children show improvements in language and symbolic thinking, but their logical thinking is not developed, and they rely on perceptions, which can be easily misguided by appearances. With proper training, children in the preoperational stage can identify patterns with symbols and use symbols or simple words to outline sequences and algorithm designs. Research indicates that children as young as 3 years old can engage in unplugged computational thinking activities, and children ages 4–6 years old can build and program basic robotics and coding projects, indicating that 3–4-year-olds could begin to use the Dataying framework [5, 6].

Method

Design

This research combined the conceptual framework approach from methods in previous research [7], [8], [9] to develop the Dataying framework. Nine stages were developed and used in order to create the blueprint for the study illustrated in Figure 1. A conceptual framework is a structured approach to organize and understand complex ideas and then utilize a visual representation to illustrate the topic elements and their relationships [7]. 4 Literature reviews implemented using the research method that focus on using different angles to explore existing evidence to formulate understanding of a topic.

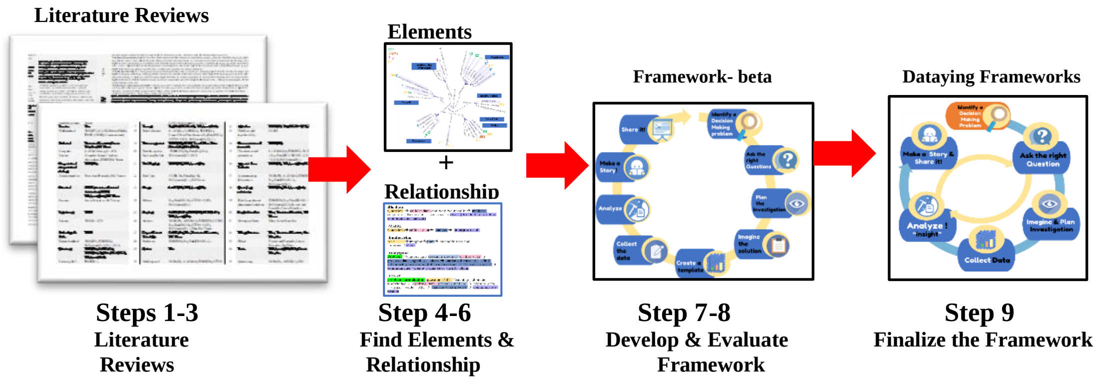


Figure 1. Design Steps for the Dataying Framework

The following nine stages were followed to develop the framework:

1. **Extracting Elements I:** In this first stage, literature review 1 was used to identify data science elements from preschool professional organizations. The purpose of this step was to understand the current status of data science in early childhood. Data saturation of the results determined the need for further reviews from different perspectives.
2. **Extracting Elements II:** This second stage identified data science elements from definitions in literature review 2. The primary goal was to develop a common general understanding between multidisciplinary fields. The identified disciplines were the search keywords for the next step.
3. **Extracting Elements III:** The goal of this step was to identify K–12 data science elements from frameworks and standards from other professional institutes (literature review 3).
4. **Filtering Elements:** This stage included clustering, grouping, and eliminating redundant elements.
5. **Validating Elements**: This fifth stage utilized literature review 4 to identify evidence of elements suitable for early childhood and then linked the evidence to the elements.
6. **Missing Elements:** This stage identified new elements from literature review 3 that could enhance children’s utilization of a framework.

7. **Connection of the Elements:** This stage identified relationships between the elements and their coherent order.
8. **Validate the Framework:** The framework was evaluated using feedback from early 2 childhood teachers by interview them.
9. **Developing the Framework:** putting together all the elements

Keywords, Database, and Criteria

This study used keywords to investigate and search the literature reviews. The keywords for literature review 1 were “Standards||Frameworks” + “preschooler||Early childhood||young children” + “Data Science.” The keywords for literature review 2 were “What is Data Science||Definition||History+ Data Science” + “Precollege||K-12 + “Data Science,” while the keywords for literature review 3 were “Standards||Frameworks” + “Precollege||k-12” + “Data Science” + “nothing|| Mathematics|| Statistics|| Computer science|| Engineering|| Technology ||Science|| Business.” The keywords for literature review 4, which included elements and disciplines from the three previous reviews, were “Science|| Math|| Engineering|| Technology|| Computational thinking|| STEM” + “preschooler||Early childhood||young children” + “Standards||Frameworks.” The search engines used were Google, ScienceDirect, and IEEE Xplore, and the source types were conference papers and proceedings, blogs, government and official websites, and scholarly journals. The results were filtered by isolating abstracts and titles that did not match searching criteria, in addition surfing website pages and search boxes for the searching criteria. A second round of filtering utilized references from papers that addressed precollege-aged students since the focus of this research was early childhood. Any study that did not include elements of data science was excluded.

Results

Table 1 presents the merged results of literature review 1 and 3, including the data science elements and their resources. Table 2 lists the elements and their resources before the elimination steps of framework development, and Figure 2 shows the elements after elimination and clusters; elements with evidence are highlighted in green.

Literature Review 1 – Data Science for Preschoolers

The search using IEEE Xplore revealed 86 results, and the ScienceDirect search yielded 65 results, with all excluded expected one, which indicated the need for further investigation. The paper “Data Science K-10 Big Ideas” provided a comprehensive overview of the fundamental skills students should learn to become proficient in data science [10]. The paper also outlined four key concepts that should be taught in data science curriculum for kindergarten through 10th grade, including topics such as data collection and representation, data analysis and interpretation, and ethical considerations in the use of data. The included paper was developed by Youcubed, a nonprofit organization that provides mathematics education resources for teachers, students, and parents. “Data Science K-10 Big Ideas” was constructed as a contribution from statistics and computer science professional organizations.

Literature Review 2 – Data Science Elements - Exploring Definition

The term *data science* was first used in the 1960s by statistician Peter Naur in his research on the use of computers for scientific data processing. The term gained widespread use in the late 1990s and early 2000s as the amount of available data began to grow rapidly and new technologies made it possible to store, process, and analyze this data at scale [11]. As data science has become a multidisciplinary field, the breadth and depth of the roles of data science underscore the complexity and subsequent disagreement regarding the definition of the term. Data science has no official definition [11]. Cassie Kozyrkov, the Chief Decision Scientist at Google, defines data science as the use of data to make informed decisions, with a focus on technical and business skills to solve real-world problems. She also highlights the importance of understanding the limitations and biases in data, as well as the ethics of data-driven decision-making and size of the data set [12]. Dhar, in his book *Data Science for Business,* defines data science as a field that uses data, algorithms, and systems to extract knowledge and insights from data in various forms. He explains that data science is not just about analyzing data, but it is also about solving problems and making decisions based on the insights gained from data [13]. Provost and Fawcett's definition of data science highlights the interdisciplinary nature of the field and the importance of combining technical and domain expertise to extract insights and knowledge from data. They emphasize the use of data science to make data-driven decisions and the importance of understanding the ethics and limitations of data-driven decision making [13]. Similarly, Hal Varian, the Chief Economist at Google, defines data science as the process of extracting insights and knowledge from data. He considers data science to be an interdisciplinary field that blends computer science, statistics, and domain knowledge. Varian emphasizes the importance of having technical and business skills to effectively analyze data and extract insights [14]. Comparatively, Yong Cui proposed a formula for a simple definition of data science, suggesting that it can be used as easy-to-hard strategy formula: Data + Tools -> Implications. He asserts that data is at the core of data science and that the essential tools for data science include hardware, software, and techniques, such as machine-learning algorithms. In the formula, *Implications* are the outcomes of data science that can be viewed through two lenses: temporally (insights into the past, present, and future) and spatially (influences in various fields and daily life). The formula serves as a starting point for people with no background in data science [15].

Literature Review 3 – Data Science Framework and Standards for K–12

Diverse communities, professional organizations, and experts have contributed to the published literature that proposes pathways and elements to promote data science in K–12 education. These contributions have come from science, statistical, math, and computer science fields. A primary contribution from the science field has been the Next Generation Science Standards (NGSS), which outline eight essential, grade-level practices of science and engineering for all students in kindergarten through fifth grade in the United States [16]. In addition, a recent consensus report from the National Academies of Sciences, Engineering, and Medicine (NASEM) proposed nine main foundations to develop data science skills, or data acumen, the ability to make good judgments about the use of data to support problem solutions [17].

From the statistical field, the American Statistical Association is committed to enhancing data science through statistics education to foster statistical and data science literacy at all levels. The Association published a report, "Guidelines for Assessment and Instruction in Statistics Education Report II (GAISE II)," that proposed a data science framework with four essential concepts and 22 examples of framework application and assessment for three progressively conceptual structure levels (A, B, and C) [18]. Similarly, in their paper "Investigating Data Like a Data Scientist: Key Practices and Processes," Hollylynne S. Lee et el. developed a framework using the work of statistics educators and researchers to investigate how data science practices can inform work in K–12 education. Their framework builds fundamental practices and processes from data science [19].

The math field has contributed to data science research via the Common Core State Standards Initiative (CCSSI), which is a joint project to develop common K–12 reading and math standards designed to prepare students for college and careers. The CCSSI includes a data science section for elementary students that focuses on data collection, data type, function, analysis type, and sample [20]. Similarly, the Launch Years Data Science Course Framework provides broad guidelines for student learning outcomes, organizing topics thematically instead of focusing on specific skills [21].

Table 1. Literature Review Elements and Resources

Src	Elements
Big Data	***BIG1 Formulate statistical investigative questions: Big1.K1.1** Develop curiosity through noticing and wondering about data rich situations. **Big1.K1.2** Generate ideas and ask questions. The teacher helps refine, direct and create statistical investigative questions. **Big1.K1-2.1** Develop curiosity through noticing and wondering about data rich situations. **Big1.K1-2.2** The teacher helps refine, direct and create statistical investigative questions. * **BIG2 Collect/consider data: Big2.K1.1** Consider: What is data? Understand that people collect data to answer questions and that data can vary (eg objects have different colors or sizes). **Big2.K1.2** Develop strategies to collect and organize data – eg. sort collections of objects into categories that they have chosen. **Big2.K1-2.1** Learn about what counts as data and understand that people collect data to answer questions, and that data can vary (eg objects have different colors or sizes). **Big2.K1-2.2** Work with categorical and numerical (whole number) data. **Big2.K1-2.3** Consider and decide: What data will answer my question? **Big2.K1-2.4** Collect survey data (eg. favorite pets) or use data given by teacher (eg. ladybug data cards). * **BIG3 Analyze data: Big3.K1.1** Develop ways to represent data eg with tally marks, or as pictures or a drawing. **Big3.K1.2** Notice, describe and analyze patterns **Big3.K1.3** Recognize variability by noticing eg different sizes across a collection of items like buttons or blocks. **Big3.K1-2.1** Develop ways to represent data as tally marks, drawings, or digitally on a tablet or laptop. **Big3.K1-2.2** Students use data visualizations to look for patterns that allow them to make predictions. **Big3.K1-2.3** Students notice the likelihood of various outcomes, and variation across them. **Big3.K1-2.4** Explore mode — thinking conceptually about the data point(s) that happen the most. * **BIG4 Interpret and communicate: Big4.K1.1** Decide key results to summarize from an investigation and answer initial questions. **Big4.K1.2** Communicate results eg with a data report, a poster, a video. **Big4.K1.3** Decide key results to summarize from an investigation and answer initial questions. **Big4.K1-2.1** Decide key results to report that answer students' initial questions. **Big4.K1-2.2** Make a data report: poster, video, news article, etc. **Big4.K1-2.3** Make predictions using the terms: "likely, unlikely, certain, and impossible".
Launch Years Data Science	(1)Use collaboration and communication, (2) Utilize resources to overcome obstacles, (3) Recognize and improve individual behaviors., (4) Understand the role of data in the world, (5) Ask data-based questions, (6) Collect and manage data, (7) Explore data to make sense of and represent the story that the data are telling, (8) Analyze data to create data-based arguments and to reach data-based conclusions, (9) Understand limitations of data sources and data-based conclusions, (10) Use data to make predictions, (11) Use data to inform decision making, (12)Problem Solving, (13) Context and Interdisciplinary Connections, (14)Communication, Technology,
Hollylynne	***O1 FRAMETHEPROBLEM:** O1.1, consider real-world phenomena & broader issues related a problem. O1.2 Pose investigative questions, O1.3 anticipate potential data and strategies ***O2 CONSIDERANDGATHERDATA:** O2.1 Understand possible attributes, measurements, and data collection, O2.2 Evaluate and use appropriate design and techniques to collect or source data, O2.3, consider sample size, access, storage, and trustworthiness of data. * **Q3 Process Data**: O3.1 Organize, Structure, clean and transform data in efficient and useful ways. Consider additional data cases or attributes. * **O4 Explore & Visualize Data:** O4.1, Construct meaningful visualizations, statistic or dynamic, O4.2 Compute meaningful statistical measures, O4.3 Explore and analyze data for potential relationships or patterns that address the problem. * **O5 Consider Models:** O5.1 Analyze & Identify models that address the problem, O5.2 Consider assumption and context of the models O5.3 Recognize possible limitations * **O6 Communicate & Propose Action:** O6.1 Craft a data story to convey insight to stakeholder audiences, O6.2 justify claims with evidence from data and propose possible action, O6.3 Address uncertainty and potential bias in the analysis
K-12 Computer Science Framework	**(S1) Data Collection:** Everyday digital devices collect and display data over time. The collection and use of data about individuals and the world around them is a routine part of life and influences how people live. **(S2) Storage:** Computers store data that can be retrieved later. Identical copies of data can be made and stored in multiple locations for a variety of reasons, such as to protect against loss. **(S3) Visualization and Transformation**: Data can be displayed for communication in many ways. People use computers to transform data into new forms, such as graphs and charts. **(S4) Inference and Models:** Data can be used to make inferences or predictions about the world. Inferences, statements about something that cannot be readily observed, are often based on observed data. Predictions, statements about future events, are based on patterns in data and can be made by looking at data visualizations, such as charts and graphs.
IDSSP - International Data Science in Schools Project	**D1 Problem elicitation and information D1.1** Listening skills **D1.2** Questioning Skills **D1.3** Developing domain knowledge **D2** Getting the Data **D2.1** Data Management **D2.2** data Harvesting and wrangling/munging **D2.3** Sampling **D2.4** Experimental design **D3** Exploring the data **D3.1** Exploratory tools **D3.2** Visualization **D3.3** Algorithms **D3.4** Coding **D3.5** Data summaries **D4** Analyzing the data **D4.1** visualization **D4.2** developing and testing models **D4.3** algorithms **D4.4** coding **D4.5** Identifying aberrant phenomena **D4.6** Prediction **D4.7** quantifying uncertainty **D5** Communication the results **D5.1** Presentation Skills **D5.2** Plain Language Skills **D5.3** Presentation graphics **D5.4** Client focus
DREMEME - Development and Research in Early Math Education	***Data in the Preschool Classroom:*** *DR.1 Identifying Variation, Classifying, and Sorting DR.2 Understanding the Question DR.3 Use questioning to help children think about what they want to find out. DR.4 Gathering, Representing, and Interpreting the Data* ***Important Ideas about Data*** *DR.5 Classifying and Sorting DR.6 Comparing DR.7 Counting DR.8 Measuring DR.9 Representation of data DR.10 Unit size* ***Solving Problems and Answering Questions with Data*** *DR.11 Understand the question DR.12 Form a hypothesis and collect appropriate data DR.13Test the hypothesis DR.14 Analyze the data DR.15 Sometimes analysis presents the need for more data collection* ***Development of Children's Thinking on Measurement & Data*** *DR.16 A Thinking Story about Seriation and Measurement (question as a story)*
COR Advantage - HighScope	***APPROACHES TO LEARNING:** A. Initiative and planning, B. Problem solving with materials, C. Reflection ***SOCIAL AND EMOTIONAL DEVELOPMENT:** D. Emotions, E. Building relationships with adults F. Building relationships with other children, G. Community, H. Conflict resolution ***PHYSICAL DEVELOPMENT AND HEALTH:** I. Gross-motor skills, J. Fine-motor skills, K. Personal care and healthy behavior ***LANGUAGE, LITERACY, AND COMMUNICATION:** L. Speaking, M. Listening and comprehension, N. Phonological awareness O. Alphabetic knowledge, P. Reading, Q. Book enjoyment and knowledge, R. Writing ***MATHEMATICS:** S. Number and counting, T. Geometry: Shapes and spatial awareness, U. Measurement V. Patterns, W. Data analysis ***CREATIVE ARTS:** X. Art, Y. Music, Z. Movement, AA. Pretend play ***SCIENCE AND TECHNOLOGY:** BB. Observing and classifying, CC. Experimenting, predicting and drawing conclusions, DD. Natural and physical world EE. Tools and technology ***SOCIAL STUDIES:** FF. Knowledge of self and others, GG.Geography, HH. History * **ENGLISH LANGUAGE LEARNING (ELL):** II. Listening to and understanding English, JJ. Speaking English
CCSSI - Common Core State Standards Initiative	1 data collection, 2 data type, 3 function, 4 analysis type, 5 sample
NGSS - Next Generation Science Standards	(P1) Asking Questions and Defining Problems (P2) Developing and Using Models (P3) Planning and Carrying Out Investigations (P4) Analyzing and Interpreting Data (P5) Using Mathematics and Computational Thinking (P6) Constructing Explanations and Designing Solutions (P7) Engaging in Argument from Evidence (P8) Obtaining, Evaluating, and Communicating Information
NASEM - National Academies of Sciences, Engineering, and Medicine	(1) Mathematical foundations, (2) Computational foundations, (3) Statistical foundations, (4) Data management and curation, (5) Data description and visualization, (6) Data modeling and assessment, (7) Workflow and reproducibility, (8) Communication and teamwork, (9) Domain-specific considerations, and (10) Ethical problem solving
K-12 Computer Science Framework for preschooler	Patterns, Problem solving, Representation and sequencing.
STEM + C	C1 Abstraction, C2 Algorithms, C3 Automation, C4 Data Collection, C5 Data Analysis and C6 Data representation, C7 Debugging/Troubleshooting C8 Pattern Recognition, C9 Problem Decomposition, C10 Parallelization, C11 simulation
GAISE - American Statistical Association	**G1** Formulate Statistical Investigation Question **G2** Collect/ Consider the Data **G3** Analyze the Data **G4** Interpret the results
Engineering cycle - Marina Bers	1 Ask, 2 Imagine, 3 Plan, 4 Create, 5 Test and 6 improve and 7 finally share
Alabama Early childhood Standard	*M.P.3.1 Match, sort, place in a series, and regroup objects according to attributes (color, shape, size, etc.).* *M.P.3.2 Describe, duplicate, and extend simple patterns using a variety of materials or objects* *M.P.3.3 Recognize and identify patterns in the environment* *M.P.5.1 Use math vocabulary to compare sets of objects with terms such as more, less, equal to, greater than, fewer.* *M.P.5.2 Classify objects using more than one attribute* *M.P.5.3 Sort and classify objects using self-selected criteria* *M.P.5.4 Develop ability to collect, describe, and record information through drawings, maps, charts and graphs.* *S.P.1.1 Use senses to gather information, classify objects, observe processes, and describe materials* *S.P.1.3 Record observations using simple visual tools such as drawings, graphs, charts, logos.* *S.P.1.4 Describe simple cause and effect relationships.* *T.P.1.1 Demonstrate basic knowledge of computer skills.* *T.P.1.3 Demonstrate knowledge of the use of technology as a communication system of the world.*

The primary contribution of the computer science field has been the K–12 Computer Science framework, which includes four sub-concepts that outline data and analysis [22]. The concept outcomes require students to think about accuracy, quality, and quantity of data, as well as how data visualization can influence conclusions. The data science community has also shared a data science disciplinary diagram that suggests data science fields [23]. In addition, the International Data Science in Schools Project (IDSSP) promotes the integration of data science education into K–12 schools globally, specifically a "discrete" data science education that addresses data science through the lens of other subjects [24].

Table 2. Elements and Resources Before Elimination Steps

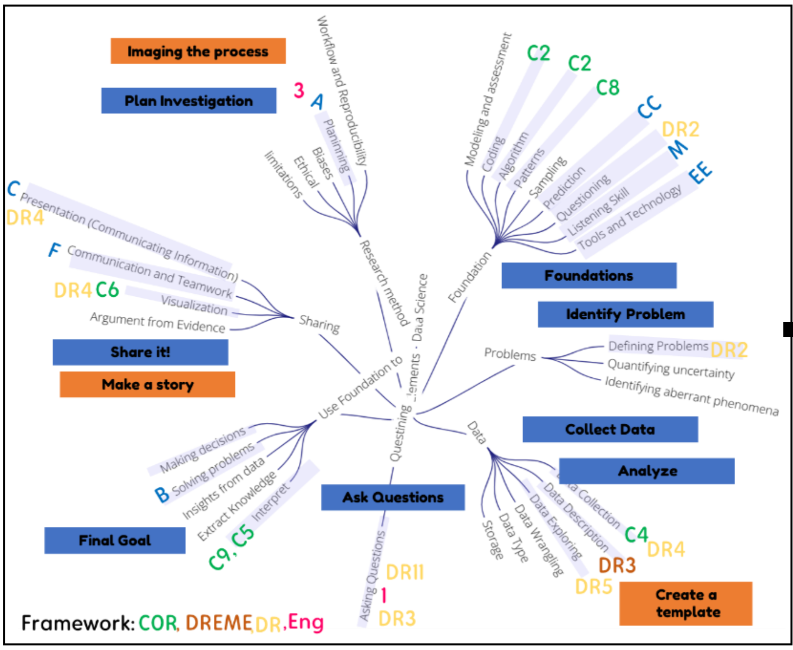
#	ELEMENTS	EVIDENCE	#	ELEMENTS	EVIDENCE	#	ELEMENTS	EVIDENCE
1	Interdisciplinary /multidisciplinary	Dhar, Provost and Fawcett's, Hal Varian	16	Defining Problems	NGSS(P1), IDSSP(D1, D1.3), STEM + C(C9), Hollylynne(O1.1, Q2.1)	31	Coding	IDSSP(D3, D4)
2	Business	Dhar	17	Planning	Big Data(BIG2.k.1.2), Hollylynne(O1.3)	32	Algorithm	IDSSP(D3, D4), STEM + C(C2), Dhar
3	Mathematical	NGSS(P5), k-12CS(National Math), IDSSP, CCSSI(k-5 measurement)	18	Data Collection	k-12CS(S1), GAISE(G2), IDSSP(D2), Launch Years Data Science, Big Data(BIG2), STEM + C(C4), Hollylynne(O2, Q2.1)	33	Workflow and reproducibility	CCSSI
4	Technical	Provost and Fawcett's, Cassie Kozyrkov, Hal Varian	19	Data management	IDSSP(D2), CCSSI, Launch Years Data Science, Big Data(BIG2),	34	Communication and teamwork	IDSSP(D5)
5	Computer Science	k-12CS(Code.org, k-12CS, Computer Science Teachers Association), IDSSP, Hal Varian	20	Data description , Exploring the data	IDSSP(D3), Big Data(Big3-k-2), Hollylynne(Q2.3, Q4.3)	35	Visualization and presentation	k-12CS(S3,S4), IDSSP(D3, D4, D5.3), Big Data(BIG3, BIG4-K1.2), STEM + C(C6), Hollylynne(Q4.1)
6	Computational /computational thinking	NGSS(P5)	21	Data Wragling	IDSSP(D2), Hollylynne (Q3)	36	Argument from Evidence	NGSS(P7), Hollylynne(Q3.1, O6.2)
7	domain expertise	Provost and Fawcett's, Hal Varian	22	Data Type	CCSSI, Big Data(Big2-K), Hollylynne(Q2.1), Hal Varian	37	Communicating Information	NGSS(P8), k-12CS(S3), IDSSP(D5.2), Big Data(BIG4.K1.2), Hollylynne(Q6.1)
8	Statistical	GAISE (American Statistical Association), IDSSP, Hollylynne(Q4.1), Hal Varian	23	Storage	k-12CS(S2), Hollylynne(Q2.3)	38	Quantifying uncertainty	IDSSP(D4), Hollylynne(Q4.3)
9	Business	Cassie Kozyrkov, Hal Varian	24	Patterns	Big Data(BIG3-k), STEM + C(C8), Hollylynne(Q4.3)	39	Identifying aberrant phenomena/limitations	IDSSP(D4), Big Data(BIG4.K.1.1), Hollylynne(Q5.3), Provost and Fawcett's, Cassie Kozyrkov
10	Engineering	NGSS	25	Sampling	GAISE (K-2+), IDSSP(D2), CCSSI(k-2+), Hollylynne(Q2.3)	40	Insights from data	Dhar, Provost and Fawcett's, Hal Varian
11	Science	NGSS, k-12CS	26	Analysis	NGSS(P4), GAISE(G3), IDSSP(D4), CCSSI, Big Data(BIG3-k), STEM + C(C5), Hollylynne(Q5.1), Dhar, Hal Varian	41	Solving problems	Dhar, Cassie Kozyrkov
12	Technology & Tools	NGSS	27	Interpret/Extract Knowledge	GAISE(G4), Dhar, Provost and Fawcett's, Hal Varian	42	Making decisions	Dhar, Provost and Fawcett's, Cassie Kozyrkov
13	Reseach method	NGSS(P3), IDSSP(D2), Hollylynne(O1.2, O2.1, O2.2)	28	Prediction	IDSSP(D4), Big Data(BIG4.K.1.2)	43	Ethical	Provost and Fawcett's, Cassie Kozyrkov, Big Data
14	Questining	NGSS(P1), GAISE (G1), IDSSP(D1), Big Data(BIG1,BIG2), Hollylynne(O1.2)	29	System	Dhar	44	Biases	Hollylynne(Q2.3, O6.3), Cassie Kozyrkov
15	Listening Skill	IDSSP(D1)	30	Modeling and assessment	NGSS(P2), k-12CS(S4), IDSSP(D4), STEM + C(C11), , Hollylynne(Q5.1)	45	Questioning	NGSS(P1), GAISE(G1), BIG Data(BIG1.K1.2), Hollylynne(O1.2)

Literature Review 4 - Early Childhood Evidence for the Elements

This study collected evidence from professional early childhood organizations through identifying data Science standard and framework from engineering, CT, math, and CS fields. For example, the Child Observation Record (COR) Advantage Model helps teachers gain high-quality insight into students' positive developments using 36 elements from child development skills, such as problem solving skills, tools and technology, and data analysis developed by HighScope Educational Research Foundation [25]. Another identified framework was STEM Plus Computation (STEM + C), which is CT, engineering, and CS model for K–12 STEM education. The goal of this framework is to apply STEM and CT concepts to solve real-world problems [26]. Similarly, Engineering Cycle, developed by Marina Bers, helps early childhood learners solve a problem by building a physical item in six steps [27]. The Development and Research in Early Math Education (DREME) network was created to advance the field of early mathematics learning research and improve young children's opportunities to develop math skills. They proposed needed foundations to understand data, especially measured data, including Data in the Preschool Classroom, Important Ideas about Data, Solving Problems and Answering Questions with Data, and Development of Children's Thinking on Measurement & Data [28]. This research also investigated the Alabama Early Childhood Standard, which includes all milestones for early childhood development stages. Some milestones are elements related to data science foundations as analyzed by HighScope's COR guidelines [29]. The K–12 Computer Science Framework also develops practices for CS in preschools, stating that early childhood education is not a departure from traditional CS which would be developmentally appropriate. This framework utilizes four play-based pedagogies (patterns, problem solving, representation, sequencing) embedded within the core content areas of math, literacy, science, and social and emotional learning [22].

Filtering, Validating, and Adding

The literature reviews identified 45 elements (Table 2). Figure 2 shows the final analysis results of the clustering, grouping, and elimination of redundant elements. Elements were eliminated and merged by similarities, then clustered by purpose, and then validated for suitability by comparing with literature review 4 were highlighted by gray and the reference linked to it each standard has its own color. Missing elements recommended by literature review 4 are shown in orange in the figure. Overall, 11 stages were identified: Foundation, Identify Problem, Collect Data, Analyze, Create a Template, Ask Questions, Final Goal, Make a Story, Share It!, Plan Investigation, and Imaging the Process.



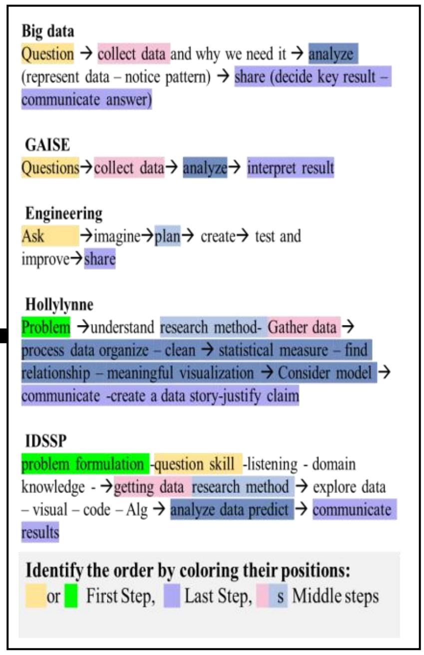


Figure 2. Datatying Elements (After Step 7)

Figure 3. Relationship Between Elements

Connecting Elements

The considered sequence depends on a child's ability to cognitively progress through the Datatying framework. The first and last steps are in agreement with the literature reviews, as shown in Figure 3. The sequence always begins with identifying or understanding a given problem and ends with sharing. Because children's problem-solving skills start developing at age 3 and identifying suitable decision-making problems require high levels of thinking, young children would be overwhelmed from the level of the cycled required to collect and analyze data. Therefore, Understanding Given Problem is more suitable for the Datatying framework than Asking Questions. The last step in the Datatying framework, Sharing and Presenting, requires the child to make a connection between the information and use a story to present their understanding. Consequently, the logical order of steps is Collect, Analyze, Make a Story, and Share. Plan, Imagine, and Create fit in the middle of the framework, where a child can plan by asking, imagine by showing how will it happen and what is needed, and create a template to collect the data. The foundational skills needed for Datatying are not included in the diagram, but to determine the suitable activities when developing a Datatying lesson.

Framework Validation and Creation

Framework evaluation feedback was collected from two STEM early childhood teachers who recommended decreasing the number of steps since each task was a semi-new task for the students. They also recommended introducing one new topic between familiar activities to focus training on the new skills. They also suggested merging the Imagine, Create Template, and Plan steps and replacing the Create Template step with a ready-made template that can intentionally teach the analysis skills. In addition, the cycle should be limited to two cycles with small data sets that allow children to see the same data twice but look at something different which another fundamental skill needs to be considered. Figure 4. Datataying Framework Version 1 (Before Step 8 show the Datataying framework before and after validation.

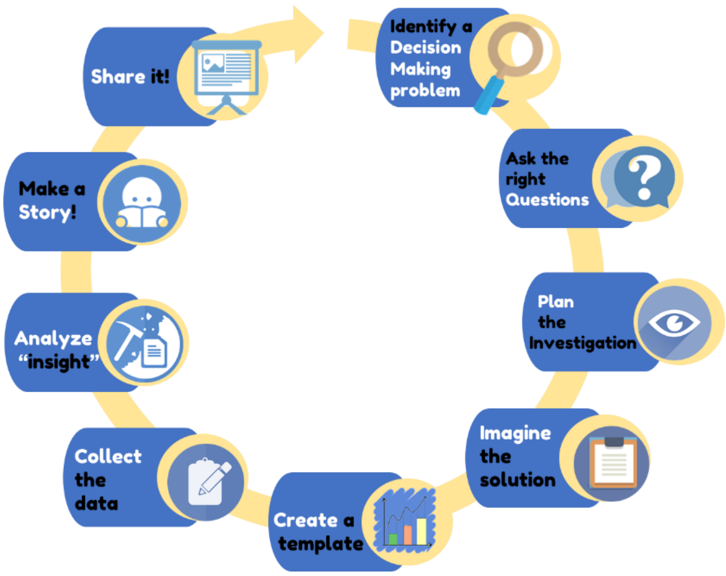


Figure 4. Datataying Framework Version 1 (Before Step 8)

Figure 5. Datataying Framework

Summary of Findings

The final Dataying framework, shown in Figure 4. Dataying Framework Version 1 (Before Step 8 consists of six progression stages to teach Dataying thinking through prepared decision-making problems with data sets. The Dataying thinking skills comprise a set of foundations from 12 subjects summarized from the literature reviews presented in Figure 6.

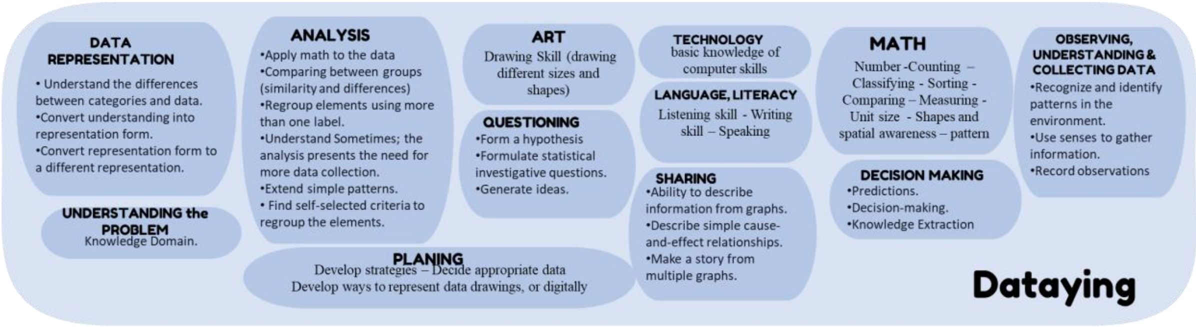


Figure 6. Dataying Foundations

Identify Decision-Making Problem

Engineering and coding problems typically follow steps that eventually result in physical or digital projects, such as teaching a robot to dance or creating an app. However, a “problem” in early childhood education typically refers to a dilemma that can be solved through thinking and implementation, meaning this model for early childhood education required a data-driven problem, which included identifying the appropriate decision-making problem. Very young children are still developing their decision-making skills and reasoning, so they cannot be expected to think by themselves and identify a problem that requires data science. Therefore, the teacher must share an appropriate level that requires fundamental Dataying skills. Choosing the right problem is the key to successful training, so the decision-making problem can be defined as a problem that requires the collection of data to gain insight on how to make a design. Compared to solving a math problem, a decision-making problem is not finalized with a quantitative answer—it requires data collection and then insightful predictions to make a decision or extract knowledge. Consider the following early childhood prediction example: Do you think Laila would like us to surprise her by taking her to Chuck E. Cheese for her birthday? Decision Making: What would be a good gift for Laila’s birthday? A bad problem requires one quantitative answer in one cycle, such as: How many gifts can fit in a box? Good problems require a child to make a fair decision while choosing for others, specifically one quantitative answer developed through several cycles and data collection. For example: What would be a good gift for your sick grandma? The answer should be only one gift. The child needs to understand that people collect data to answer questions, and that data can vary around the world.

Asking the Right Questions

Asking the right question can shape the succession of the experiment. The child must be guided to formulate statistically investigative questions that answer the problem. Brainstorming can help generate ideas. Prior to the Questions step, the child must demonstrate an understanding of the proposed problem. Considering the previous example of determining a gift for a sick grandmother, the right questions can be Q1: What does Grandma like?, or Q2: What does Grandma need?. An example of a bad question would be Q3: What does Grandma hate?. This step teaches the child that there is always more than one right answer and more than one way to approach the right answer.

Plan the Investigation

The step Plan the Investigation is similar to academic research methods but for children. The child must plan (1) where to get the data (e.g., mother, father, grandfather), (2) how to reach the source (e.g., when I go home, call them), (3) how the data will be collected (e.g., paper and pen, taking a picture, etc.), (4) how the data will be organized (e.g., list, draw in circle), and (5) how to prepare the needed material.

This study proposes Insight-Detective steps to help young children learn using basic counting, comparing, and pattern recognition. The Insight-Detective method utilizes small data sets that can teach Datatying concepts with minimum record to help the child recognize the overall picture to enhance future analysis and simplify the math required to solve the problem. The collected data should be from two resources, where each source provides two elements that can be logically grouped into a 2x2 visualization to help children focus on exploring rather than focusing on math where the answer of the hardest comparing is always less or equal than 4 which is suitable for 4 years old math ability. In addition, writing numbers and drawing which is still suitable. The second analysis transforms the 2x2 visualization into three groups: one group with two elements and two groups with one element each. Because decision making requires math, this method reduces the number of required steps to solve the problem by limiting the process to two cycles. This method also creates two figures from the same elements with different labels to help children understand data representation. For example, the mother and the father are two resources that can each give two options. Assuming the mother gave an orange and flowers and the father gave a banana and cards, the first analysis grouped the resources "mother" and "father," while the second analysis grouped the usage (e.g., "food," "plants," "papers"). Figure 7. Insight-Detective Worksheet shows the Insight-Detective worksheet plan template developed using this concept.

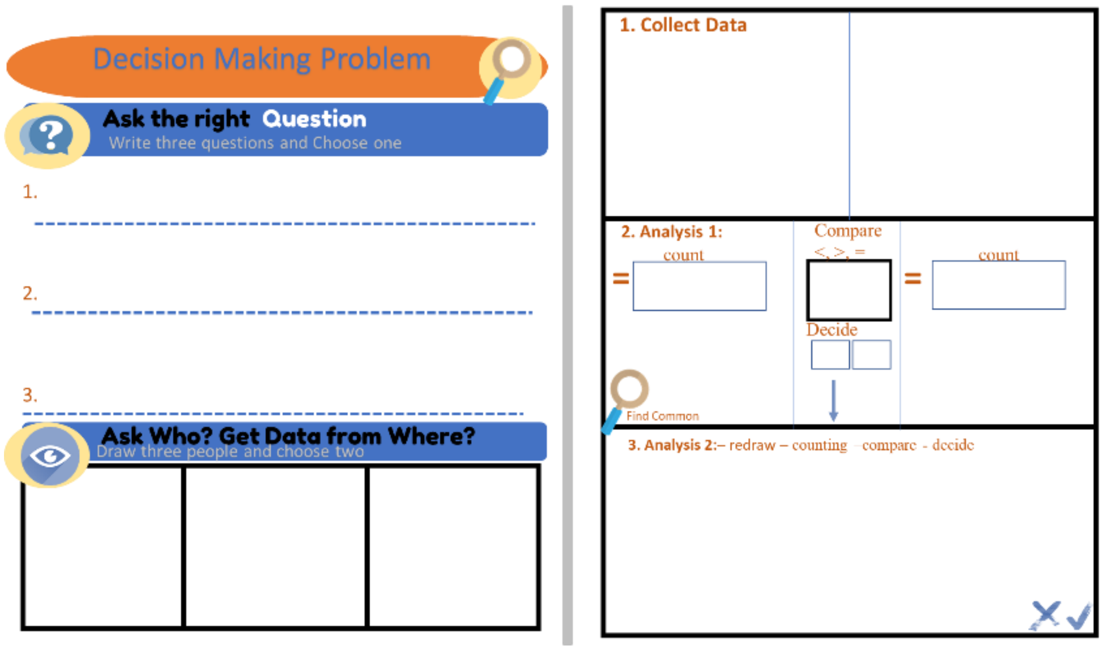


Figure 7. Insight-Detective Worksheet

Collect the Data

Data collection requires the child to convert understanding from seeing and hearing into representable shapes, such as marks, drawings, writing or digital tablet. In other words, the child must use their senses to gather data into several representation.

Analyze

In the Analysis step, a child attempts to make sense of the data until insight is obtained. The main goal of this step is to drive and support the analysis skill to train insight finding. Suggested steps for analyzing in early childhood education include identifying ways to categorize, counting and recording the number of data in each category, comparing numbers between categories, making a conclusion, reinvestigating (requires the child to look for common factors in the data, such as usage, size, color, weight, or taste), regrouping the data using identified elements, and deciding yes after identifying going to step 2-4. For example, if the mother gave an orange and flowers and the father gave a banana and cards, the data could be categorized as the mother's answers and the father's answers, each with two elements. Since both have two options, counting does not help with the decision making, and identifying commonalities between the answers is not helpful because all the options are distinct. However, considering the answers in light of usage could lead to food, flower, and card categories, which could create a hypothesis according to the mother's and the father's opinions. Because the example states that the grandmother is sick, the question could be related to whether an orange or a banana is better for sick people. The child would then be asked to circle the chosen answer.

Make a Story and Share It!

In this final step, the child must utilize the first and second analyses and the drawing to create a story. The child tells a story from the Dataying framework, including identifying what the problem was, who was involved, and how the decision was made. The child then shares the drawing.

Limitations and Future Work

This investigation only included five primary disciplines, but other disciplines may also utilize data science concepts. In addition, future work should use exploratory mixed-method research to validate the framework, and the work should be expanded to include integration paths. This literature review prompted the development of a data science tool for early childhood education, specifically young children who are still developing their reading abilities. This tool could be built from drag-and-drop concepts inspired by the ScratchJr app.

Acknowledgment

This work was funded by the National Science Foundation (NSF) with Grant No DRL GEGI008182. However, the authors alone are responsible for the opinions expressed in this work and do not reflect the views of the NSF.

References

[1] "Occupational outlook handbook - Data Scientist." United States Department of Labor. https://www.bls.gov/ooh/math/data-scientists.htm

[2] A. Swan and S. Brown, "The skills, role and career structure of data scientists and curators: An assessment of current practice and future needs," 2008.

[3] J. Louie, "Critical data literacy: Creating a more just world with data," ed: National Academies of Sciences, Engineering, and Medicine.

[4] Z. Drozda, D. Johnstone, and B. Van Horne, "Previewing the National Landscape of K-12 Data Science Implementation," ed: Workshop on Foundations of Data Science for Students in Grades K-12, 2022.

[5] S. Papadakis, "Robots and robotics kits for early childhood and first school age," 2020.

[6] M. U. Bers, "Coding and computational thinking in early childhood: the impact of ScratchJr in Europe," *European Journal of STEM Education*, vol. 3, no. 3, p. 8, 2018.
[7] S. Bas and G. Tegan. "What Is a Conceptual Framework? | Tips & Examples." https://www.scribbr.com/methodology/conceptual-framework/
[8] C. Carroll, M. Patterson, S. Wood, A. Booth, J. Rick, and S. Balain, "A conceptual framework for implementation fidelity," *Implementation science*, vol. 2, pp. 1-9, 2007.
[9] S. A. Malallah, "Developing computational thinking best practices for early childhood education in Kuwait and United States," Kansas State University, 2022.
[10] youcubed. "Data Science K-10 Big Ideas " https://www.youcubed.org/data-big-ideas. (accessed 2023).
[11] D. Donoho, "50 years of data science," *Journal of Computational and Graphical Statistics*, vol. 26, no. 4, pp. 745-766, 2017.
[12] A. Z. Spector, P. Norvig, C. Wiggins, and J. M. Wing, "Data Science in Context: Foundations, Challenges, Opportunities," 2022.
[13] F. Provost and T. Fawcett, *Data Science for Business: What you need to know about data mining and data-analytic thinking*. " O'Reilly Media, Inc.", 2013.
[14] B. Baumer, "A data science course for undergraduates: Thinking with data," *The American Statistician*, vol. 69, no. 4, pp. 334-342, 2015.
[15] C. Yong. "A Minimalism Approach to Understand What Data Science Is." https://towardsdatascience.com/what-is-data-science-4565e457cc17
[16] "Analyzing and Interpreting Data." Next Generation Science Standards (NGSS) https://ngss.nsta.org/Practices.aspx?id=4
[17] E. National Academies of Sciences, and Medicine, *Data science for undergraduates: Opportunities and options*. National Academies Press, 2018.
[18] A. Bargagliotti *et al.*, "Pre-K–12 guidelines for assessment and instruction in statistics education II (GAISE II): A framework for statistics and data science education," in *American Statistical Association*, 2020.
[19] H. LEE, G. MOJICA, E. THRASHER, and P. Baumgartner, "Investigating data like a data scientist: Key practices and processes," *Statistics Education Research Journal*, vol. 21, no. 2, pp. 3-3, 2022.
[20] C. C. S. S. Initiative., "Common Core State Standards for MATHEMATICS," ed.
[21] L. Years. "Data Science Course Framework." https://www.utdanacenter.org/our-work/k-12-education/launch-years-initiative/launch-years-initiative-course-frameworks
[22] K.-C. S. F. S. Committee, *K-12 computer science framework*. ACM, 2016.
[23] Denitsay. "How to take on the Data Science career path right after college?" (accessed 2023).
[24] IDSSP, "Curriculum frameworks for introductory data science," 2019.
[25] HighScope. "Child Observation Record." https://highscope.org/wp-content/uploads/2018/03/COR-Advantage-to-KDIs_March-2018-SS.pdf
[26] M. M. Hynes *et al.*, "Inspiring young children to engage in computational thinking in and out of school (research to practice)," in *2019 ASEE Annual Conference & Exposition*, 2019.
[27] M. U. Bers, *Coding as a playground: Programming and computational thinking in the early childhood classroom*. Routledge, 2017.
[28] DREME. "What Children Know and Need to Know about Data." https://prek-math-te.stanford.edu/measurement-data/what-children-know-and-need-know-about-data (accessed 2023).
[29] HighScope. "Alignment of COR Advantage with Alabama early childhood standards." https://highscope.org/wp-content/uploads/2018/03/2012-Alabama-Preschool-to-COR-Advantage-1.5_7-5-16.pdf